Statistical Methods in Medical Research 2005; **14**: 553–565

A multilevel modelling solution to mathematical coupling

Andrew Blance, Yu-Kang Tu Leeds Dental Institute, University of Leeds, Leeds, UK and Biostatistics Unit, Centre of Epidemiology and Biostatistics, University of Leeds, Leeds, UK and **Mark S Gilthorpe** Biostatistics Unit, Centre of Epidemiology and Biostatistics, University of Leeds, Leeds, UK

Owing to mathematical coupling, statistical analyses relating change to baseline values using correlation or regression are erroneous, where the statistical procedure of testing the null hypothesis becomes invalid. Alternatives, such as Oldham's method and the variance ratio test, have been advocated, although these are limited in the presence of measurement errors with non-constant variance. Furthermore, such methods prohibit the consideration of additional covariates (e.g., treatment group within trials) or confounders (e.g., age and gender). This study illustrates the more sophisticated approach of multilevel modelling (MLM) which overcomes these limitations and provides a comprehensive solution to the analysis of change with respect to baseline values. Although mathematical coupling is widespread throughout applied research, one particular area where several studies have suggested a strong relationship between baseline disease severity and treatment effect is guided tissue regeneration (GTR) within dental research. For illustration, we use GTR studies where the original data were found to be available in the literature for reanalysis. We contrast the results from an MLM approach and Oldham's method with the standard (incorrect) approach that suffers from mathematical coupling. MLM provides a robust solution when relating change to baseline and is capable of simultaneously dealing with complex error structures and additional covariates and/or potential confounders.

1 Introduction

Clinical researchers are keen to know whether the patients with more severe disease at baseline will obtain greater treatment benefits from interventions, that is, whether treatments work better in patients with more serious conditions. This is known as *differential* baseline effects on treatment outcomes. The most commonly used statistical approach to this problem is to test the association between change in clinical parameters of disease and their baseline values. However, statistical analyses relating change to baseline values using correlation or regression, to test differential baseline effects on treatment outcomes, are highly questionable because of mathematical coupling.[1,2] In general, mathematical coupling occurs when one variable contains directly or indirectly the whole or part of another, and the two variables are then analysed using correlation or regression.[3–8] The most common form of mathematical coupling in clinical research

Address for correspondence: Dr Mark S Gilthorpe, Biostatistics Unit, Centre of Epidemiology and Biostatistics, University of Leeds, 30/32 Hyde Terrace, Leeds LS2 9LN, UK. E-mail: m.s.gilthorpe@leeds.ac.uk

 10.1191/0962280205sm418oa

occurs when investigating change (e.g., following an intervention) in relation to initial or baseline values (i.e., prior to the intervention).[4] So, the statistical procedure of testing the null hypothesis – that the coefficient of correlation or the slope of regression is zero – becomes inappropriate,[8] and conclusions from studies that analyse change in this way are highly questionable.

A simple approach to overcome the problem of analysing change against baseline value was first proposed by Oldham,[3] with the variance ratio test[9] being proposed sometime later (although this test was invented earlier than Oldham's method but not explicitly proposed for this problem until much later). Neither method permits the simultaneous consideration of additional covariates (e.g., treatment group or potential confounders) in the usual way that multiple regression would. Furthermore, the underlying premise of both solutions requires a constant variance structure for measurement errors, which may not always be valid. So, the use of Oldham's method, or any alternative that depends on similar assumptions, is limited and an alternative strategy is required. This article introduces a multilevel modelling approach (MLM)[10] and presents the circumstances under which this is reasonable,[11–13] using an illustration from the field of dental research.

Guided tissue regeneration (GTR) is a dental surgical method of placing a barrier membrane to selectively promote growth of periodontal (gum) tissue in wound healing. The two most commonly used clinical measures of periodontal disease status are: pocket probing depth (PPD) and clinical attachment level (CAL). PPD is the distance from gingival margin (where the tooth emerges proud of the gum) to the bottom of gingival pocket (space between the gum and the tooth). CAL is the distance from the cemento-enamel junction (transition point between exposed tooth and root) to the bottom of gingival pocket. Using correlation and/or regression, many studies have shown that the treatment effects of GTR are highly associated with baseline disease severity. For instance, PPD reduction or CAL gain (i.e., change in PPD or CAL) is highly correlated with initial PPD or CAL.

In our previous studies, it was shown that evidence from the periodontal literature investigating GTR, regarding the dependence of treatment effects on the baseline disease severity, is potentially erroneous owing to mathematical coupling.[1,2] Where available, data from studies using GTR to treat intrabony defects are re-evaluated using MLM. Results are contrasted to analyses that ignore mathematical coupling (an incorrect analysis) and those that use Oldham's method (a correct analysis when the error variance is constant). Within a range of clinical scenarios, the assumption of constant error variance may be unreasonable. In periodontology, for instance, measurement errors may be a function of the outcome.[14] Although our illustration does not seek to explore different error variance structures, MLM can be extended to address this. The purpose of this study is to introduce the MLM strategy to this ongoing problem.

We initially outline the issue of mathematical coupling, followed by the methods of both Oldham and the variance ratio test, highlighting their limitations, and detail the MLM alternative. GTR data available from the literature for re-analysis are introduced. The MLM results are compared with Oldham's method, and both are contrasted with the uncorrected correlation. Key results are discussed and final conclusions are drawn.

2 Defining observed values

Let the true (unobserved) pre/post-treatment values be denoted by X_0 and X_1, respectively. The observed pre-treatment values are $x_0 = X_0 + e_0$, where e_0 are the measurement errors for X_0. The observed post-treatment values are $x_1 = X_1 + e_1$, where e_1 are the measurement errors for X_1. As both e_0 and e_1 are random errors, they have zero means with variances denoted δ_0^2 and δ_1^2, respectively. δ_0^2 and δ_1^2 are uncorrelated with each other and uncorrelated with X_0 and X_1, respectively. Therefore, the variance of x_0 (s_0^2) is the sum of the variance of X_0 (σ_0^2) and the variance of e_0 (δ_0^2); similarly, the variance of x_1 (s_1^2) is the sum of variance of X_1 (σ_1^2) and the variance of e_1 (δ_1^2).

3 Mathematical coupling

Change between observed pre-treatment clinical values (x_0) and observed post-treatment values (x_1) following an intervention is often explored using correlation/regression. However, assessing change ($x_1 - x_0$) in relation to baseline (x_0) using correlation/regression is flawed because of mathematical coupling, and this is readily appreciated by considering the theoretical value of the correlation between $x_1 - x_0$ and x_0:[3]

$$\mathrm{Corr}(x_1 - x_0, x_1) = \frac{s_1 - rs_0}{\sqrt{s_0^2 + s_1^2 - 2rs_0s_1}}$$

where s_0^2 is the variance of x_0, s_1^2 the variance of x_1 and r the correlation between x_0 and x_1.

When assessing differences of repeated measures on the same individuals, r is often positive (between zero and 1), although it rarely attains unity because of measurement errors or heterogeneous treatment responses. The correlation between $x_1 - x_0$ and x_0 is, therefore, more often positive than negative. With consistent error variances across measurements (i.e., $\delta_0^2 = \delta_1^2$) and no genuine 'baseline effect' ($\sigma_0^2 = \sigma_1^2$), we assume $s_0^2 = s_1^2$. Thus, if x_0 and x_1 are poorly correlated because of either the treatment response being heterogeneous and/or measurement errors being large, r is closer to zero than to 1 (i.e., regression to the mean), this 'spurious' positive association is large, approaching $1/\sqrt{2} \approx 0.7$. Conversely, if x_0 and x_1 are strongly correlated (i.e., the treatment effect is predictable and measurement errors are small), this spurious association is less marked, although still present. Thus, heterogeneity in treatment response and/or large measurement errors exacerbates mathematical coupling.

4 Oldham's method

Oldham's method is to evaluate the correlation between change ($x_1 - x_0$) and the arithmetic mean of the pre- and post-treatment values, $(x_1 + x_0)/2$. The expected

(theoretical) value of this correlation is:[3]

$$\text{Corr}\left[x_1 - x_0, \frac{(x_1 + x_0)}{2}\right] = \frac{s_1^2 - s_0^2}{\sqrt{(s_0^2 + s_1^2)^2 - 4r^2 s_0^2 s_1^2}}$$

where s_0^2, s_1^2 and r are defined previously. If there were no intervention and no differences in the measurement error variances across measurement occasions, the pre- and post-treatment variances would be identical. In contrast, as argued by many statisticians such as Hotelling,[15] Oldham[3] and Freidman,[16] if there was a genuine differential baseline effect on treatment outcome, the variances of pre-treatment and post-treatment values would be different. For instance, suppose periodontal defects with greater pre-treatment PPD can attain greater pocket depth reduction, the variance of post-treatment PPD should become smaller than that of pre-treatment PPD, because the values of post-treatment PPD will become 'closer' to each other as deeper pockets decrease more than shallow ones. Therefore, to test whether or not there is a differential baseline effect on the treatment outcome, we should test the differences in the pre-treatment and post-treatment variances rather than test whether the correlation between change and pre-treatment values is different from zero.

Oldham's method only provides an exception to the rule for the effects of mathematical coupling, where the impact of mathematical coupling is annulled. This is because the correlation between the difference, $x_1 - x_0$, and the arithmetic mean, $(x_1 + x_0)/2$, preserves *orthogonality* (i.e., they are independent) under the null hypothesis that there is no difference in the pre-treatment and post-treatment variances. Hence, the two terms remain coupled because of their formulaic relationship through x_0 and x_1, but mathematical coupling no longer creates any deleterious effects.

Oldham's method of analysing change with respect to baseline is also the basis of the Bland and Altman plots[17] for investigating agreement. Although not explicitly detailed by either Oldham or Bland and Altman, their approach readily extends to simple regression (i.e., where no additional covariates are considered). However, to include additional covariates, where these are also co-correlated with the mean, the exception to the rule is compromised because of the distortion caused by collinearity.[18] Extending Oldham's method to multiple regression is generally not a sound strategy.[13]

Oldham's method has been questioned among statisticians, and it has occasionally been criticized as being inferior to Blomqvist's formula[7,19] which was proposed in 1977 as a method to correct the bias owing to regression to the mean in the regression slope of analysing change against initial value.[20] Blomqvist's formula is given as:[21]

$$b_{\text{true}} = \frac{b_{\text{observed}} - k}{1 - k}$$

where k is the ratio of the measurement error variance (δ_0^2) to that of the observed variance (s_0^2).

In a widely cited article by Hayes, it was shown that if (1) individuals have been selected on the basis of high initial values or (2) the 'true' treatment effect differs across

individuals, then Oldham's method is biased (towards a negative association in Hayes article, although change was defined as $x_1 - x_0$, so the bias is positive in this article). Therefore, Hayes recommended Blomqvist's formula which seemed to perform better than Oldham's method for these two circumstances. However, although we agree with Hayes that Oldham's method will lead to biased results in scenario (1), there is a misunderstanding surrounding the circumstances in scenario (2) – where Oldham's method in fact gives rise to correct results.

The apparent contradiction between Oldham's method and Blomqvist's formula for scenario (2) arises owing to each method addressing different questions of the study data. Blomqvist's formula corrects the bias caused by measurement errors in the pre-treatment values to give an unbiased estimate of change conditional on the pre-treatment values, that is, Blomqvist's formula should be viewed as a correction method to solve the problem of measurement errors in explanatory variable, but this method does not intend to test whether there are differential baseline effects. Non-zero estimates by Blomqvist's formula do not imply, as has been incorrectly inferred, that there are *differential* baseline effects on the treatment outcome. Although Blomqvist's formula corrects for measurement errors and/or biological variation in pre-treatment values, it does not correct for the impact of regression to the mean caused by heterogeneous responses of the patients to the treatments. Unless the non-perfect correlation between x_0 and x_1 is due *only* to measurement errors (i.e., all patient responses to treatment are identical), Blomqvist's formula and Oldham's method give rise to different results because they address different questions and there is no inconsistency in scenario (2).

In summary, Blomqvist's formula provides an unbiased estimate of the slope for change regressed on baseline by correcting for measurement errors in x_0, but does not correct for regression to the mean, and so cannot be used to infer whether differential baseline effects are present or not; Oldham's method, in contrast, correctly tests for baseline effects, but cannot estimate the slope of change regressed on baseline.

5 The variance ratio test

The variance ratio σ_0^2/σ_1^2 has also been proposed as an appropriate test of differential baseline effect on treatment outcome.[9] As with Oldham's method, this approach also tests the equality of the correlated variances, such as the variances of two repeated measurements, yielding a t-distribution with $n - 2$ (where n is the sample size) degrees of freedom:[9]

$$t = \frac{(s_1^2 - s_0^2)\sqrt{n-2}}{2s_1 s_0 \sqrt{1-r^2}}$$

where s_0^2, s_1^2 and r are defined previously. The same assumptions and hence limitations apply to the variance ratio test as for Oldham's method, and it is impossible to consider the impact of additional covariates using this approach.

6 The MLM approach

The broader methodological principles of MLM are described in detail elsewhere.[22–25] For this problem, only a two-level model is needed, written as

$$\text{outcome}_{ij} = \beta_{0ij} + \beta_{1j}\,\text{occ}_{ij} + \sum_{m=2}^{N} \beta_m \text{covariates}_{mij} \tag{1}$$

where the outcome is either PPD or CAL on the ith occasion (level-1, $i = 0, 1$) for the jth patient (level-2, $j = 1, \ldots, n$, where n is the number of subjects); $\beta_{0ij} = \beta_0 + e_{0ij} + u_{0j}$ is the intercept term comprising a mean β_0 with normally distributed random variation *within* subjects ($e_{0ij} \sim N(0, \sigma^2_{e0})$ where σ^2_{e0} is the occasion-level variance) and normally distributed random variation *between* subjects ($u_{0j} \sim N(0, \sigma^2_{u0})$ where σ^2_{u0} is the subject-level variance); the regression slope $\beta_{1j} = \beta_1 + u_{1j}$ comprises a mean β_1 with normally distributed random variation between subjects ($u_{1j} \sim N(0, \sigma^2_{u1})$ where σ^2_{u1} is the slope variance across subjects) giving rise to a covariance σ_{u01} across subjects between the random intercept (u_{0j}) and the random slope (u_{1j}); occ_{ij} denotes measurement occasion centred about zero, for example, $\text{occ}_{0j} = -0.5$ and $\text{occ}_{1j} = +0.5$; and covariates_{mij} ($m \geq 2$) are additional covariates (i.e., explanatory variables) with regression coefficients β_m, for example, treatment arm of a trial (treatment/placebo), patient characteristics (age, gender, and so on) and/or interaction terms.

The outcomes and associated variance structure are defined differently within this multilevel framework than previously for the standard regression/correlation approach, and it is this that overcomes the problem of mathematical coupling. In this framework, modelling random intercepts and random slopes mean that each subject has a different baseline PPD (or CAL) score, while simultaneously exhibiting different changes in PPD (or CAL) following the intervention;[26] the slope β_1 represents the mean change in outcome owing to GTR treatment across all subjects.

The dummy coding of occ_{ij} will affect the results of the MLM approach. When occ_{ij} is coded as $\text{occ}_{0j} = -0.5$ and $\text{occ}_{1j} = 0.5$, the intercept β_0 represents the average of pre- and post-treatment values (Figure 1(a)) – the variance of the intercept is thus the variance of the average of pre- and post-treatment values. The slope β_1 represents the change in outcome between occasions – the variance of β_1 thus represents the variance of change. By centring occ_{ij}, the covariance between intercept (β_0) and change (β_1) in the MLM approach is akin to Oldham's method and theoretically will give rise to identical results. In contrast, if occ_{ij} coded as $\text{occ}_{0j} = 0$ and $\text{occ}_{1j} = 1$, the intercept β_0 represents pre-treatment values (Figure 1(b)) and the results of the MLM approach will be identical to the conventional approach of correlating change with baseline. When occ_{ij} is coded as $\text{occ}_{0j} = -1$ and $\text{occ}_{1j} = 0$, the intercept β_0 represents post-treatment values, and the results of the MLM approach will be identical to a conventional approach but where the correlation is between post-treatment values and change (Figure 1(c)). The covariance structure of the MLM approach thus depends on how occ_{ij} is coded.

With constant measurement error variance, Equation (1) has four random parameters to be estimated ($\sigma^2_{e0}, \sigma^2_{u0}, \sigma^2_{u1}$ and σ_{u01}) with only three degrees of freedom within the

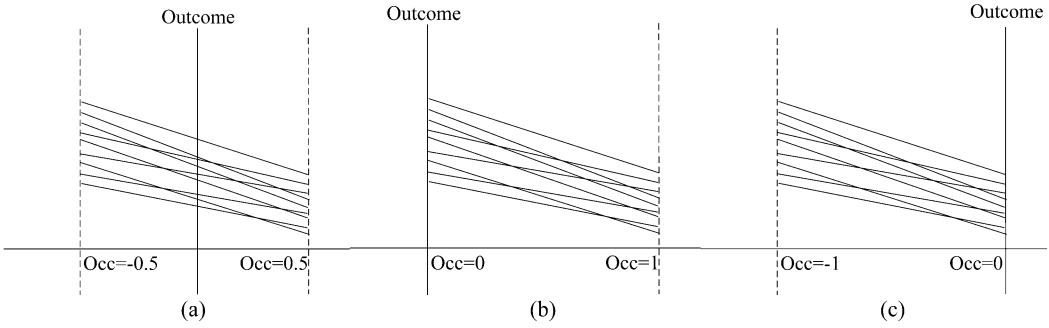

Figure 1 (a) When occ_{ij} is coded as $occ_{0j} = -0.5$ and $occ_{1j} = 0.5$, the intercept β_0 represents the average of pre-treatment and post-treatment values; (b) when occ_{ij} is coded as $occ_{0j} = 0$ and $occ_{1j} = 1$, the intercept β_0 represents the pre-treatment values only; (c) when occ_{ij} is coded as $occ_{0j} = -1$ and $occ_{1j} = 0$, the intercept β_0 represents the post-treatment values only.

data (variation on each occasion and variation in change (i.e., response to treatment) between occasions). It is, therefore, necessary to make further assumptions to reduce, by 1, the number of random parameters to be estimated. Such assumptions must be reasonable for the context in which the multilevel model is applied. For most clinical studies we are often unable to distinguish between measurement error (σ^2_{e0}) and biological variation (σ^2_{u0}), and hence these are typically treated as synonymous. We may thus estimate either σ^2_{e0} or σ^2_{u0}, although not both, by constraining one to be zero.[13] This does not affect our interpretation of either variance estimate, as the selected non-zero variance parameter represents the combined effects of measurement error and individual variation within the study group as a whole. To aid interpretation, it is convenient to constrain σ^2_{e0} to be zero, else one might have zero variance (σ^2_{u0}) and non-zero covariance (σ_{u01}) across subjects. Adopting this constraint simplifies the model to no longer be hierarchical, although this only occurs because we adopt the simplified scenario of ignoring measurement error. When extending this approach to consider explicitly measurement error, this is introduced via further non-zero constraints on the level-1 variance parameter, thereby returning to a hierarchical framework.

In any event, whatever the constraints adopted, the test of differential baseline effect on the outcome is derived from the covariance between the random intercept and random slope:[24]

$$\frac{\sigma_{u01}}{\sqrt{\sigma^2_{u0}\sigma^2_{u1}}}$$

and 95% confidence intervals are calculated using closed formulae.[27]

Models were fitted in the multilevel software *MLwiN*[28] using maximum likelihood estimation (the restricted iterative generalized least-squares procedure within *MLwiN*). As explained previously, fitting the simple model, without additional covariates, produces equivalent results to Oldham's method, providing the occasion covariate is

centred. However, in contrast to Oldham's method, MLM facilitates the simultaneous consideration of additional covariates in the usual way for single-level multiple regression, as this approach is free of mathematical coupling.

Within MLM, collinearity between the occasion covariate (occ_{ij}) and any additional covariates could only occur when the latter varies between measurement occasions. This is not feasible for most covariates, such as treatment regime (treatment/placebo), or for other confounders, such as patient characteristics (age, gender, and so on), or for any interactions among these terms. Furthermore, non-constant error variance can be modelled explicitly using the MLM approach, where an *a priori* understanding of the measurement error structure can be imposed, although this was not explored here.

7 The GTR literature

The periodontal literature on GTR shows that among a variety of clinical and host variables, the baseline disease severity, measured by clinical and radiographic parameters, is strongly associated with the treatment outcome using correlation/regression. The *Journal of Periodontology, Journal of Clinical Periodontology* and *Journal of Periodontal Research* were electronically and hand-searched between 1986 (when GTR was first proposed) and December 2002, for original data on the treatment effect of GTR in intrabony defects. Only human studies were included that contained actual measurements of PPD and/or CAL for each site, or in which these measurements could be derived from other clinical parameters. If a study contained two groups of patients using different treatment protocols, these were treated as two separate studies. Randomized controlled trials and case series were all included. Individual case reports and case series with less than nine cases were not included owing to consideration of sample size and statistical power.

Ten studies provided data of PPD and because two studies compared two different barrier membranes, these were each divided into two, yielding a total of 12 studies for PPD. There were 11 studies providing data of CAL, one study compared both GTR alone and GTR combined with bone grafting, and two studies compared two different barrier membranes, these three studies were also divided, yielding a total of 14 studies for CAL. The 14 studies identified as providing data for MLM re-analysis are listed in the appendix. These were previously re-analysed using Oldham's method and the variance ratio test,[2] with no difference in the inferences for either approach, although Oldham's method yields a correlation coefficient, which is more useful in the direct interpretation of any relationship between change and baseline. For this reason, we contrast the results of the MLM approach to Oldham's method only.

8 Comparison of statistical methods to determine 'baseline' effects

Of the 12 studies that explored PPD, 10 initially demonstrated a significant correlation between baseline value and subsequent change (unadjusted correlation). In contrast, both Oldham's method and the MLM approach identified three studies as yielding any significant baseline effect (Table 1). Of the 14 studies investigating CAL, nine initially

Table 1 Reanalysis of the data for the outcome variable PPD[a]

Study	*N*	Unadjusted correlation			Oldham/MLM method		
		$corr_1$	*P*-value	CI	$corr_2$	*P*-value	CI
7	47	***0.823***	***<0.001***	(***0.701***, ***0.898***)	***0.457***	***0.001***	(***0.196***, ***0.658***)
14	26	***0.743***	***<0.001***	(***0.499***, ***0.877***)	0.261	0.197	(−0.140, 0.589)
5	25	***0.666***	***<0.001***	(***0.368***, ***0.840***)	−0.079	0.707	(−0.460, 0.326)
12	23	***0.911***	***<0.001***	(***0.799***, ***0.962***)	***0.735***	***<0.001***	(***0.463***, ***0.880***)
1	19	***0.685***	***0.001***	(***0.336***, ***0.869***)	***0.493***	***0.032***	(***0.050***, ***0.774***)
6-i	10	***0.896***	***<0.001***	(***0.611***, ***0.975***)	0.438	0.205	(−0.264, 0.837)
13	10	***0.707***	***0.022***	(***0.139***, ***0.925***)	0.239	0.506	(−0.460, 0.755)
8-i	10	0.571	0.084	(−0.091, 0.883)	0.117	0.747	(−0.553, 0.695)
8-ii	10	***0.798***	***0.006***	(***0.339***, ***0.950***)	0.340	0.336	(−0.368, 0.799)
6-ii	9	0.598	0.089	(−0.110, 0.903)	−0.197	0.612	(−0.761, 0.538)
9	9	***0.829***	***0.006***	(***0.366***, ***0.963***)	0.111	0.776	(−0.597, 0.722)
11	9	***0.739***	***0.023***	(***0.148***, ***0.941***)	−0.018	0.963	(−0.674, 0.654)

[a]Studies are identified in the appendix (*6/8-i*, Guidor; *6/8-ii*, ePTFE); *N*, number of defects in each study; $corr_1$ and $corr_2$ are Pearson correlation coefficients for pre-Tx PPD versus ΔPPD and [pre-Tx PPD + post-Tx PPD]/2 versus Δ PPD, respectively, where Tx is treatment and Δ the change; $corr_2$ is also equivalent to the correlation between the random coefficients for the intercept and linear time derived from the multilevel models; shaded figures in bold italic are significant at the 5% level.

demonstrated a significant correlation between baseline and subsequent change. In contrast, both Oldham's method and the MLM approach identified four studies that showed any significant association (Table 2). One previously non-significant result became significantly negative for both Oldham's method and the MLM approach.

Table 2 Reanalysis of data for the outcome variable CAL[a]

Study	*N*	Unadjusted correlation			Oldham/MLM method		
		$corr_1$	*P*-value	CI	$corr_2$	*P*-value	CI
14	26	***0.530***	***0.005***	(***0.180***, ***0.761***)	−0.165	0.421	(−0.519, 0.238)
12	23	***0.889***	***<0.001***	(***0.752***, ***0.952***)	***0.676***	***<0.001***	(***0.365***, ***0.851***)
3	20	***0.527***	***0.017***	(***0.111***, ***0.786***)	0.233	0.322	(−0.233, 0.613)
4	20	***0.654***	***0.002***	(***0.298***, ***0.850***)	0.417	0.067	(−0.031, 0.726)
1	19	***0.581***	***0.009***	(***0.173***, ***0.819***)	***0.465***	***0.045***	(***0.013***, ***0.759***)
10-i	11	0.345	0.299	(−0.321, 0.783)	0.042	0.902	(−0.572, 0.626)
10-ii	11	***0.777***	***0.005***	(***0.332***, ***0.939***)	***0.625***	***0.040***	(***0.040***, ***0.891***)
2	10	0.079	0.828	(−0.579, 0.675)	−0.111	0.759	(−0.692, 0.557)
6-i	10	***0.696***	***0.025***	(***0.117***, ***0.922***)	0.573	0.083	(−0.088, 0.884)
13	10	***0.829***	***0.003***	(***0.417***, ***0.958***)	0.380	0.279	(−0.328, 0.815)
8-i	10	0.455	0.186	(−0.244, 0.843)	0.019	0.958	(−0.618, 0.641)
8-ii	10	***0.865***	***0.001***	(***0.516***, ***0.968***)	0.602	0.065	(−0.044, 0.893)
6-ii	9	0.176	0.651	(−0.553, 0.752)	***−0.720***	***0.029***	(***−0.936***, ***−0.107***)
9	9	0.453	0.221	(−0.302, 0.859)	0.210	0.588	(−0.528, 0.767)

[a]Studies are identified in the appendix (*6/8/10-i*, Guidor; *6/8/10-ii*, ePTFE); *N*, number of defects in each study; $corr_1$ and $corr_2$ are Pearson correlation coefficients for pre-Tx CAL versus ΔCAL and [pre-Tx CAL + post-Tx CAL]/2 versus ΔCAL, respectively, where Tx is treatment and Δ the change; $corr_2$ is also equivalent to the correlation between the random coefficients for the intercept and linear time derived from the multilevel models; shaded figures in bold italic are significant at the 5% level.

9 Discussion

Despite repeated warnings, the inappropriate application of statistics in clinical research continues unabated.[29–31] The phenomenon of mathematical coupling is an example of how methodological dilemmas languish within what otherwise seems straightforward and widely used statistical procedures. To confuse matters, some studies suffer a conceptual confusion between mathematical coupling and regression to the mean,[7,32] where the latter arises because of the presence of measurement error. In general, within any correlation/regression analysis of change and baseline, both regression to the mean and mathematical coupling are present and cannot be separated. Adjustment for measurement errors alone is insufficient. This article does not aim to adjust for measurement errors; rather it addresses the problem of mathematical coupling in the presence of measurement error and the purpose of this study was to illustrate how to overcome the problem of mathematical coupling when exploring change with respect to baseline. Adjustment for measurement errors within the MLM approach is feasible but beyond the scope of this article.

Findings presented here are consistent with evidence from a simulation study, which showed that there is a high probability of obtaining a spurious significant correlation, because of mathematical coupling, when there is no genuine association[1] – this probability approaches unity with increasing study size. It is, therefore, imperative that clinical researchers and medical statisticians recognize the dangers of using correlation/regression for analysing data that exhibit a coupled relationship. For the assessment of change with respect to baseline value, the adoption of Oldham's method is a simple though potentially naïve strategy. Furthermore, it would often be necessary to consider additional covariates. Within a randomized control trial, Oldham's method may be viable, as randomization should ensure negligible differences in mean scores between the treatment groups (hence no association between mean score and the treatment covariate). If, however, one wishes to control for other confounders, such as patient characteristics (age and gender), one cannot ensure that these are also independent of the mean scores, as patients cannot be randomized to these factors, and this could give rise to bias in the covariate estimate.[13] So, when seeking to determine the extent of any baseline effect, to control appropriately for additional covariates one should consider the MLM solution.

This study demonstrates that when there are only two repeated measurements, MLM will give rise to identical results to Oldham's method. However, when there are more than two repeated measurements, such as body growth in orthodontic data, Oldham's method will become less appropriate, because it cannot take interim measurements into account. In contrast, the MLM approach can utilize interim measurements to yield more accurate estimation of the linear trend in the change, and its correlation with the baseline values. Furthermore, this study also shows that when MLM is applied to multiple repeated measurements, proper coding of time variables is vital to yield correct results.

10 Conclusions

Researchers applying correlation/regression must remain mindful of the pitfalls that lie therein. Mathematical coupling should become widely understood and thus avoided in all future clinical research. Although Oldham's method provides a solution to the analysis of change with respect to baseline in isolation of all other considerations, this method is limited to circumstances of constant error variance and does not naturally extend to multiple regression or multiple repeated measurements. In general, the use of MLM is preferred, as this method completely removes the effects of mathematical coupling and facilitates the explicit modelling of additional covariates and complex error variance structures.

Acknowledgements

All three authors were funded by the United Kingdom government's Higher Education Funding Council for England (HEFCE). The authors would like to thank the referees for their very helpful comments on the redrafting of this article.

References

1 Tu YK, Gilthorpe MS, Griffiths GS. Is reduction of pocket probing depth correlated with the baseline value or is it "mathematical coupling"? *Journal of Dental Research* 2002; **81**: 722–26.
2 Tu Y-K, Maddick IH, Griffiths GS, Gilthorpe MS. Mathematical coupling can undermine the statistical assessment of clinical research: illustration from the treatment of guided tissue regeneration. *Journal of Dentistry* 2004; **32**: 133–42.
3 Oldham PD. A note on the analysis of repeated measurements of the same subjects. *Journal of Chronic Diseases* 1962; **15**: 969–77.
4 Archie JP. Mathematical coupling: a common source of error. *Annals of Surgery* 1981; **193**: 296–303.
5 Moreno LF, Stratton HH, Newell JC, Feustel PJ. Mathematical coupling of data: correction of a common error for linear calculations. *Journal of Applied Physiology* 1986; **60**: 335–43.
6 Stratton HH, Feustel PJ, Newell JC. Regression of calculated variables in the presence of shared measurement error. *Journal of Applied Physiology* 1987; **62**: 2083–93.
7 Hayes RJ. Methods for assessing whether change depends on initial value. *Statistics in Medicine* 1988; **7**: 915–27.
8 Andersen B. *Methodological errors in medical research*. London: Blackwell, 1990: 157–68.
9 Greenen R, van de Vijver FJR. A simple test of the law of initial value. *Psychophysiology* 1993; **30**: 525–30.
10 Gilthorpe MS, Tu YK. Mathematical coupling: a multilevel approach. *International Journal of Epidemiology* 2004; **33**: 1399–400.
11 White I. Assessing correlation between reporting errors and true values: untestable assumptions are unavoidable. *International Journal of Epidemiology* 2004; **33**: 1400–4011.
12 Rasbash J, Goldstein H. Mathematical coupling: a simpler approach. *International Journal of Epidemiology* 2004; **33**: 1401–4022.
13 Gilthorpe MS, Tu YK, Gunnell D. A coda: oversimplification, implicit assumptions, and measurement error. *International Journal of Epidemiology* 2004; **33**: 1402–403.
14 Fleiss JL, Mann J, Paik M, Goultchin J, Chilton NW. A study of inter- and intra-examiner reliability of pocket depth and attachment level. *Journal of Periodontal Research* 1991; **26**: 122–28.

15 Hotelling H. Book review: the triumph of mediocrity in business. *Journal of the American Statistical Association* 1933; **28**: 463–65.
16 Freidman M. Do old fallacies ever die? *Journal of Economic Literature* 1992; **30**: 2129–32.
17 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **1**: 307–10.
18 Glantz SA, Slinker BY. *Primer of applied regression and analysis of variance*, Second edition. New York: McGraw-Hill, 2001: 185–240.
19 Kirkwood BR, Sterne JA. *Medical statistics*. London: Blackwell, 2003.
20 Blomqvist N. On the relation between change and initial value. *Journal of the American Statistical Association* 1977; **72**: 746–49.
21 Blomqvist N. On the bias caused by regression toward the mean in studying the relation between change and initial value. *Journal of Clinical Periodontology* 1987; **14**: 34–37.
22 Gilthorpe MS, Maddick IH, Petrie A. Introduction to multilevel modelling in dental research. *Community Dental Health* 2000; **17**: 222–26.
23 Gilthorpe MS, Griffiths GS, Maddick IH, Zamzuri AT. The application of multilevel modelling to periodontal research data. *Community Dental Health* 2000; **17**: 227–35.
24 Gilthorpe MS, Cunningham SJ. The application of multilevel, multivariate modelling to orthodontic research data. *Community Dental Health* 2000; **17**: 236–42.
25 Gilthorpe MS, Griffiths GS, Maddick IH, Zamzuri AT. An application of multilevel modelling to longitudinal periodontal research data. *Community Dental Health* 2001; **18**: 79–86.
26 Gilthorpe MS, Zamzuri AT, Griffiths GS, Maddick IH, Eaton KA, Johnson NW. Unification of the "burst" and "linear" theories of periodontal disease progression: a multilevel manifestation of the same phenomenon. *Journal of Dental Research* 2003; **82**: 200–205.
27 Altman DG, Gardner MJ. Regression and correlation. In Altman DG, Machin D, Bryant TN, Gardner MJ eds. *Statistics with confidence*, Second edition. London: BMJ, 2000: 89–90.
28 Rasbash J, Browne W, Cameron B, Charlton C. *MLwiN version 1.10.0006*. London: Multilevel Models Project, Institute of Education, 2000.
29 Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials: a survey of three medical journals. *New England Journal of Medicine* 1987; **317**: 426–32.
30 Altman DG. The scandal of poor medical research. *British Medical Journal* 1994; **308**: 283–84.
31 Altman DG. Statistics in medical journals: some recent trends. *Statistics in Medicine* 2000; **19**: 3275–89.
32 Irwig L, Groeneveld H, Becklake M. Relationship of lung function loss to level of initial function: correcting for measurement error using reliability coefficient. *Journal of Epidemiology and Community Health* 1988; **42**: 383–89.

APPENDIX: Studies re-analysed in this study

1 Al-Arrayed F, Adams S, Moran J, Dowell P. Clinical trial of cross-linked human type I collagen as a barrier material in surgical periodontal treatment. *Journal of Clinical Periodontology* 1995; **22**: 371–79.
2 Chung KM, Salkin LM, Stein MD, Freedman AL. Clinical evaluation of a biodegradable collagen membrane in guided tissue regeneration. *Journal of Periodontology* 1990; **61**: 732–36.
3 De Sanctis M, Clauser C, Zucchelli G. Bacterial colonization of barrier material and periodontal regeneration. *Journal of Clinical Periodontology* 1996; **23**: 1039–46.
4 De Sanctis M, Zucchelli G, Clauser C. Bacterial colonization of bioabsorbable barrier material and periodontal regeneration. *Journal of Periodontology* 1996; **67**: 1193–200.
5 Garret S, Loos B, Chamberlain D, Egelberg J. Treatment of intraosseous periodontal defects with a combined adjunctive therapy of citric acid conditioning, bone grafting, and placement of collagenous membranes.

Journal of Clinical Periodontology 1988; **15**: 383–9.
6 Karapataki S, Hugoson A, Falk H, Laurell L, Kugelberg CF. Healing following GTR treatment of intrabony defects distal to mandibular 2nd molars using resorbable and non-resorbable barriers. *Journal of Clinical Periodontology* 2000; **27**: 333–40.
7 Laurell L, Falk H, Fornell J, Johard G, Gottlow J. Clinical use of a bioresorbable matrix barrier in guided tissue regeneration therapy. Case series. *Journal of Periodontology* 1994; **65**: 967–75.
8 MacDonald ES, Nowzari H, Contreras A, Flynn J, Morrison J, Slots J. Clinical and microbiological evaluation of a bioresorbable and a nonresorbable barrier membrane in the treatment of periodontal intrasseous lesions. *Journal of Periodontology* 1998; **69**: 445–53.
9 Mattson J, McLey L, Jabro M. Treatment of intrabony defects with collagen barrier membrane barriers. Case reports. *Journal of Periodontology* 1995; **66**: 635–45.
10 Mellado JR, Salkin LM, Freedman AL, Stein MD. A comparative study of e-PTFE periodontal membranes with and without decalcified freeze dried bone allografts for the regeneration of interproximal intraosseous defects. *Journal of Periodontology* 1995; **66**: 751–55.
11 Proestakis G, Bratthall G, Söderholm G, Kullendorff B, Gröndahl K, Rohlin M *et al*. Guided tissue regeneration in the treatment of infrabony defects on maxillary premolars. *Journal of Clinical Periodontology* 1992; **19**: 766–73.
12 Quteish D, Dolby AE. The use of irradiated-crosslinked human collagen membrane in guided tissue regeneration. *Journal of Clinical Periodontology* 1992; **19**: 476–84.
13 Silvestri M, Ricci G, Rasperini G, Sartori S, Catteneo V. Comparison of treatments of infrabony defects with enamel matrix derivative, guided tissue regeneration with a nonresorbable membrane and Widman modified flap. *Journal of Clinical Periodontology* 2000; **27**: 603–10.
14 Zucchelli G, De Sanctis M, Clauser C. Integrated connective tissue in bioresorbable barrier material and periodontal regeneration. *Journal of Periodontology* 1997; **68**: 996–1004.