



Article

Repeated Measures Regression in Laboratory, Clinical and Environmental Research: Common Misconceptions in the Matter of Different Within- and Between-Subject Slopes

Donald R. Hoover ^{1,*}, Qiuhu Shi ², Igor Burstyn ³  and Kathryn Anastos ⁴

¹ Department of Statistics and Biostatistics and Institute for Health, Health Care Policy and Aging Research, Rutgers University, Piscataway, NJ 08854, USA

² School of Health Sciences and Practice, New York Medical College, Valhalla, NY 10595, USA; Qiuhu_shi@nysmc.edu

³ Environmental and Occupational Health Dornsife School of Public Health, Philadelphia, PA 19104, USA; igor.burstyn@drexel.edu

⁴ Albert Einstein College of Medicine, Montefiore Medical Center, Bronx, NY 10467, USA; kanastos@montefiore.org

* Correspondence: drhoover@stat.rutgers.edu; Tel.: +1-732-445-2691; Fax: +1-732-445-3428

Received: 21 October 2018; Accepted: 6 February 2019; Published: 11 February 2019



Abstract: When using repeated measures linear regression models to make causal inference in laboratory, clinical and environmental research, it is typically assumed that the within-subject association of differences (or changes) in predictor variable values across replicates is the same as the between-subject association of differences in those predictor variable values. However, this is often false. For example, with body weight as the predictor variable and blood cholesterol (which increases with higher body fat) as the outcome: (i) a 10-lb. weight increase in the same adult affects more greatly an increase in cholesterol in that adult than does (ii) one adult weighing 10 lbs. more than a second indicate higher cholesterol in the heavier adult. A 10-lb. weight gain in the first adult more likely reflects a build-up of body fat in that person, while a second person being 10 lbs. heavier than the first could be influenced by other factors, such as the second person being taller. Hence, to make causal inferences, different within- and between-subject slopes should be separately modeled. A related misconception commonly made using generalized estimation equations (GEE) and mixed models on repeated measures (i.e., for fitting cross-sectional regression) is that the working correlation structure only influences variance of the parameter estimates. However, only independence working correlation guarantees that the modeled parameters have interpretability. We illustrate this with an example where changing the working correlation from independence to equicorrelation qualitatively biases parameters of GEE models and show that this happens because within- and between-subject slopes for the outcomes regressed on the predictor variables differ. We then systematically describe several common mechanisms that cause within- and between-subject slopes to differ: change effects, lag/reverse-lag and spillover causality, shared within-subject measurement bias or confounding, and predictor variable measurement error. The misconceptions we describe should be better publicized. Repeated measures analyses should compare within- and between-subject slopes of predictors and when they do differ, investigate the causal reasons for this.

Keywords: within-/between-subject associations; repeated measures; cross-sectional regression; generalized estimating equations; mixed models; working correlation structure

1. Introduction

We focus on two common misconceptions that are made in research while fitting repeated measures regression with generalized estimating equations (GEE) and mixed models (MM). Misconception-A: The association between the predictor variable and the outcome across different measures from the same subject (within-subject) is the same as the association of that variable with the outcome between measures from different subjects (between-subject). In fact, these associations often differ, which should be considered when making causal inference. For example, consider weight as the predictor and cholesterol the outcome given the well-known association of higher serum cholesterol and with greater body fat: (i) a 10 lb. increase in the same adult more likely indicates greater difference in serum cholesterol than does (ii) one adult being 10 lbs heavier than a second adult. A 10-pound weight gain in the same adult more likely reflects a build-up of body fat in that person, while the first adult being 10 pounds heavier than the second could be influenced by other factors such as the first adult being taller than the second. Misconception-B: The working correlation structure used in GEE and MM models is only a nuisance factor that impacts precision of model parameter estimates. As illustrated and explained in the next Section (and Table 1), the wrong choice for working correlation structure biases parameter estimates.

Both of these misconceptions are related, but the analytical details are complicated. To explore this further, Section 2 begins with an illustration of Misconception-B in real data. Section 2 also explains how this relates to Misconception-A and why independence working correlation must be used for creation of predictive models using “cross-sectional regression” on repeated measures. Then, Section 3 details how separation into within- and between-subject associations is needed for using repeated measures regression to make causal inference. Section 4 describes epidemiological mechanisms that can cause within- and between-subject slopes to differ. Section 6 summarizes and explores further implications for statistical practice in applied research.

2. Cross-Sectional and Between/Within-Subject Linear Models with Repeated Measures

We begin here with some notation. Consider repeated measures on n subjects denoted by $i = 1, 2, \dots, n$. The “subjects” can either be persons with longitudinal repeated measures, or, as is common in environmental epidemiology, can be cities, schools, neighborhoods, census tracts, hospitals, etc. Each subject has J_i different observations enumerated by $j = 1, \dots, J_i$. For example, these J_i different observations could be taken at times $t_{i1} < t_{i2} < \dots < t_{ij}$, on the same person when the “subject” is a person or from J_i different persons living in the same neighborhood when the “subject” is a neighborhood. For J_i constant across i , (i.e., always the same number of repeat measures for a subject), we drop the “ i ” subscript and denote J . Let us consider that the observations have continuous outcomes Y_{ij} and K predictor (or exposure) variables $X_{ij} = X_{1,ij}, X_{2,ij}, \dots, X_{K,ij}$. When $K = 1$, we drop the “ K ” enumeration, using X_{ij} for the only predictor. Linear regression models for $E[Y_{ij} | X_{ij}]$ or $E[Y_{ij} | X_{ij}]$ are fit in the analyses described here. However, the overall conclusions we obtain on these linear regression models can be generalized to discrete outcomes (i.e., logistic regression) and survival analyses.

2A Cross-Sectional (CS) Regression. The most commonly fitted linear regression model on repeated measures does not separate within- and between-subject associations and is usually written out as $Y_{ij} = \alpha + \beta_1 X_{1,ij} + \beta_2 X_{2,ij} + \dots + \beta_K X_{K,ij} + \varepsilon_{ij}$. This is denoted as “cross-sectional (CS) regression” particularly for longitudinal repeated measures. We add a subscripted “CS” to the β ’s to distinguish these slopes from between-subject (BS) and within-subject (WS) slopes defined in Section 2B. The CS regression model here is thus denoted as

$$Y_{ij} = \beta_{CS} X_{ij} = \alpha_{CS} + \beta_{1,CS} X_{1,ij} + \beta_{2,CS} X_{2,ij} + \dots + \beta_{K,CS} X_{K,ij} + \varepsilon_{ij} \quad (1)$$

where $\alpha_{CS}, \beta_{1,CS}, \beta_{2,CS}, \dots, \beta_{K,CS}$ are parameters (fixed effects), while ε_{ij} is error with $E[\varepsilon_{ij}] = 0$ that is independent between different subjects i and i' , but may be correlated for $j \neq j'$ within the same subject. It should be noted that the intercept is fixed at the same α_{CS} for each subject. Should the actual

intercepts differ between subjects (i.e., be $\alpha_{CS,i}$) as random intercepts, then for both MM and GEE, the difference $\alpha_{CS,i} - \alpha_{CS}$ is incorporated into the error term ε_{ij} of (1) and the within-subject correlation of that error [1]. Using (vs. not using) random intercepts does not influence the point estimates of α_{CS} , $\beta_{1,CS}$, $\beta_{2,CS}$, \dots , $\beta_{K,CS}$ or the variance of these estimates for mixed models [1]. However, for GEE using a different intercept on each subject (with each intercept now adding a new parameter) creates too many parameters for the asymptotic properties of GEE model to hold in our examples (and in general) which destabilizes parameter estimates [2].

Again for $K = 1$, the subscript for K is dropped and the model is $Y_{ij} = \alpha_{CS} + \beta_{CS}X_{ij} + \varepsilon_{ij}$. The main goal of CS regression is to first obtain estimates $\hat{\beta}_{CS}$ for β_{CS} and then input $\hat{\beta}_{CS}$ into (1) in order to estimate future unobserved Y 's from observed X_{ij} 's as $\hat{\beta}_{CS}X_{ij}$. Cross-sectional regression is also used to make adjusted (causal) inference on the covariate associations in $\hat{\beta}_{CS}$, but, as we show later, doing this may be problematic.

Table 1 presents parameter estimates from repeated measure cross-sectional regression (1) to a clinical measure of glomerular filtration rate (EGFR) from the Modification of Diet in Renal Disease Study (MDRD) [3]. Formula (1) with EGFR as the outcome Y and three predictor variables (X_1, X_2, X_3) = (HIV infection, serum albumin, blood urea nitrogen (BUN)) was fit to 10,782 semi-annual measures of 584 women at the Bronx-site of the Women's Interagency HIV Study (WIHS) [4]. Higher EGFR values indicate better renal function. The models assume that the within- and between-subject associations of the predictor variables are the same. We later show this assumption is incorrect. The parameter estimates of Table 1 were calculated using GEE [1] with both independence (GEE-IND) in columns 2–4 and equicorrelation (GEE-E) columns 5–7 for the working correlation structure of model residuals from repeated measures in the same person. We again note that this model (1) has a fixed intercept across all subjects with the error term being independent between different subjects. However, otherwise in Table 1 (and elsewhere in the paper) the within subject correlation structure of the error is allowed to be either (i) independent within the same subject (GEE-IND) or (ii) to have the same correlation for all outcomes within the same subject (GEE-E). The second condition (i.e., equicorrelation) is equivalent to fitting a random subject intercept model [1].

Most of today's literature providing guidance on fitting repeated measures linear regression (i.e., [5–13]) qualitatively describes working correlation as a "nuisance factor" that does not alter model parameters and states that "the working correlation that minimizes variance of parameter estimates should be chosen". However, in Table 1, the parameter estimates for BUN (per g/dL), from GEE-E, of -1.22 ; 95% confidence interval (CI) ($-1.46, -0.99$) is both qualitatively and statistically higher than the corresponding GEE-IND estimate of -1.87 ; 95% CI ($-2.12, -1.62$). For HIV, the parameter estimates of -3.86 , $p = 0.0081$ from GEE-E is qualitatively lower than that from GEE-IND -2.04 and $p = 0.19$. Clearly, changing the working correlation from independence to equicorrelation qualitatively and statistically changes the parameter estimates. Thus, this correlation structure is not a nuisance factor.

When faced with such a dilemma of qualitatively and statistically different parameter estimates from the same model fit to the same data with only the working correlation structure changed (as is shown in Table 1), investigators typically go to published guidance on which correlation structure to use. To that end, based on the within-subject correlation of residuals being 0.45 in GEE-E (and in MM-E), and the quasi-likelihood independence criteria goodness of fit statistic (QIC) = 10,836.27 for GEE-E being smaller than the QIC = 10,847.14 for GEE-IND (or the Akaike information criteria goodness of fit statistic (AIC) from a mixed model using equicorrelation (MM-E) of (AIC = 94,934.5) being smaller than AIC = 99,374.5 from a mixed model using independence (MM-IND) as shown in Table A1 in Appendix A), almost all articles providing model fitting guidance [5–13] point towards using equicorrelation as the working correlation structure. However, as the rest of Section 2 describes in detail, this guidance is problematic as only the parameter estimates obtained by using independence working correlation can have any meaning for cross-sectional regression.

But first we make two brief asides. First, we note that if MM, rather than GEE are used for Table 1, the corresponding parameter point estimates in Table 1 using independence correlation (MM-IND) and equicorrelation (MM-E) are essentially unchanged [1]. (See Appendix A for details on parameter estimates from MM fit to this data with independence and equicorrelation correlations structures). However, due to non-robustness of MM, GEE is preferable for this specific example. Second, we note that the differences observed in Table 1 occur not only between independence and equicorrelation. Any different choice of correlation structure, such as AR(1), Toeplitz, unstructured, etc. will result in different parameter estimates (results not shown). For simplicity, we focus this article on only two structures: independence and equicorrelation.

Table 1. Cross-sectional regression parameter estimates using GEE ¹ for EGFR = HIV infection, serum albumin and BUN in the Bronx WIHS.

Variable	Working Correlation Structure					
	Independence			Equicorrelation ²		
	Point Estimate	95% CI	Z-Value (p)	Point Estimate	95% CI	Z-Value (p)
HIV Infection ($\beta_{HIV,CS}$)	-2.04	(-5.07, 0.98)	-1.32 (<0.19)	-3.96	(-6.90, -1.03)	-2.65 (0.0081)
Albumin Per g/dL ($\beta_{ALB,CS}$)	-6.21	(-8.95, -3.47)	-4.44 (<0.0001)	-9.84	(-12.01, -7.68)	-8.93 (<0.0001)
BUN Per mg/dL ($\beta_{BUN,CS}$)	-1.87	(-2.12, -1.62)	-14.45 (<0.0001)	-1.22	(-1.46, -0.99)	-10.30 (<0.0001)
Quasi-Likelihood Information Criteria (QIC)	10,847.14			10,836.27		

¹ Mixed models gave essentially similar point estimates; see Appendix A. ² Interclass correlation of residuals from GEE-E was 0.45 indicating non-independence correlation was structurally correct.

2B Between-/Within-Subject Slope (BS/WS) Regression. While investigators almost never consider this in practice, it has long been noted that slopes on changes of X_{ij} within the same subject i differ from cross-sectional slopes on between subject-measure differences in X_{ij} [14–17]. To illustrate this, consider the cross-sectional model of a laboratory measure cholesterol (Y_{ij}), which is well known to be higher in people with more body fat. To that end, the predictor is body weight (X_{ij}) with $E[Y_{ij}] = \alpha_{CS} + \beta_{CS}X_{ij}$. As described in the Introduction, the cross-sectional slope β_{CS} for association of a 10 lbs. weight difference between two different adults for cholesterol is less than the slope for association of a 10 lbs. within-subject weight change for the same adult on cholesterol, which we denote as β_{WS} . Again, the reason β_{CS} is less than β_{WS} is that: (i) a 10 lb. cross-sectional weight difference between two adults often reflects greater height in one of the persons, but (ii) a 10 lbs. weight increase in the same adult is not influenced by height difference and thus is more likely due to more body fat after the 10-lbs weight gain. Thus, since greater body fat is what is directly associated with more cholesterol, the within-person association of a 10-lb. weight increase with cholesterol is greater than the cross-sectional repeated measures association with a 10-lb. weight difference between two persons.

Common within-person height creates a shared within-subject measurement bias from this extraneous factor for subject i (denoted E_i) on weight as a predictor of cholesterol. To that end, many investigators adjust weight for height using body mass index = wt/ht^2 to remove this effect of height on weight. As Figure 1a illustrates, if TX_{ij} = body mass index (wt/ht^2) were the true predictor of Y_{ij} , and H_i = height (which does not change with j in the same i), then $X_{ij} = TX_{ij} * (H_i)^2$ contains this shared within-subject measurement bias from common H_i which again we denote as E_i in Table 1a to confer it is an extraneous within-subject bias. Section 4 describes more settings where $\beta_{WS} \neq \beta_{CS}$.

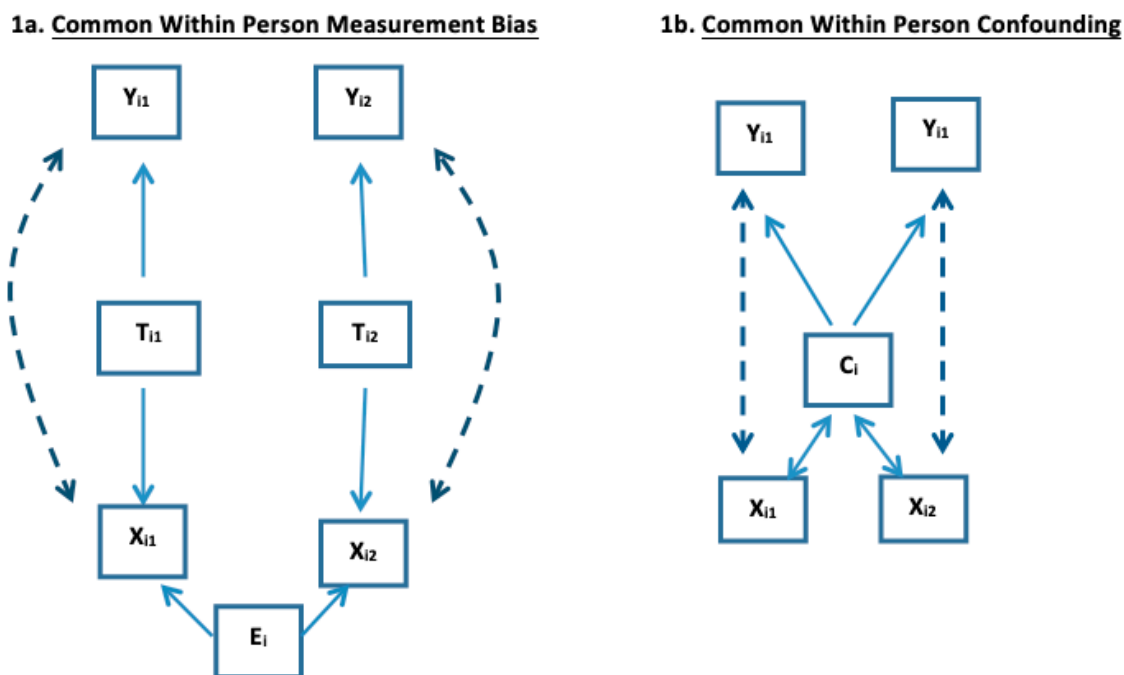


Figure 1. Illustration of common within-subject measurement bias and confounding for K = 1.

While for weight it is possible to remove the common shared within-subject bias from height by dividing by ht^2 , this is not the case for less well-understood causal relationships. Therefore, to model and account for a bias such as this, linear regression models fit for making causal inference can decompose the associations into “within-subject” slopes (β_{WS}), described above, and “between-subject” slopes (β_{BS}), described below, which capture associations of subjects’ central tendencies of the exposure. To do this, subject i means of the predictor variables $\bar{x}_i = \bar{x}_{1,i}, \bar{x}_{2,i}, \dots, \bar{x}_{K,i}$ are calculated, where $\bar{x}_{k,i} = \sum_{j=1}^{J_i} X_{k,ij} / J_i$. Then Y_{ij} is modeled as a combination of “between-subject” slopes from $\bar{x}_{k,i}$ (that could be influenced by the common person measurement bias in Figure 1) and “within-subject” slopes from deviations of $X_{k,ij}$ about $\bar{x}_{k,i}$ which will be free of such a bias, since the comparison is within person.

$$Y_{ij} = \alpha_{BS/WS} + \beta_{1,BS}\bar{x}_{1,i} + \beta_{2,BS}\bar{x}_{2,i} + \dots + \beta_{K,BS}\bar{x}_{K,i} + \beta_{1,WS}(X_{1,ij} - \bar{x}_{1,i}) + \beta_{2,WS}(X_{2,ij} - \bar{x}_{2,i}) + \dots + \beta_{K,WS}(X_{K,ij} - \bar{x}_{K,i}) + \varepsilon_{ij} \tag{2}$$

As described for (1), this is a fixed intercept model that is functionally equivalent to a random intercept model for MM. When $K = 1$, we have $Y_{ij} = \alpha_{BS/WS} + \beta_{BS}\bar{x}_i + \beta_{WS}(X_{ij} - \bar{x}_i) + \varepsilon_{ij}$. To illustrate this for our earlier example with Y_{ij} = cholesterol and X_{ij} = weight, let $\alpha_{BS/WS} = 30$, $\beta_{BS} = 0.9$ and $\beta_{WS} = 3$, such that $Y_{ij} = 30 + 0.9\bar{x}_i + 3(X_{ij} - \bar{x}_i) + \varepsilon_{ij}$. If person i had an average value of $\bar{x}_i = 210$ across all J_i measures with the j^{th} measure being $X_{ij} = 200$, then for the person-visit at time t_{ij} , $E[Y_{ij}] = 30 + 0.9(210) + 3(200-210) = 189$.

Now we make some technical asides. First, the choice of the observed $\bar{x}_{k,i}$ as the “central tendency” of $X_{k,ij}$ for subject i is necessary as $\mu_{k,i}$ a person’s “true average weight” over the entire time period is unknown, but for J_i large enough, $\bar{x}_{k,i}$ should be close to $\mu_{k,i}$. Thus, while $\beta_{k,WS}$ only captures association with within-subject change in $X_{k,ij}$, $\beta_{k,BS}$ inherently contains some $\beta_{k,WS}$ from deviation of $(\bar{x}_{k,i} - \mu_{k,i})$; especially for small J_i . This situation is described for occupational epidemiology research, where often an average of personal exposure measurements is computed as estimate of true exposure of a “subject”, defined as either an individual, or group of individuals that share a job [18]. Second, the implicit assumption that $\beta_{k,WS}$ is well defined may also not always be true. For example, “ $\beta_{k,WS}$ ” could differ by time separation $t_j - t_j'$. Perhaps for $k =$ weight, a weight gain

of 10 lbs. in one month creates a shock that hyper-elevates cholesterol, but a 10 lbs. weight gain over 12 months does not, in which case $\beta_{k,WS} \left| (t_{ij} - t_{ij'}) = 1 \right. > \beta_{k,WS} \left| (t_{ij} - t_{ij'}) = 12 \right.$. Third, if the investigator is only interested in the within-subject slopes he/she can substitute as a *fixed effect* a different subject intercept $\alpha_{WS,i}$ for the between-subject slopes in (2) with the model reducing to $Y_{ij} = \alpha_{BS,i} + \beta_{1,WS}(X_{1,ij} - \bar{x}_{1,i}) + \beta_{2,WS}(X_{2,ij} - \bar{x}_{2,i}) + \dots + \beta_{K,WS}(X_{K,ij} - \bar{x}_{K,i}) + \varepsilon_{ij}$.

Despite these technical caveats, the within- vs. between-subject decomposition in (2) is used to test whether $\beta_{k,BS} = \beta_{k,WS}$ so that, as shown in Section 2C, they also equal $\beta_{k,CS}$ and thus the separated *WS* vs. *BS* decomposition can be collapsed to (1). Due to the orthogonal decomposition of $X_{k,ij}$ about $\bar{x}_{k,i}$ this previous test for collapsing the within- vs. between-subject decomposition is a two-sample z-test of parameter estimates from fitted models comparing $\left| \hat{\beta}_{k,BS} - \hat{\beta}_{k,WS} \right| / \sqrt{\text{Var}(\hat{\beta}_{k,BS}) + \text{Var}(\hat{\beta}_{k,WS})}$ to $Z_{1-\alpha/2}$ [17]. The within- vs. between-subject decomposition is mostly used for inference on adjusted (causal) associations of the $X_{k,ij}$'s on Y_{ij} 's. It is typically not used to produce models to estimate future unknown Y_{ij} from known $X_{k,ij}$ as such estimation often only happens in settings where just one observation per subject is available, hence $X_{k,ij} \equiv \bar{x}_{k,i}$.

We refit the analyses of Table 1 to illustrate that the impact of choice of correlation structure (i.e., GEE-IND vs. GEE-E working correlation structure) is eliminated in our example after making a within- vs. between-subject decomposition. Please note that there were no new HIV infections after study entry; so $X_{HIV,ij} \equiv \bar{x}_{HIV,i}$ meaning that the within-subject association of change of HIV infection status cannot be modeled. For within-subject associations of BUN and albumin, GEE-IND and GEE-E gave identical point estimates, because centering about $\bar{x}_{k,i}$ makes comparisons entirely within-subject and invariant to these correlation structure choices (although within-subject estimates could differ slightly if autoregressive (AR (1)) or other formulations for intra-subject correlation of residuals had been used). There were small GEE-IND vs. GEE-E differences on the between-subject slopes as was observed elsewhere [19]. For example, the point estimate for between-subject HIV status is -1.16 ; 95% CI $(-4.21, 1.88)$ in the GEE-IND of Table 2 versus -1.57 ; 95% CI $(-4.47, 1.33)$ with GEE-E.

From now on, we only examine GEE-IND results for within- between-subject decomposition models, as GEE-E results are similar. For BUN and GEE-IND, the within-subject $\hat{\beta}_{BUN,WS} = -1.11$, 95% CI $(-1.34, -0.88)$ is qualitatively and statistically closer to 0 than is the corresponding between-subject slope $\hat{\beta}_{BUN,BS} = -2.72$, 95% CI $(-3.10, -2.33)$. However, serum albumin goes the other way: the within-subject slope $\hat{\beta}_{ALB,WS} = -10.70$, 95% CI $(-12.99, -8.40)$ is statistically further from 0 than is the corresponding between-subject GEE-IND $\hat{\beta}_{ALB,BS} = -3.27$ with a 95% CI $(-7.88, 1.33)$ that overlaps 0. The QIC is lower (10,857.62) for equicorrelation than for independence (10,866.64) which perhaps now indicates an advantage to the former correlation structure in this setting where the slopes have been correctly decomposed.

One might wonder how to interpret differences in the within- and between-subject slopes for causal inference, including the reasons that these slopes were different? This in part will depend on the hypotheses of interest (and we did not have any for this illustrative example). However, general rules also apply, although we are unaware of any systematic exploration of reasons why the between-subject slopes β_{BS} (or β_{BS} for $K = 1$) could differ from within-subject slopes β_{WS} (or β_{WS} for $K = 1$). and the resultant implications for causal inference. Before outlining these rules, it is important to note an important relationship among cross-sectional, within-subject and between-subject slopes.

Table 2. Within- and between-subject decomposition regression parameter estimates using GEE ¹ for EGFR = HIV infection, serum albumin and BUN in the Bronx WIHS.

Variable	Working Correlation Structure						
	Compartment	Independence			Equicorrelation		
		Point Estimate	95% CI	Z-Value (p)	Point Estimate	95% CI	Z-Value (p)
HIV Infection	Between-subject (β _{HIV,BS})	-1.16	(-4.21, 1.88)	-0.75 (0.45)	-1.57	(-4.47, 1.33)	-1.06 (0.29)
	NA ²	---	---	---	NA ²	---	---
Albumin Per g/dL	Between-subject (β _{ALB,BS})	-3.27	(-7.88, 1.33)	-1.39 (0.16)	-2.71	(-7.00, 1.57)	-1.24 (0.21)
	Within-subject (β _{ALB,WS})	-10.70	(-12.99, -8.40)	-9.16 (<0.0001)	-10.70	(-12.99, -8.40)	-9.16 (<0.0001)
BUN Per mg/dL	Between-subject (β _{BUN,BS})	-2.72	(-3.10, -2.33)	-13.89 (<0.0001)	-2.65	(-3.01, -2.08)	-14.21 (<0.0001)
	Within-subject (β _{BUN,WS})	-1.11	(-1.34, -0.88)	-9.31 (<0.0001)	-1.11	(-1.34, -0.88)	-9.31 (<0.0001)
Quasi-Likelihood Information Criteria (QIC)			10,866.64			10,857.62	

¹ Mixed models gave essentially similar point estimates. See Appendix A ² There is no within-subject variation for HIV infection status.

2C Relationship between β_{CS}, β_{WS} and β_{BS}. Now β_{CS} averages β_{WS} and β_{BS} according to relative variances of the subject means (i.e., the \bar{x}_i) vs. the variance of the repeated measures about those sample means (i.e., the $X_{ij} - \bar{x}_i$) [17]. For example, with $K = 1$, if $\sigma_{\bar{x}}^2$ is the population variance of the within-person mean \bar{x}_i and $\sigma_{X-\bar{x}}^2$ is the population variance of the deviations of differences of the repeat measures X_{ij} from their \bar{x}_i , then

$$\beta_{CS} = \beta_{BS}\sigma_{\bar{x}}^2 / (\sigma_{X-\bar{x}}^2 + \sigma_{\bar{x}}^2) + \beta_{WS}\sigma_{X-\bar{x}}^2 / (\sigma_{X-\bar{x}}^2 + \sigma_{\bar{x}}^2) \tag{3}$$

In the previous example of weight and cholesterol with $\beta_{BS} = 0.9$, $\beta_{WS} = 3$ and $Y_{ij} = 30 + 0.9\bar{x}_i + 3(X_{ij} - \bar{x}_i) + \varepsilon_{ij}$, if $\sigma_{\bar{x}}^2 = 400$ and $\sigma_{X-\bar{x}}^2 = 100$, then from (3) $\beta_{CS} = 0.9*400 / (100+400) + 3*100 / (100+400) = 1.32$. If the between-person sample means are more homogeneous in weight with $\sigma_{X-\bar{x}}^2 = 200$ but the within- person $\sigma_{\bar{x}}^2$ is still 100, then again using (3) β_{CS} moves closer to β_{WS} ; $\beta_{CS} = 0.9*200 / (100+200) + 3*100 / (100+200) = 1.60$.

2D Working Correlation Structures for Model Residuals Other than Independence Can Lead to Unusable Results for Cross-Sectional Regression. As noted earlier, fitting both MM and GEE repeated measure regression models involves specification of correlation (or working correlation) structure of ε_{ij} within the same subject i . We denote the working correlation structure by matrix V_i . Typical choices for V_i are the ones we used in the illustrative examples of Tables 1 and 2; equicorrelation (E), with correlation of ε_{ij} and $\varepsilon_{i'j'}$ for $j \neq j'$ always the same value ρ (this common value of ρ is estimated in the model fitting process based on the residuals in the model fitting process), and independence (IND), with correlation of ε_{ij} and $\varepsilon_{i'j'} \equiv 0$. However, other structures are used such as AR(1) where correlation of ε_{ij} and $\varepsilon_{i'j'}$ is $\rho^{|j-j'|}$ with the value of ρ being estimated from the residuals [1]. Again, current guidance [5–13] emphasizes choosing the V_i that most closely fits the true covariance structure of the residuals within i and/or by model fit criteria such as having lowest QIC for GEE and AIC for MM, because doing so often improves precision of the model parameter estimates. However, we just observed that this approach may be wrong for CS regression, because using any correlation structure other than IND can introduce structural bias into $\hat{\beta}_{CS}$ [20,21] and, unfortunately, AIC and QIC do not account for this bias.

To that end, Pepe and Anderson (1994) [20], developed a general rule for when IND is (and is not) the only correlation structure that should be used for CS regression that we now present. Specifically, they show that if a predictor X_{ij} varies (i.e., takes on different values) within the same subject i and,

$$E[Y|X_{ij}] \text{ depends on } X_{k,j} \text{ for any } k \text{ of a different replicate } j' \text{ in } i \tag{4}$$

then, no matter what true correlation structure of ϵ_{ij} among repeated measures within a subject is, GEE-IND gives unbiased estimates for β_{CS} , but any MM or GEE model not using $V_i = IND$, gives biased estimates of β_{CS} . Thus, the only working correlation structure that should be used to estimate β_{CS} is $V_i = IND$. However, if (4) does not hold, then any working correlation structure obtains unbiased estimates for β_{CS} in which case, choosing the V_i that most accurately fits the correlation structure of ϵ_{ij} minimizes the variance of $\hat{\beta}_{CS}$.

Our paper only focuses on equicorrelation as the alternate to independence in order to keep the presentation from becoming too cumbersome, given the large number of possible correlation structures. However, the previous paragraph and (4) apply to any non-independence correlation structure.

As one (of many) examples of where (4) holds, let $k = 1$ and Y_{ij} and X_{ij} be the degree of airway obstruction and inhalation of tobacco smoke of subject i at time j , respectively. One would expect that, because smoking effect on the lung is cumulative, historical smoking in a current smoker or non-smoker would lead to poorer lung function. Thus, $E[Y_{ij} | X_{ij'}]$ for a smoker at time $j' < j$ would be poorer irrespective of X_{ij} .

We now present an easier way to visualize (4). If repeated measures j and j' are thought of as “siblings” and the predictors as “exposures” then (4) means that even after considering the “self-exposure” of the current measure j through X_{ij} the outcome Y has “**C**onditional **D**ependence **O**n **S**ibling **E**xposures” (Co-DOSE) (i.e., on $X_{k,i,j'}$). Thus, the sibling exposure $X_{k,i,j'}$ could be thought of as a Co-DOSE beyond the “dose” from the “self-exposure”. Hence, from now on we use the term Co-DOSE to denote that (4) occurs.

Also, while this point has not been very well made, for CS regression, Co-DOSE in (4) largely occurs if and only if within- and between-subject slopes differ. If within- and between-subject slopes differ for any predictor (i.e., $\beta_{BS} \neq \beta_{WS}$) then Co-DOSE (4) happens. However, if the within- and between-subject slopes are equal for all predictors (i.e., $\beta_{BS} = \beta_{WS}$) then Co-DOSE (4) does not occur. More details on this and an illustration are given in Appendix B, but one trivial case arises if the predictors are invariant within the same subject (i.e., $X_{i1} \equiv X_{i2} \equiv \dots \equiv X_{ij_i} \equiv \bar{x}_i$) such that the within-subject slopes are not defined (since $X_{ij} - \bar{x}_i \equiv 0$) and for the same reason Co-DOSE in (4) cannot occur. While the mathematical details are beyond this paper, if $\beta_{BS} \neq \beta_{WS}$ and $V_i = IND$, then non-zero covariance $\rho_{ij} > 0$ besides adjusting for within- i collinearity of ϵ_{ij} also over-weights the β_{WS} relative to β_{BS} in (3), thereby pushing CS regression parameter estimates away from β_{CS} towards β_{WS} [17]. Since robust covariance methods exist to adjust for impact of misspecification of $V_i = IND$ from collinearity of the residuals ϵ_{ij} 's on variance estimates, in particular for GEE [1], $V_i = IND$ can eliminate bias in estimating β_{CS} while providing conservative variances for the parameter estimates.

2E Implications for Applied Research and Statistical Practice. Much of what has been presented above is not commonly understood and implemented in applied research and statistical practice. CS models are typically fit, with β_{CS} interpreted to also be β_{BS} and β_{WS} , without checking if these slopes are equal. Non-independence V_i is often used for CS regression without checking if Co-DOSE (in (4)) exists. Perhaps in part this occurs because systematic epidemiological descriptions of causal mechanisms

for why between- and within-subject slopes can differ are lacking, which hinders awareness of this possibility. We endeavor to fill this gap in Section 3.

3. Epidemiological Reasons for Between- and Within-Subject Slopes to Differ

To make it easier for investigators to identify what could cause $\beta_{k,WS} \neq \beta_{k,BS}$ (or equivalently Co-DOSE) in a given setting, we classify major reasons why this can happen. For simplicity, let $K = 1$ unless otherwise noted, as the following principles extend to multivariate settings.

3A. Change Effects. We propose that the effect of a longitudinal within-subject change in the predictor X could have a greater (or less) direct impact on Y than a long-term standing difference in X between two different subjects (hence $\beta_{WS} \neq \beta_{BS}$) and define this as a (c.f. short term) “change effect”. Returning to the example of weight and cholesterol, consider two identical twins, “A” has lived his adult life at $\bar{x}_i = 190$ lbs. and “B” at $\bar{x}_i = 180$ lbs. If “B” undergoes a short-term weight gain of 10 lbs. to 190 ($X_{i2} - \bar{x}_i = 10$), assuming \bar{x}_i not impacted by the rapid change, while A remains at 190 lbs. ($X_{ij} - \bar{x}_i = 0$), the shock or corollaries of this rapid change in B may raise his cholesterol level above that of A’s even though they both now weigh 190 lbs., meaning that $\beta_{WS} > \beta_{BS}$ and Co-DOSE in (4) occurs. However, it should be noted that as was mentioned in Section 2B, in this setting, β_{WS} would be somewhat undefined if, e.g., a 10 lbs. gain in a shorter time period (i.e., 1 month) increases β_{WS} more than does a 10 lbs. gain over a longer time period (i.e., 12 months).

3B Lag Causality of X on Future Y . The effect of historical levels of X on Y may independently project into the future (i.e., beyond that effect of the current level of X). For example, consider an HIV-infected person and two time points $t_1 < t_2$; let X_{ij} be HIV viral load and Y_{ij} be CD4 count. High HIV levels destroy CD4 blood cells into the future. Therefore, as illustrated in Figure 2a, high HIV viral load at t_1 may affect CD4 loss from t_1 to t_2 so that even if the person’s HIV viral load is low at t_2 , the high viral load at t_1 is predictive of lower CD4 at t_2 through higher viral load at t_1 having created more CD4 destruction between t_1 and t_2 (i.e., lag causality of X at t_1 on Y at t_2). Thus, $Y_{i2} | X_{i2}$ at t_2 is not independent of X_{i1} at t_1 ; Co-DOSE in (4) occurs and the within- and between- subject slopes differ ($\beta_{WS} \neq \beta_{BS}$). In Figure 2a,b, E_{i2} denotes that X_{i2} differs from X_{i1} due to an extraneous process that is causing X_i to change over time. Lag causality is often considered when serial measures of X represent long-term environmental exposures (such as air pollution and cigarette smoke) that effect chronic conditions Y (such as lung function) are obtained [1,18,22].

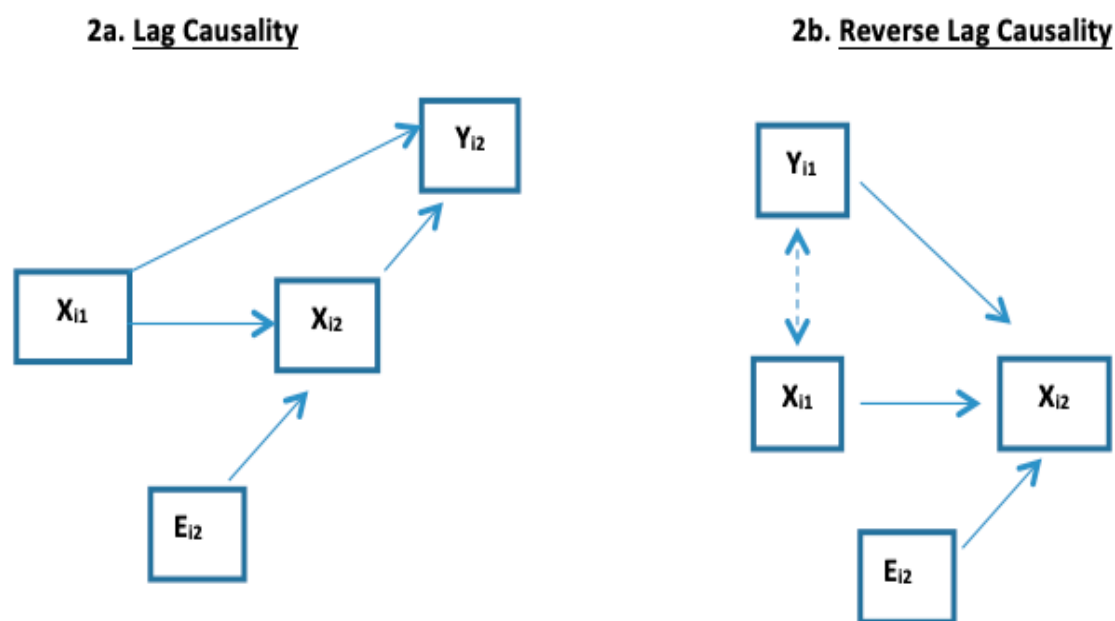


Figure 2. Illustration of lag causality and reverse-lag causality for $K = 1$.

3C Reverse-Lag Causality of X on Future Y. The setting in Section 3B also manifests in the opposite direction if X is being used as to estimate Y that is causal for future X. Reversing the previous example with X now being CD4 used to predict HIV viral load as Y, as Figure 2b illustrates, high viral load (Y_{i1}) at t_1 may have degraded the CD4 count from t_1 to t_2 . Thus, $Y_{i1} | X_{i1}$ at t_1 is not independent of X_{i2} at t_2 : Co-DOSE in (4) occurs and within- and between-subject slopes differ ($\beta_{WS} \neq \beta_{BS}$).

3D Spillover Causality of X on Adjacent Y. An analogous setting to those of 3B and 3C can also manifest in repeated measure cross-sectional settings based on geographical proximities. Let the subjects i now be cities and j enumerate different neighborhoods in these cities. The repeated measures are average air pollution (X_{ij}) of neighborhood j in city i and average lung function of all residents living within neighborhood j of city i (Y_{ij}). A resident living in neighborhood j may work in a different neighborhood j' of the same city and thus have “spillover exposure” to air in the neighborhood they work in, for a given city i , thus $Y_{ij} | X_{ij}$ is not independent of $X_{ij'}$ and hence Co-DOSE in (4) occurs.

3E Common Within-Subject Measurement Bias. Shared *within-subject measurement bias* occurs if all repeat measures from the same subject have the same correlated measurement bias. This was the setting described in Section 2B and Figure 1a with weight as exposure for cholesterol. Here with weight as a surrogate for body fat, the measurement bias was mediated by height with taller adults being heavier independently of body fat than shorter adults, which leads to $\beta_{WS} > \beta_{CS}$ and Co-DOSE in (4) when weight was a predictor of cholesterol. In this setting, height is a *measurement bias* not a confounder as height itself is not associated with cholesterol. We now present a similar setting where the un-modeled variable is a confounder.

3F Common Within-Subject Confounding. Figure 1b shows *common within-subject confounding*, that causes $\beta_{WS} \neq \beta_{CS}$ and Co-DOSE in (4). This phenomenon is diagrammatically similar to *common measurement bias* that was described in Section 2B. However, rather than a common measurement bias, the extraneous factor, shared by the repeated measures of the same subject, is a confounder that is associated with both X and Y. For example, let the confounder variable C_i be sex of subject i (which does not change with j) not be in the model and the outcome Y_{ij} be a linear score for male pattern baldness at time j with again X_{ij} being weight at time j . Adult men are both on average heavier and, independently of weight, have greater male pattern baldness than do adult women. So C_i is associated with both the exposure and the outcome. Here a 10 lbs. weight difference in two adults, but not a within-adult increase of 10 lbs., could be informative of the heavier adult more likely being male. Hence for this example, $\beta_{WS} = 0$ (assuming within adult weight does not influence baldness), but $\beta_{CS} > 0$ (and thus $\beta_{BS} > 0$) as males are more likely to be both heavier and bald compared to women. Hence also $\beta_{CS} > 0$, reflecting unaddressed between-subject confounding from heavier adults more likely being men.

Similarly, Mancl, Leroux and DeRouen proposed that in a study with repeated dental predictor and outcome pairs as (X_{ij}, Y_{ij}) measured on teeth (i.e., enumerated by j) on the same persons (i.e., enumerated by i) that better compliance with dental treatment by some persons was a confounder that could lead to differences in slopes within and between subjects [19]. In a non-longitudinal setting where i denotes clusters (for example schools) and j denotes repeated subjects within that cluster (for example students), common within-subject confounding is referred to as “contextual effects” [23,24]. For example, as Robinson (1950) [14] observed, when X was race of the student and Y was achievement-score, a higher \bar{x}_i (here: portion of a school’s students that were non-White) indicated weaker financial support for that school (weaker financial support being the confounder) and thus worse achievement-scores overall for that school: β_{BS} was negative. However, within the same school, race had no impact on the achievement score ($\beta_{WS} = 0$). Begg and Parides [25] identify a similar setting in birthweight and intelligence quotient in families.

3G Measurement Error in X_{ij} Makes $E[Y_{ij} | X_{ij}]$ Dependent on X_{ij} In many settings, the predictor we observe is $X = TX + M$ where TX is the true value of the predictor and M is measurement error that is independent of TX (i.e., classical measurement error). It has been shown that, measurement error in X that is either independent of [26], or correlated with Y [27], biases estimates for the slope that

relates TX with Y . Measurement error can arise either from imprecision in an analysis instrument, such as in a machine quantifying components of serum, or *in data collection process*, such as the chemical composition of blood samples being non-informatively influenced by diurnal and other nuisance processes. If X_{ij} is incorrectly quantified due to such measurement error, then Co-DOSE in (4) occurs and the observed within- and between-subject slopes differ, because, as illustrated in Appendix C, the biases being created from the measurement error distribute differentially to different slopes. As Figure 3 shows and the paragraph below it describes using an illustrative example, if X_{i1} incompletely measures the true state TX_{i1} (i.e., true BUN) due to classical measurement error as the extraneous influence then X_{i2} is informative for TX_{i1} even after considering X_{i1} . Please note that in Figure 3 there are two times subscripts on the extraneous influence, because E_{i1} and E_{i2} are two independent measure errors.

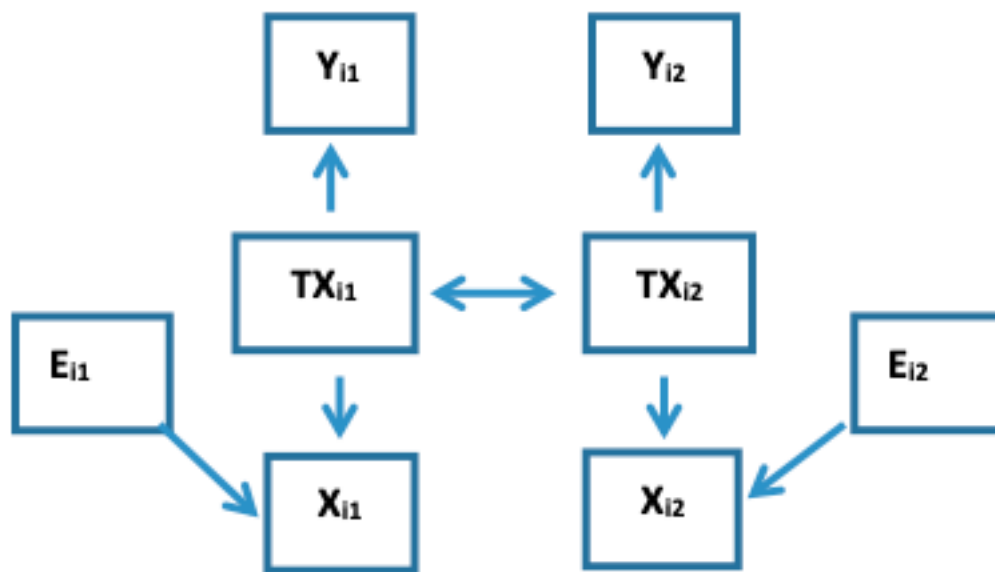


Figure 3. Illustration of residual association with independent measure error in X for K = 1.

For example, going back to the analysis of Table 1, let X_{ij} be BUN and Y_{ij} be EGFR. Consider two persons who have BUN of $X_{i1} = 10$ mg/dL measured with error today. Also assume that the true BUN state changes slowly. If so, and after 6 months one of these persons measures $X_{i2} = 20$ mg/dL while the other measures $X_{i2} = 5$ mg/dL, we can then surmise that since BUN changes slowly, it is more likely that the true BUN today (TX_{ij}) of the former person is > 10 mg/dL and that of the latter is < 10 mg/dL. Thus, since (i) EGFR (Y_{i1}) directly depends on TX_{i1} not X_{i1} , and (ii) X_{i2} is informative on TX_{i1} after considering X_{i1} , then (iii) $Y_{i1} | X_{i1}$ is not independent of X_{i2} and similarly $Y_{i2} | X_{i2}$ not independent of X_{i1} meaning Co-DOSE in (4) occurs and the observed within- between-subject slopes differ. Appendix C shows that measurement error in the exposure that is independent of the outcome pushes both β_{WS} and β_{BS} towards 0, but more so for β_{WS} . Such tempering from averaged measurement error has been proposed as a reason $|\beta_{WS}| < |\beta_{BS}|$ was observed in dental research [19] and occupational epidemiology [28,29].

However, if M_{ij} is correlated with Y_{ij} (most likely being correlated with measurement error on Y_{ij} [27]) the tempering of β 's from M_{ij} will not be to 0. For example, consider $TX = CD8$ and $TY = CD4$ cells which together are the almost exclusive components of serum lymphocytes (TZ) (i.e., $TY \approx TZ - TX$). Physiologically, TZ is constrained to create a negative β_{BS} , β_{WS} and β_{CS} for TY_{ij} on TX_{ij} : subjects with a higher CD8 component of serum lymphocytes by converse must have lower CD4 components. However, the measured lymphocyte count (Z) is subject to a correlated measurement error that equally spreads onto X and Y . For example, if a person is dehydrated, the entire measured lymphocyte (meaning both $CD8 = X$ and $CD4 = Y$) portion of blood becomes artificially higher due to reduction of the percentage of water in the blood. If a person has a high (or low) measured lymphocyte

count $Z_{ij} = TZ_{ij} + M_{ij}$ due to such measurement error, then M_{ij} contributes to both CD4 (X_{ij}) and CD8 (Y_{ij}), making both simultaneously artificially higher (or lower). Consequently, within person, a higher measured CD4 count due to positive M_{ij} is associated with higher measured CD8. Because in this case the measurement error is shared, naïve regression analysis tends to draw β_{WS} towards being positive. On the other hand, β_{BS} , which tempers down M_{ij} on both X and Y through averaging as shown in Appendix C, is less affected by the shared bias due to measurement error.

We have only considered classical measurement error so far. The other common type of measurement error is known as Berkson error [30]. It is approximated by some exposure assessment procedures commonly used in environmental and occupational epidemiology (see semi-ecological design and group-based exposure assessment) [18]. While this is an aside to the main points of this paper, when Berkson measurement error exists, only the between-subject slope, β_{BS} , is estimable. More details are in Appendix D.

4. Predictors Having Co-DOSE Will Bias Adjusted Parameter Estimates of Other Predictors Not Having Co-DOSE When Included Together in Cross-Sectional Regression When $V_i \neq \text{IND}$ Is Used

Going back to Table 1, it was shown earlier that the point estimate from GEE-IND $\hat{\beta}_{HIV,CS}$ for the adjusted cross-sectional association of HIV with EGFR is still consistent for $\beta_{HIV,CS}$. However, HIV infection status was constant over all replicates within the same subject, and therefore cannot have Co-DOSE in (4) as the entire effect of HIV is mediated between-subject, not within-subject. Consequently, the question arises whether the adjusted estimate from a non-independence correlation structure (say for example $\hat{\beta}_{HIV,CS-E}$) can be biased for $\beta_{HIV,CS}$. Please note that for this section, we use $\hat{\beta}_{XXX,CS}$ and $\hat{\beta}_{XXX,CS-E}$ to denote estimates for adjusted cross-sectional association for variable XXX from models using independence and equicorrelation structures, respectively. The added designation of “E” (CS-E) in the subscript for equicorrelation, but none for independence correlation, is made because the equicorrelation estimate (but not the independence estimate) can be asymptotically biased. The specific question addressed here is: could including BUN and albumin that each have Co-DOSE in the model bias the corresponding estimate for cross-sectional adjusted HIV association from using equicorrelation ($\hat{\beta}_{HIV,CS-E}$) so that it no longer is consistent for $\beta_{HIV,CS}$ in the multivariate model, even though HIV itself is not Co-DOSE? This is important, because in Table 1, $\hat{\beta}_{HIV,CS}$ of -2.04 95% CI (-5.07 0.98) qualitatively differs from $\hat{\beta}_{HIV,CS-E}$ of -3.96 (-6.90 , -1.03) with only $\hat{\beta}_{HIV,CS-E}$ statistically ($p < 0.01$) differing from 0.

We believe that $\hat{\beta}_{HIV,CS-E}$ for HIV in Table 1 is biased away from $\beta_{HIV,CS}$. To help make this point, Table 3 presents normative data broken down by HIV status of the subjects. First we note from Table 1 that $\hat{\beta}_{BUN,CS-E}$ is biased higher for $\beta_{BUN,CS}$ (with GEE-E $\hat{\beta}_{BUN,CS-E} = -1.22 > \hat{\beta}_{BUN,CS} = -1.87$, $p < 0.0001$ from GEE-IND), while from Table 3, those who are HIV+ have higher mean BUN (12.94 vs. 12.10, $p < 0.0001$ from GEE-IND). Thus, the full apparent “negative effect” of the higher BUN in HIV+ subjects from $\beta_{BUN,CS}$ is underestimated by $\hat{\beta}_{BUN,CS-E}$ and this pushes $\hat{\beta}_{HIV,CS-E}$ down to compensate. Second, similarly, also from Table 1, $\hat{\beta}_{ALB,CS-E}$ is biased lower for $\beta_{ALB,CS}$ (with GEE-E $\hat{\beta}_{ALB,CS-E} = -9.84 < \hat{\beta}_{ALB,CS} = -6.21$), while from Table 3, HIV+ individuals have lower mean albumin (3.97 vs. 4.14, $p < 0.0001$ from GEE-IND). Thus, the apparent “positive effect” of the lower albumin in HIV+ subjects from $\beta_{ALB,CS}$ is overestimated by $\hat{\beta}_{ALB,CS-E}$, which pushes $\hat{\beta}_{HIV,CS-E}$ further down to compensate. Now we consider these two biases together as illustrated in Figure 4. These two deficits act jointly to push $\hat{\beta}_{HIV,CS-E}$ downwards from the true adjusted $\beta_{HIV,CS}$. Therefore, non-independence V_i can bias multivariate cross-sectional parameter estimates of variables that do not carry Co-DOSE in (4) when other variables in the model carry Co-DOSE.

Table 3. Means ± standard deviation of EGFR serum albumin and BUN broken down by HIV status across all repeated measures used in Tables 1 and 2.

Variable	For HIV + Subjects (496 persons 7326 Replicates)	For HIV - Subjects (178 persons 3456 Replicates)
EGFR	90.3 ± 27.2	92.4 ± 25.0
BUN	12.94 ± 5.71	12.10 ± 5.30
Albumin	3.97 ± 0.44	4.14 ± 0.36

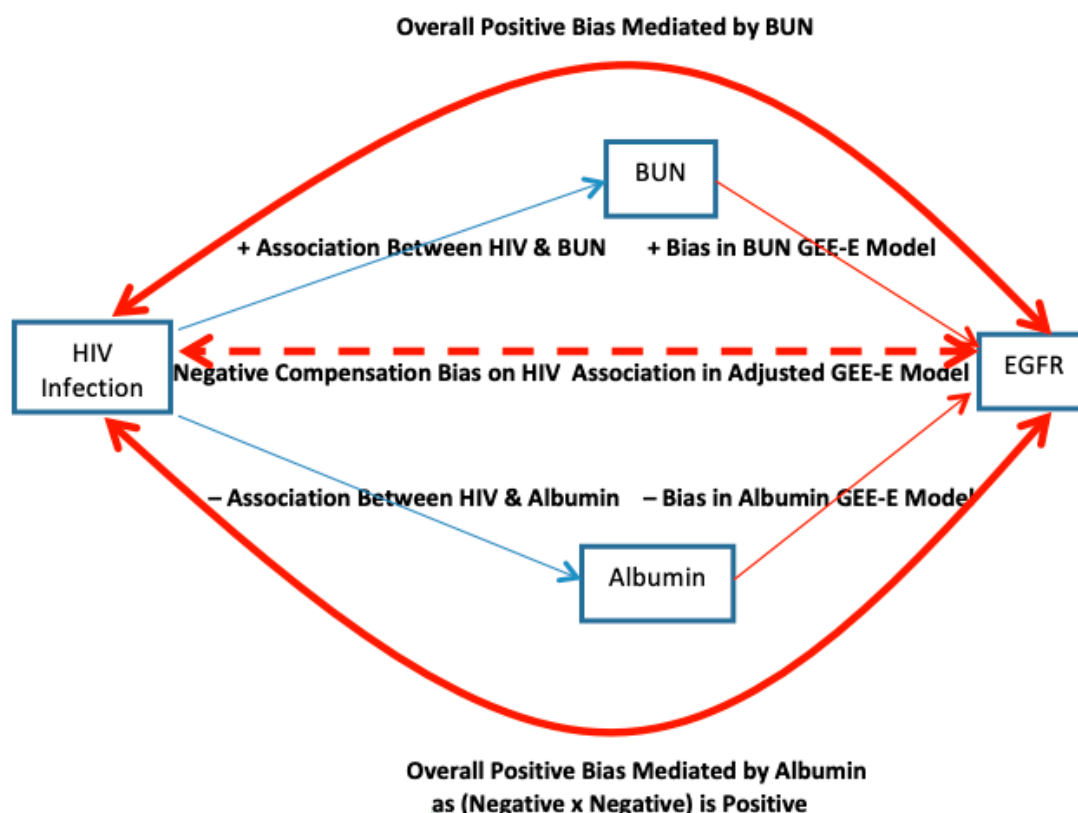


Figure 4. Compensating bias pathways on time invariant HIV estimate from failure to use an independent working correlation structure in repeated measures GEE.

5. Discussion

Numerous published papers fit GEE and MM cross-sectional regression models with repeated measures having time varying predictors that either use non-independence working correlations structures or do not state the correlation structure. These papers, which continue to be published, do not show awareness of the points presented in Sections 1–4, above. Specifically, they:

- (a) Neither specify whether the coefficients of interest are β_{CS} , β_{WS} or, β_{BS} nor check whether $\beta_{WS} = \beta_{BS}$;
- (b) Make potentially invalid interpretations of β_{CS} from MM and GEE using non-independence correlation V_i 's; and/or;
- (c) Do not justify the choice of non-independence working correlation structures V_i in light of potential differences between β_{WS} , β_{BS} and β_{CS} .

We have identified almost 45 such papers including some authored by us prior to becoming aware of these issues. This is almost certainly only a fraction of the total number of such papers.

Yet papers published up to 65 years ago either warn against using non-independence working correlation structure in cross-sectional regression with repeated measures [19,20], or instruct to decompose the associations into within-subject (β_{WS}) and between-subject (β_{BS}) slopes to make causal inference [14–17]. Numerous examples where $\beta_{WS} \neq \beta_{BS} \neq \beta_{CS}$ have been presented [14–20,22–25]. While it was not covered in our paper, this includes fitting GEE models of binary outcomes where the issues discussed here also apply [19,31]. However, these points are still not well known or emphasized in statistical software documentation and papers providing guidance on GEE and MM analyses (i.e., [5–13]).

One problem that impedes acceptance of within- and between-subject decomposition is that it necessitates much more complicated models that are difficult to explain. Still, some air pollution epidemiologic studies have attempted within- and between-subject decompositions using cities as the subject and neighborhoods as the repeated measures within the city [32–34]. Most often in these studies, the magnitude was greater for within-subject slope $|\beta_{WS}| > |\beta_{BS}|$ but sometimes $|\beta_{BS}| > |\beta_{WS}|$ was observed meaning that possibly multiple causes for slope differences are involved. Those papers that did attempt to explain the reasons for the differences described only “common within-subject confounding” (Section 3E) as a potential reason; such as un-modeled pollutants that were correlated between (but not within) cities with the modeled pollutants of interest. Other studies in environmental research have considered the mechanism described in Section 3B, namely, lag causality in longitudinal analyses of association of air pollution on health measures [1]. Nevertheless, having to explain complicated and unknown mechanisms for biases such as these can appear to detract from the main purpose of the research and cast doubt on the overall findings, making the paper harder to publish. In other words, there appears to be neither incentive, nor guidance on how to engage with these issues for applied researchers.

We concur with others [19,20], that cross-sectional regression with repeated measures should use independence as the default working correlation unless justification is given to use other V_i . While non-independence V_i can improve precision and thus be desirable [21], they can considerably bias estimates for cross-sectional parameters, β_{CS} , including perhaps towards what the investigator wants to see. For example, in Table 1, $p < 0.01$ was observed for association with HIV with worse EGFR in GEE-E compared to the more appropriate $p = 0.19$ from GEE-IND. An investigator who was expecting HIV to be associated with worse EGFR might thus be tempted to use the results from GEE-E for this reason.

While showing this is beyond the scope of our paper, when V_i is not independence, factors such as the values of J_i and magnitude/structure of ε_{ij} strongly influence parameter estimate values for β_{CS} from the miss-fitted cross-sectional models, allowing the miss-fitted estimate to arbitrarily range from β_{CS} to β_{WS} [17]. Standardization is important and, as such factors will arbitrarily vary between studies, parameter estimates of β_{CS} become harder to compare across studies when V_i differs at discretion of investigators. Therefore, the working correlation structure used in cross-sectional regressions using repeated measures should always be justified and reported.

We also concur with others [14–19,23–25] that despite the difficulties in identifying why within- and between-subject slopes differ, causal inference analyses with repeated measures should initially make such decompositions. Investigators should then be wary if there are qualitative differences between β_{WS} and β_{BS} . For example, Table 2 with 584 subjects and 10,782 measurements demonstrated need for β_{WS} , β_{BS} decomposition to make causal inference (as well as for using GEE-IND in cross-sectional regression). However, a smaller study could have been less clear-cut. If the same point estimates for β_{WS} and β_{BS} seen in Table 2 were observed but did not statistically differ, one would be tempted to merge β_{WS} and β_{BS} into a combined β_{CS} at least for some variables, because standard model-fitting practice promotes parsimony when statistical significance is not observed. This would

be particularly true if for a given variable, k , neither $\hat{\beta}_{k,WS}$ nor $\hat{\beta}_{k,BS}$ statistically differed from 0, but $\hat{\beta}_{k,CS}$ did. If such collapsing is done, it may still be important to report $\hat{\beta}_{k,WS}$ and $\hat{\beta}_{k,BS}$ for comparison to future studies and target potential mechanisms for within- between-subject slope differences as described in Section 3.

Unfortunately, the within- and between-subject slope decomposition expands required analyses and presentation. Statistical software mostly does not have standard subroutines to do this. Decomposition can be tedious if $\bar{x}_{k,i}$ is recalculated to maintain orthogonal decomposition of $X_{k,ij}$ as new models are fit if observations are excluded from the J_i due to missing values of newly included variables. The fact that the $\bar{x}_{k,i}$ are ill-defined by averaging the $X_{k,ij}$ rather than being true means for subject i creates confusion about interpretation of $\hat{\beta}_{BS}$ that can also be influenced by within-subject slopes as was noted in Section 2B.

When $\hat{\beta}_{BS}$ and $\hat{\beta}_{WS}$ differ, the causal mechanisms as to why this happens should be explored. For example, in our analysis presented in Table 2 with EGFR as the outcome, for BUN the between-subject slope $\hat{\beta}_{BUN,BS} = -2.72$ (from GEE-IND) was statistically further from 0 in the expected direction of association than was the within-subject slope $\hat{\beta}_{BUN,WS} = -1.11$. However, the albumin went the other way: between-subject slope $\hat{\beta}_{ALB,BS} = -3.27$ was statistically closer to 0 than was within-subject slope $\hat{\beta}_{ALB,WS} = -10.70$ with again both slopes being in the expected direction from zero. So what are the potential reasons for this? While lag/reverse-lag causality (Section 3B,C) between BUN and creatinine (the main component of calculated EGFR) could reduce magnitude of $\beta_{BUN,WS}$ vs. $\beta_{BUN,BS}$, this was unlikely given the separation of visits was 6 months and internal biochemistry operates over shorter time periods. However, independent measurement error on BUN (Section 3G) would temper $|\beta_{BUN,WS}|$ towards 0 relative to $|\beta_{BUN,BS}|$. To that end, several articles find greater coefficient of variation [35,36], within-person change [35,36], assay error [36], and sample degradation for BUN vs. albumin measures [37], all of which could reflect BUN having larger independent measurement error than does albumin that would selectively attenuate $\hat{\beta}_{BUN,WS}$ towards 0 (i.e., more than it did to $\hat{\beta}_{ALB,WS}$). Conversely, serum creatinine and albumin are both constrained into the intravascular fluid compartment and will non-informatively increase together with greater hydration and decrease with less hydration of this compartment, inducing positively correlated measurement error, as in the case for measured CD4 and CD8 cells in the last paragraph of Section 3G. As creatinine factors inversely into the EGFR calculation, this would constitute negative correlation of measurement error between albumin and EGFR and selectively bias $\hat{\beta}_{ALB,WS}$ to be more negative than $\hat{\beta}_{ALB,BS}$. However, BUN, which crosses across all body compartments, is less subject to such correlation in measurement error with creatinine and thus with EGFR.

As is illustrated in the previous paragraph, we believe that the systematic epidemiological description of reasons for within- and between subject slopes to differ in Section 3 will provide some basis for future studies to explore this. That may lead to greater recognition and understanding of this phenomenon. However, our list of reasons for these slopes to differ may not be exhaustive. Furthermore, these mechanisms are quite complicated including that limited resources may be available to investigate them in given studies given the other tasks that need to be done and limited funding/personnel.

When between- and within-subject slopes differ, $\beta_{BS} \neq \beta_{WS}$, it is unclear which is the “least confounded or biased”, including the possibility that by “averaging” the different biases in each would make β_{CS} be the least biased. There may be a heuristic perception that by “matching within the same subject”, β_{WS} is superior to β_{BS} and β_{CS} , but this is not necessarily true as measurement error in X (Section 3G) and lag/reverse-lag and spillover causality (Section 3B–D) can in fact bias β_{WS} to a larger degree than they do for β_{BS} and β_{CS} .

6. Conclusions

It has been known for decades by some that when exposures vary within subjects in repeated measures regression then, (i) cross-sectional regression using $V_i =$ independence working correlation should be the default for building a model to estimate a future unknown Y as the goal, and (ii) within- and between-subject decompositions of slopes should at least initially be fit when building models for causal inference. Yet this advice rarely makes it into published guidelines and hence is not heeded, perhaps in part due to complexity of the settings where within- and between-subject slopes differ and limited substantive study of the mechanisms that cause such differences. In general, analysts should explore and quantify reasons for biases that can occur in such study designs. To that end, analyses using repeated measures regression should investigate if within- and between-subject slopes differ and when they do, try to identify the reasons for this.

Author Contributions: Conceptualization, D.R.H.; Methodology, D.R.H., Q.S. and I.B.; Software, D.R.H. and Q.S.; Validation, D.R.H., Q.S., and I.B.; Formal Analysis, D.R.H. and Q.S.; Investigation, D.R.H. and I.B.; Resources, K.A.; Data Curation, K.A. and Q.S.; Writing—Original Draft Preparation, D.R.H.; Writing—Review & Editing, D.R.H., and I.B.; Funding Acquisition, K.A.

Funding: This research was funded by the Women’s Interagency HIV Study (WIHS) Collaborative Study Group at New York City/Bronx Consortium, which was funded by the National Institute of Allergy and Infectious Diseases U01-AI-35004.

Acknowledgments: Data in this manuscript were collected by the Women’s Interagency HIV Study (WIHS) Collaborative Study Group at New York City/Bronx Consortium, which was funded by the National Institute of Allergy and Infectious Diseases U01-AI-35004. We are indebted to the participants of this study, many of whom have now devoted over 15 years of their life to this effect.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

Abbreviations

AIC	Akaike Information Criteria
AR(1)	Autoregressive Order 1
BS	Between-Subject
BUN	Blood Urine Nitrogen
Co-DOSE	<u>C</u> onditionally <u>D</u> ependent <u>O</u> n <u>S</u> ibling <u>E</u> xposure
CS	Cross-sectional
E	Equicorrelation
EGFR	Estimated Glomerular Filtration Rate
GEE	Generalize Estimation Equations
IND	Independent
MM	Mixed Models
QIC	Quasi-likelihood Information criteria
WIHS	Women’s Interagency HIV Study
WS	Within-Subject

Appendix A. Results of Our Example Obtained Using Mixed Models

While mixed models are not appropriate for the cross-sectional regression of this example (and often are not appropriate for cross-sectional regression using repeated measures in general), they are often used for this purpose in practice. We have made the case that the biases described in this paper applying to GEE for CS regression also apply to mixed models CS regression. Thus, Table A1 below presents the parameters for CS regression of the Bronx WIHS example in Table 1 as estimated by mixed models using independence and equicorrelation working correlation structures. To that end, the reader can confirm that the parameter estimates from the mixed models in Table A1 are almost identical to those from the GEE model with the same correlation structure in Table 1. This includes that the estimates obtained using independence correlation here are also qualitatively different than those

from using equicorrelation. We caution the reader; however, that the confidence intervals and p-values reported in these tables are meaningless irrespective of the biases reported on in this paper, since, unlike GEE, mixed models are not robust to misspecification of the correlation structure. Misspecified correlation structure for this example is clearly the case for independence although such a claim is more debatable for equicorrelation.

Table A1. Parameter estimates for the cross-sectional regression of Table 1 EGFR = HIV infection, serum albumin and BUN in the Bronx WIHS using mixed models.

Variable	Working Correlation Structure					
	Independence			Equicorrelation		
	Point Estimate	95% CI ¹	Z-Value (p) ¹	Point Estimate	95% CI ¹	Z-Value (p) ¹
HIV Infection ($\beta_{HIV,CS}$)	-2.04	(-5.04, -1.04)	-4.02 (<0.0001)	-3.99	(-7.04 -0.93)	-2.55 (0.01)
Albumin Per g/dL ($\beta_{ALB,CS}$)	-6.21	(-7.30 -5.11)	-11.04 (<0.0001)	-9.89	(-11.03, -8.73)	-16.90 (<0.0001)
BUN Per mg/dL ($\beta_{BUN,CS}$)	-1.87	(-1.95, -1.79)	-44.68 (<0.0001)	-1.22	(-1.30, -1.13)	-29.08 (<0.0001)
Akaike Information Criteria (AIC)	99,374.5			94,934.5		

¹ The confidence interval and p-values for independence working correlation structure in particular but also arguably for equicorrelation as well overestimate the precision of the parameter estimates. Unlike GEE, mixed models are not robust to misspecification of the correlation structure.

Table A2. Parameter estimates for the within- between-subject decomposition regression of Table 2 EGFR = HIV infection, albumin and BUN in the Bronx WIHS using mixed models.

Variable	Compartment	Working Correlation Structure					
		Independence			Equicorrelation		
		Point Estimate	95% CI ¹	Z-Value (p) ¹	Point Estimate	95% CI ¹	Z-Value (p) ¹
HIV Infection	Between-subject ($\beta_{HIV,BS}$)	-1.16	(-4.21, 1.88)	-2.30 (0.02)	-1.57	(-4.47, 1.32)	-1.06 (0.29)
	NA ²	---	---	---	NA ²	---	---
Albumin Per g/dL	Between-subject ($\beta_{ALB,BS}$)	-3.28	(-4.78, -1.77)	-4.26 (0.16)	-2.72	(-7.00, 1.57)	-1.32 (0.19)
	Within-subject ($\beta_{ALB,WS}$)	-10.70	(-12.24, -9.15)	13.60 (<0.0001)	-10.67	(-11.86, -9.48)	-17.57 (<0.0001)
BUN Per mg/dL	Between-subject ($\beta_{BUN,BS}$)	-2.72	(-2.84, -2.60)	-44.87 (<0.0001)	-2.65	(-2.95, -2.35)	-17.10 (<0.0001)
	Within-subject ($\beta_{BUN,WS}$)	-1.11	(-1.22, -0.99)	-18.59 (<0.0001)	-1.11	(-1.20, -1.03)	-25.72 (<0.0001)
Akaike Information Criteria (AIC)		98,451.8			94,824.7		

¹ The confidence interval and p-values for independence working correlation structure in particular but also arguably for equicorrelation overestimate the precision of the parameter estimates. Unlike GEE, mixed models are not robust to misspecification of the correlation structure. ² There is no Within-subject Variation for HIV Infection Status.

Mixed models may be more appropriate for within- between- subject decomposition models than they are for CS regression provided the correct correlation structure of the residuals is used. Table A2 below thus presents the parameters for within- between- subject decomposition regression of the Bronx WIHS example in Table 2 as estimated by mixed models using independence and equicorrelation. As was the case with Table A1 above compared to Table 1, the reader can again confirm here that the parameter estimates from the mixed models in Table A2 are almost identical to those from the GEE model with the same correlation structure in Table 2. This also includes that the mixed model

parameter estimates under independence and equicorrelation are at worst qualitatively similar and often close to identical for the two correlation structures. We caution the reader; however, that, as with Table A1, the confidence intervals and p-values in Table A2 should be interpreted cautiously, because mixed models are not robust to misspecification of correlation structure of the residuals. While independence correlation is clearly not correct (as the within subject residuals for this example had a large positive correlation) it can be argued that equicorrelation might be correct. However, looking into that is beyond the scope of this paper.

Appendix B. Homology between Co-DOSE in (4) Occurring with between- and within-Subject Slopes Being the Same or Differing

Figure A1 illustrates using the example of Section 2B (with $K = 1$) that Co-DOSE in (4) occurs if $\beta_{WS} \neq \beta_{BS}$. Remember that in this example, $\beta_{BS} = 0.9$, $\beta_{WS} = 3$, $\beta_{CS} = 1.60$. Now let $J = 2$. So for the between / within-subject decomposition model; $Y_{ij} = 30 + 0.9\bar{x}_i + 3(X_{ij} - \bar{x}_i) + \varepsilon_{ij}$. If the overall mean of X_{ij} for all repeat measures in the sample was 180 (i.e., $\sum_{i=1}^n \sum_{j=1}^2 X_{ij} / (2n) = 180$) then the full cross-sectional model is $E[Y_{ij}] = -96 + 160(X_{ij})$. If a subject's two weight measures are $X_{i1} = 200$ and $X_{i2} = 220$, then for the first measure, the cross-sectional model estimates $E[Y_{ij}] = -96 + 160(200) = 224$. However, since $X_{i2} = 220$ and $\bar{x}_i = 210$, as we saw earlier within- between-subject decomposition gives; $E[Y_{ij}|X_{i1}, X_{i2}] = 30 + 0.9(210) + 3(200 - 210) = 189$. Thus, $E[Y_{i1} | X_{i1}]$ is not independent of X_{i2} since X_{i2} is informative of where \bar{x}_i falls and the slope for $(X_{ij} - \bar{x}_i)$ is different than the slope for \bar{x}_i when $\beta_{WS} \neq \beta_{BS}$. However, if $\beta_{WS} = \beta_{BS} = \beta_{CS}$, then X_{i2} is non-informative on $Y_{i1} | X_{i1}$ as $E[Y_{ij}] = \alpha_{CS} + \beta_{CS}X_{i1} = \alpha_{CS} + \beta_{CS}(X_{i1} - \bar{x}_i) + \beta_{CS}\bar{x}_i = \alpha_{WS/BS} + \beta_{WS}(X_{i1} - \bar{x}_i) + \beta_{BS}\bar{x}_i$ since $\beta_{WS} = \beta_{BS} = \beta_{CS}$.

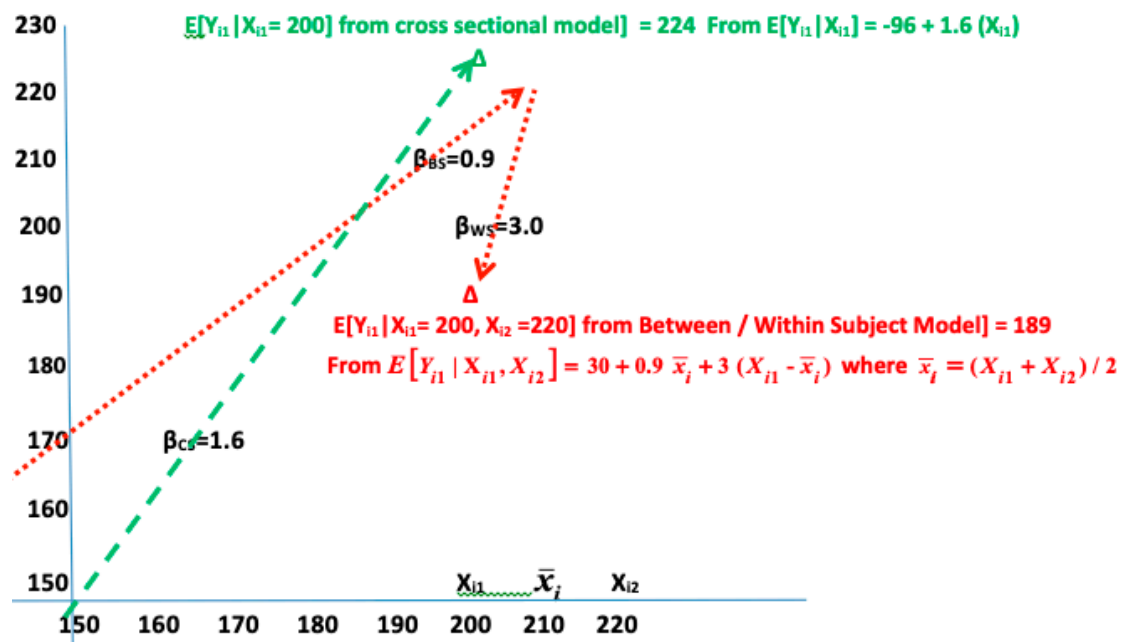


Figure A1. Illustration that $E\{Y_{i1} | X_{i1}\}$ is not independent of X_{i2} in Cross-sectional regression for $K = 1$ when $\beta_{BS} \neq \beta_{WS}$ in a between/within decomposition model with $J_i = 2$, $X_{i1} = 200$ and $X_{i2} = 220$.

As $J_i \equiv 2$, in the prior example, the second observation was deterministic for \bar{x}_i . However, for $J_i > 2$, while additional X_{ij} go into computation of \bar{x}_i these are still informative on relative contributions of $\beta_{BS}\bar{x}_i$ and $\beta_{WS}(X_{ij} - \bar{x}_i)$ on $E[Y_{ij} | X_{ij}]$.

Whether or not Co-DOSE in (4) occurs also informs if $\beta_{BS} = \beta_{WS}$. If for a given j , $E[Y_{ij} | X_{ij}]$ is independent of all other $X_{ij'}$, then $E[Y_{ij} | X_{ij}]$ is independent of $\bar{x}_i = \sum_{j=1}^{J_i} X_{ij} / J_i$, which only happens if

$\beta_{BS} = \beta_{WS}$. However, if $E[Y_{ij} | X_{ij}]$ is not independent of other $X_{ij'}$, then; (i) if $\text{Corr}[(Y_{ij}, \bar{x}_i) | X_{ij}] \neq 0$, β_{WS} (if well defined) $\neq \beta_{BS}$, (ii) otherwise if $\text{Corr}[(Y_{ij}, \bar{x}_i) | X_{ij}] = 0$ then β_{WS} is not well defined.

Appendix C. Illustration That Classical Measurement Error Which Is Independent of the Outcome Pushes β_{WS} and β_{BS} to Zero with Greater Impact on β_{WS}

To illustrate this for the classical measurement error setting with $K = 1$, let there always be the same number of replicates, J , per subject and assume that the true data-generating mechanism, i.e., in the absence of measurement error, involves $\beta = \beta_{WS} = \beta_{BS} = \beta_{CS}$. For example, let $E[Y_{ij}] = \beta TX_{ij} = \beta(TX_{ij} - \mu_i) + \beta\mu_i$, where μ_i is true mean exposure of the i^{th} subject, for simplicity the intercept is 0. However, we only observe $X_{ij} = TX_{ij} + M_{ij}$, where M_{ij} is measurement error with $E[M_{ij}] = 0$, variance σ_M^2 that is independent across all i 's and j 's and also independent from Y_{ij} . It also would often be assumed that $M_{ij} \sim N(0, \sigma_M^2)$, but we do not invoke this assumption here. When we use X_{ij} instead of TX_{ij} in regression, the observed estimates of β_{WS} , β_{BS} , and β_{CS} will not be equal to their true values (i.e., as obtained with TX_{ij}), but instead will equal different values β_{WS}^* , β_{BS}^* , and β_{CS}^* from X_{ij} being watered down by the independent measurement error as shown below. In this special case that $\beta = \beta_{WS} = \beta_{BS} = \beta_{CS}$, we show that we expect $\beta_{WS}^* \neq \beta_{BS}^* \neq \beta_{CS}^*$. We also reproduce a known result that under classical measurement error observed β^* 's are attenuated towards 0 with respect to the true β 's. Furthermore, let TX_{ij} vary with j within i as follows; $TX_{ij} = TC_i + TR_{ij}$ where TC_i is a central tendency of TX for subject i , while TR_{ij} is within- subject i repeated visit variation in TX_{ij} across the j 's. Let σ_C^2 and σ_R^2 be variances of TC_i and TR_{ij} , respectively. Now, using the identity that the slope of the regression line for $Y = \alpha + \beta X$ is the covariance of X and Y divided by the variance of X (i.e., $\text{Cov}(X, Y) / \text{Var}(X)$), we derive:

$\text{Var}(X_{ij}) = (\sigma_C^2 + \sigma_R^2 + \sigma_M^2)$, $\text{Cov}(X_{ij}, Y_{ij}) = \beta_{CS}(\sigma_C^2 + \sigma_R^2)$, so that

$$\beta_{CS}^* = \beta_{CS}(\sigma_C^2 + \sigma_R^2) / (\sigma_C^2 + \sigma_R^2 + \sigma_M^2);$$

$\text{Var}(\bar{x}_i) = (\sigma_C^2 + \sigma_R^2/J + \sigma_M^2/J)$, $\text{Cov}(\bar{x}_i, Y_{ij}) = (\beta_{BS} \sigma_C^2 + \beta_{WS} \sigma_R^2/J)$, so that

$$\beta_{BS}^* = (\beta_{BS} \sigma_C^2 + \beta_{WS} \sigma_R^2/J) / (\sigma_C^2 + \sigma_R^2/J + \sigma_M^2/J);$$

$\text{Var}(X_{ij} - \bar{x}_i) = ((J - 1)/J)(\sigma_R^2 + \sigma_M^2)$, $\text{Cov}(X_{ij} - \bar{x}_i, Y_{ij} - \bar{y}_i) = \beta_{WS} \sigma_R^2 ((J - 1)/J)$, so that

$$\beta_{WS}^* = \beta_{WS} \sigma_R^2 / (\sigma_R^2 + \sigma_M^2).$$

Thus, for example if $\beta_{WS} = \beta_{BS} = \beta_{CS} = 5$ we therefore have from the above formulas

$$\beta_{CS}^* = \frac{5(\sigma_C^2 + \sigma_R^2)}{\sigma_C^2 + \sigma_R^2 + \sigma_M^2}, \beta_{BS}^* = \frac{5\sigma_C^2 + 5\sigma_R^2/J}{\sigma_C^2 + \sigma_R^2/J + \sigma_M^2/J}, \text{ and } \beta_{WS}^* = \frac{5\sigma_R^2}{\sigma_R^2 + \sigma_M^2}.$$

Continuing with this numeric example, let $\sigma_C^2 = 8$, $\sigma_R^2 = 2$ and $\sigma_M^2 = 10$ and $J = 5$. The entire variance of X_{ij} is 20 of which half, 10, is from measurement error, 8 is variation of the central tendency of X between-subjects and 2 is variation of X within-subject. Then $\beta_{CS}^* = 5(10)/20 = 2.5$, $\beta_{BS}^* = 5(8.4)/10.4 = 4.03$ and $\beta_{WS}^* = 5(2)/12 = 0.83$. Considering that, without measurement error, true between and within person slopes are both 5, measurement error has greatly attenuated $\beta_{WS}^* = 0.83$ towards 0, while $\beta_{BS}^* = 4.03$ is the least tempered. This happens because β_{BS}^* most fully retains the common signal in X , but tempers M through averaging, while β_{WS}^* more fully retains M while excluding the between-subject signal in X .

Appendix D. Only β_{BS} Is Estimable When Under Berkson-type Measurement Error

Since we have brought up classical measurement error we should also discuss the other common type of measurement error known as Berkson error. For Berkson error the measurement error is independent of the observed value (i.e., X), but is not independent from the true value (TX). Such a situation is approximated when a common value is reported for all J_i replicates in the same subject i . For example consider a study of radiation contamination of the milk supply with $i = \text{community}$

and $j =$ child within the community. Now the true value of daily milk consumption for each child, TX_{ij} , is unknown, but the average daily consumption of milk across all children in that community, μ_i , is known (or estimated with high degree of certainty as \bar{x}_i , with a caveat noted below) and is thus substituted for X_{ij} . With Berkson-type error, the common within subject mean, μ (or \bar{x}_i), rather than the different X_{ij} is observed across all j replicates. Thus, when Berkson-type error exists, the fitted model estimates β_{BS} by default and both β_{WS} and β_{CS} (which require knowledge of TX_{ij}) are not identifiable. However, in practice of observational rather than laboratory studies Berkson-type error may coexist with classical measurement error in different ratios. This was described by Berkson (1950) [30] as “modified controlled experimentation”. If so, then the estimate for β_{BS} is likely attenuated (i.e., as a β_{BS}^* similar to what has been described in Appendix C for classical measurement error) due to the classical measurement error component of \bar{x}_i that is computed from X_{ij} . More formal exploration of this hybrid quasi-Berkson-type error is given in context of occupational epidemiology in Kim et al. [18].

References

1. Diggle, P.J.; Heagarty, P.; Liang, K.Y.; Zeger, S.L. *Analysis of Longitudinal Data*; Oxford Press: New York, NY, USA, 2002.
2. Yang, Y.; Xie, M. Asymptotics for generalized estimating equations with large cluster sizes. *Ann. Stat.* **2003**, *31*, 310–347. [CrossRef]
3. Levey, A.S.; Bosch, J.P.; Lewis, J.B.; Greene, T.; Rogers, N.; Roth, D. A more accurate method to estimate glomerular filtration rate from serum creatinine: A new prediction equation. Modification of Diet in Renal Disease Study Group. *Ann. Intern. Med.* **1999**, *130*, 461–470. [CrossRef] [PubMed]
4. Barkan, S.E.; Melnick, S.L.; Preston-Martin, S.; Weber, K.; Kalish, L.A.; Miotti, P.; Young, M.; Greenblatt, R.; Sacks, H.; Feldman, J. The Women’s Interagency HIV Study. WIHS Collaborative Study Group. *Epidemiology* **1998**, *9*, 117–125. [CrossRef] [PubMed]
5. Littell, R.C.; Pendergast, J.; Natarajan, R. Modelling covariance structure in the analysis of repeated measures data. *Stat. Med.* **2000**, *19*, B1793–B1819. [CrossRef]
6. SPSS Inc. Linear Mixed-Effects Modeling in SPSS: Introduction to the MIXED Procedure). Available online: http://www.spss.ch/upload/1126184451_Linear%20Mixed%20Effects%20Modeling%20in%20SPSS.pdf (accessed on 11 January 2019).
7. Cui, J. QIC program and model selection in GEE analyses. *Stat. J.* **2007**, *7*, 209–220. [CrossRef]
8. Gardiner, J.C.; Luo, Z.; Roman, L.A. Fixed effects, random effects and GEE: What are the differences? *Stat. Med.* **2009**, *28*, 221–239. [CrossRef] [PubMed]
9. Shults, J.; Sun, W.; Tu, X.; Kim, H.; Amsterdam, J.; Hilbe, J.M.; Ten-Have, T.A. Comparison of several approaches for choosing between working correlation structures in generalized estimating equation analysis of longitudinal binary data. *Stat. Med.* **2009**, *28*, 2338–2355. [CrossRef]
10. Cheng, J.; Edwards, L.J.; Maldonado-Molina, M.M.; Komro, K.A.; Muller, K.E. Real Longitudinal Data Analysis for Real People: Building a Good Enough Mixed Model. *Stat. Med.* **2010**, *29*, 504–520. [CrossRef] [PubMed]
11. Gosh, M. Criteria to Select a Working Correlation Structure for the Generalized Estimating Equations Method in SAS. *J. Stat. Software* **2014**, *57*, 1–10. [CrossRef]
12. Tiwar, P.; Shukla, G. Approach of Linear Mixed Model in Longitudinal Data Analysis Using SAS. *J. Reliabil. Statist. Stud.* **2011**, *4*, 73–84.
13. Robinson, W.S. Ecological correlations and the behavior of individuals. *Am. Sociol. Rev.* **1950**, *15*, 351–357. [CrossRef]
14. Cronbach, L.J. *Research on Classifications and Schools. Formulations of Questions Designs and Analyses. Occasional Paper*; Stanford Evaluation Consortium: Stanford, CA, USA, 1976.
15. Firebaugh, G. A rule for inferring individual level relationships from aggregate data. *Am. Sociol. Rev.* **1978**, *43*, 557–572. [CrossRef]
16. Scott, A.J.; Holt, D. The effect of two-stage sampling on ordinary least square methods. *J. Am. Stat. Assoc.* **1982**, *77*, 848–854. [CrossRef]
17. Kim, H.M.; Richardson, D.; Loomis, D.; Van Tongeren, M.; Burstyn, I. Bias in the estimation of exposure effects with individual- or group-based exposure assessment. *J. Expo. Sci. Environ. Epidemiol.* **2011**, *21*, 212–221. [CrossRef] [PubMed]

18. Mancl, L.A.; Leroux, B.G.; DeRouen, T.A. Between-subject and within-subject statistical information in dental research. *J. Dent Res.* **2000**, *79*, 1778–1781. [[CrossRef](#)] [[PubMed](#)]
19. Pepe, M.S.; Anderson, G.L. A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Commun. Stat. Simul.* **1994**, *23*, 939–951. [[CrossRef](#)]
20. Mancl, L.A.; Leroux, B.G. Efficiency of regression estimates for clustered data. *Biometrics* **1996**, *52*, 500–511. [[CrossRef](#)]
21. Schildcrout, J.S.; Heagarty, P.J. Regression analysis of longitudinal binary data with time-dependent environmental covariates: bias and efficiency. *Biostat* **2005**, *6*, 633–652. [[CrossRef](#)]
22. Burstein, L. The analysis of multilevel data in educational research and evaluation. *J. Educ. Stat.* **1980**, *3*, 347–383. [[CrossRef](#)]
23. Raudenbusch, S.W.; Bryk, A.S. *Heirarchical Linear Models: Applications and Data Analysis Methods*, 2nd ed.; Sage Publications: London, UK, 2002.
24. Begg, M.D.; Parides, M.K. Separation of individual-level and cluster-level covariate effects in regression analysis of correlated data. *Stat. Med.* **2003**, *22*, 2591–2602.
25. Fuller, W.A. *Measure Error Models*; Wiley: New York, NY, USA, 1987.
26. Rifkin, R.D. Effects of Correlated and Uncorrelated Measurement Error on Linear Regression and Correlation in Medical Method Comparison Studies. *Stat. Med.* **1995**, *14*, 789–798. [[CrossRef](#)] [[PubMed](#)]
27. Preller, L.; Kromhout, H.; Heederik, D.; Tielen, M.J. Modeling long-term average exposure in occupational exposure-response analysis. *Scand. J. Work Environ. Health* **1995**, *21*, 504–512. [[CrossRef](#)] [[PubMed](#)]
28. Tielemans, E.; Kupper, L.L.; Kromhout, H.; Heederik, D.; Houba, R. Individual-based and group-based occupational exposure assessment: some equations to evaluate different strategies. *Ann. Occup. Hyg.* **1998**, *42*, 115–119. [[CrossRef](#)]
29. Berkson, J. Are There Two Regressions? *J. Am. Stat. Assoc.* **1950**, *45*, 164–180. [[CrossRef](#)]
30. Neuhaus, J.M.; Kalbfleisch, J.D. Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics* **1998**, *54*, 638–645. [[CrossRef](#)] [[PubMed](#)]
31. Zhang, J.; Hu, W.; Wei, F.; Wu, G.; Korn, L.R.; Chapman, R.S. Children’s Respiratory morbidity prevalence in relation to air pollution in four Chinese cities. *Environ. Health Perspect.* **2002**, *110*, 961–967. [[CrossRef](#)] [[PubMed](#)]
32. Miller, K.A.; Siscovick, D.S.; Sheppard, L.; Sheppard, K.; Sullivan, J.H.; Anderson, G.L.; Kaufman, J.D. Long term exposure to Air Pollution and Incidence of Cardiovascular Events in Women. *N. Eng. J. Med.* **2007**, *350*, 447–458. [[CrossRef](#)] [[PubMed](#)]
33. Pan, G.; Zhang, S.; Feng, Y.; Takahashi, K.; Kagawa, J.; Yu, L.; Wang, P.; Liu, M.; Liu, Q.; Hou, S.; et al. Air pollution and children’s respiratory symptoms in six cities of Northern China. *Respir. Med.* **2010**, *104*, 1903–1911. [[CrossRef](#)]
34. Crouse, D.L.; Peters, P.A.; Villeneuve, P.J.; Proux, M.A.; Shih, H.H.; Goldberg, M.S.; Johnson, M.; Wheeler, A.J.; Allen, R.W.; Atari, D.O.; et al. Within- and between-city contrasts in nitrogen dioxide and mortality in 10 Canadian cities; a subset of the Canadian Census Health and Environment Cohort (CanCHEC). *J. Expo. Sci. Environ. Epidemiol.* **2015**, *25*, 482–489. [[CrossRef](#)] [[PubMed](#)]
35. Morrison, B.; Shenklin, A.; McLelland, A.; Robertson, D.A.; Barrowman, M.; Graham, S.; Wuga, G.; Cunningham, K.J.M. Intra-individual variation in commonly analyzed serum constituents. *Clin. Chem.* **1979**, *25*, 1799–1805.
36. Lacher, D.A.; Hughes, J.P.; Carroll, M.P. *Biological variation of laboratory analytes based on the 1999-2002 National Health and Examination Survey*; National Health Statistics Reports; National Center for Health Statistics: Hyattsville, MD, USA, 2010.
37. Cuhadar, S.; Atay, A.; Koseoglu, M.; Dirican, A.; Har, A. Stability studies of common biochemical analytes in serum separator tubes with or without gel barriers subjected to various storage conditions. *Biochemia Media* **2012**, *22*, 202–214. [[CrossRef](#)]

