REPEATED MEASURES REGRESSION IN LABORATORY, CLINCAL AND ENVIROMENTAL RESEARCH - DIFFERENT BETWEEN/WITHIN SUBJECT SLOPES AND COMMON MISCONCEPTIONS

Donald R. Hoover[1], Qiuhu Shi[2], and Kathryn Anastos[3]

[1]Department of Statistics and Biostatistics and Institute for Health, Health Care Policy and Aging Research, Rutgers University, Piscataway, NJ, US, [2]School of Health Sciences and Practice/New York Medical College, Valhalla, NY, US, [3]Albert Einstein College of Medicine/Montefiore Medical Center, Bronx

Corresponding Author – Donald R Hoover

Email: drhoover@stat.rutgers.edu

Tel: 732-445-2691 (USA)

Fax: 732-445-3428 (USA)

Running Title – Repeated Measures Regression Common Misconceptions

Acknowledgements

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 UO1-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. We also thank Howard Strickler, Xiaonan Xue, Steve Gange, Peter Bacchetti, Minge Xie & Zhiqiang Tan for their helpful comments.

ABSTRACT

When using repeated measures linear regression models to make causal inference in laboratory, clinical and environmental research, it is often assumed that the Within Subject association of differences (or changes) in predictor value across replicates is the same as the Between Subject association of differences in those predictor values. But this is often false, for example with body weight as the predictor and blood cholesterol the outcome i) a 10 pound weight increase in the same adult more greatly a higher increase in cholesterol in that adult than does ii) one adult weighing 10 pounds more than a second reflect increased cholesterol levels in the first adult as the weigh difference in i) more closely tracks higher body fat while that in ii) is also influenced by heavier adults 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 (MM) on repeated measures (i.e. for fitting Cross Sectional Regression) is that the working correlation structure used only influences variance of model parameter estimates. But only independence working correlation guarantees the modeled parameters have any interpretability. We illustrate this with an example where changing working correlation from independence to equicorrelation qualitatively biases parameters of GEE models and show this happens because Between and Within Subject slopes for the predictor variables differ**.** We then describe several common mechanisms that cause Within and Between Subject slopes to differ as; change effects, lag/reverse lag and spillover causality, shared within subject measurement bias or confounding, and predictor variable measurement error. The misconceptions noted here should be better publicized in laboratory, clinical and environmental research. Repeated measures analyses should compare Within and Between subject slopes of predictors and when they differ, investigate the reasons this has happened.

KEY WORDS – Between/Within Subject Associations , Repeated Measures, Cross Sectional Regression, Generalized Estimating Equations, Mixed Models, Working Correlation

HIGHLIGHTS – When using repeated measures with time varying predictors variables in laboratory, clinical and environmental research:

- Cross sectional regressions with any working correlation structure other than independence often give non-meaningful results
- Between/Within subject decomposition of slopes should be undertaken when making causal inferences
- Investigators should investigate the reasons Between and Within Subject slopes differ if this occurs

ABBREVIATIONS

AR(1) – Autoregressive Order 1
BS – Between Subject
BUN – Blood Urine Nitrogen
Co-DOSE – Conditionally Dependent On Sibling Exposure
CS – Cross Sectional
E – Equicorrelation
EGFR – Estimated Glomerular Filtration Rate
GEE – Generalize Estimation Equations
IND – Independent
MA(2) – Moving Average Order 2
MM – Mixed Models
WIHS – Women's Interagency HIV Study
WS – Within Subject

1. Introduction

Two common misconceptions made in laboratory, clinical and environmental research fitting repeated measures regression with Generalized Estimating Equations (GEE) and Mixed Models (MM) are: Misconception-A: The association between the predictor variable and 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 can differ which should be considered when making causal inference. For example with weight as the predictor and cholesterol the outcome, i) a 10 lb. increase within the same person more likely indicates greater difference in serum cholesterol than does ii) one person being 10 lbs. heavier than another as i) more likely reflects body fat gain while ii) also can indicate that the heavier person is taller. 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 Table 1 in the next paragraph illustrates and Section 2 explains, a wrong choice for working correlation biases parameter estimates. Both of these misconceptions are related to each other, but the analytical details are complicated. So we begin with a direct illustration of Misconception -B that we later show relates to Misconception -A.

Table 1 presents parameter estimates from repeated measures linear regression using a clinical/laboratory measure glomerular filtration rate (EGFR) from the MDRD Formula[1] as the outcome Y and three laboratory predictor variables (X_1, X_2, X_3)=(HIV infection, serum albumin, blood urea nitrogen (BUN)) fit to 10,782 semiannual measures of 584 women at the Bronx-site of the Women's Interagency HIV Study (WIHS)[2]. Higher EGFR values indicate better renal function. The models fit in this Table assume that the Between and Within 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[3] with both independence (GEE-IND) in columns 2-4 and equicorrelation (GEE-E) columns 5-7 for the working correlation of model residuals of repeated measures in the same person. (The row notation used in this Table for the slope parameters; $\beta_{CS,,HIV}$ $\beta_{CS,,ALB}$ and $\beta_{CS,,BUN}$) is explained later in the paper).

INSERT TABLE 1 HERE

Most of today's literature providing guidance on fitting repeated measures linear regression (i.e. [4-12]) qualitatively describes working correlation as a "nuisance factor" that does not alter model parameters and states *the working correlation that minimizes variance of parameter estimates should be chosen*. However, in Table 1, the parameter estimate 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 estimate 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 changes the parameters; thus correlation structure is not merely a nuisance factor. We note that if Mixed Models[3] 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 (data not shown).

Faced with this confusing dilemma, 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 quasilikelihood 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) (data not shown)) most articles providing model fitting guidance [4-12] point towards using equicorrelation working correlation.

However, as Section 2 describes in detail, this guidance is problematic; only the parameter estimates obtained by using independence working correlation can have any meaning here as the model did not separately fit Within and Between Subject Slopes. Further as Section 3 explains, if the investigators' goal is causal inference, then decomposition of the parameters into Within and Between subject associations (or slopes) is needed. Sections 4 and 5 present epidemiological factors that cause Within and Between subject slopes to differ and some consequences of this. The Discussion (Section 6) summarizes and explores further implications for our example and other laboratory, clinical and environmental settings.

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 $i=1,2,\ldots n$. For most laboratory and clinical analyses the "subjects" will be persons with longitudinal repeated measures. But for environmental analyses "subjects" can be something else such as cities, schools, hospitals, etc. Each subject has J_i different observations enumerated by $j= 1, \ldots, J_i$ often taken at times $t_{i1} < t_{i2} <\ldots <t_{i,Ji}$, on the same person when person is the "subject". But for some studies, replicates are taken at the same time, such as from J_i neighborhoods in the same city when city is the "subject". For J_i constant across i, we denote J. The observations have continuous outcomes Y_{ij} and K predictor (or exposure) variables $\underset{\sim}{X}_{ij} = X_{1,ij}, X_{2,ij},\ldots.X_{K,ij}$. When K=1 we drop the "K" enumeration, using X_{ij} for the only predictor. Linear models for E[Y_{ij}| $\underset{\sim}{X}_{ij}$] or E[Y_{ij}| X_{ij}] are fit.

<u>2A Cross Sectional Regression (CS) Regression</u>. The most commonly fitted linear model does not separate "Within" and "Between" subject associations and is usually written out as Y_{ij} = $\alpha + \beta_1 X_{1,ij} + \beta_2 X_{2,ij} + \ldots. + \beta_K X_{K,ij} + \varepsilon_{ij}$. This is denoted "Cross Sectional Regression" for longitudinal repeated measures and we use this same nomenclature for settings where the repeated measures are not longitudinal. We also 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 Cross Sectional Regression model here is denoted:

$$Y_{ij} = \underset{\sim}{\beta}_{CS} \underset{\sim}{X}_{ij} = \alpha_{CS} + \beta_{1,CS} X_{1,ij} + \beta_{2,CS} X_{2,ij} + \ldots. + \beta_{K,CS} X_{K,ij} + \varepsilon_{ij} \qquad (1)$$

where $\alpha_{CS}, \beta_{1,CS}, \beta_{2,CS}, \ldots., \beta_{K,CS}$ are unknown parameters, while ε_{ij} is error with E[ε_{ij}]= 0 that is independent between different subjects i and i', but may be correlated for $j \neq j'$ within the same subject. Again for K=1 the subscript for K is dropped; the model is, $Y_{ij} = \alpha_{CS} + \beta_{CS} X_{ij} + \varepsilon_{ij}$. The goal of Cross Sectional Regression is most appropriately to obtain estimates $\hat{\underset{\sim}{\beta}}_{CS}$ for $\underset{\sim}{\beta}_{CS}$ to input $\hat{\underset{\sim}{\beta}}_{CS}$ into (1) order to estimate unobserved Y's from observed $\underset{\sim}{X}_{ij}$'s as $\hat{\underset{\sim}{\beta}}_{CS} \underset{\sim}{X}_{ij}$. Cross Sectional Regression is also used to make adjusted (causal) inference on the covariate associations in $\hat{\underset{\sim}{\beta}}_{CS}$ but as we show later, doing this may be problematic.

2B Between/Within Subject Slope (BS/WS) Regression. While regression models fit in laboratory, clinical and environmental studies typically do not consider this, it has long been noted that slopes on changes of X_{ij} within the same subject i differ from cross sectional associations[13-16]. To illustrate this consider the cross sectional model of a laboratory measure cholesterol (Y_{ij}) on the clinical outcome of body weight (X_{ij}); $E[Y_{ij}] = \alpha_{CS} + \beta_{CS} X_{ij}$. As described in the first paragraph of the Introduction, the cross sectional slope β_{CS} for association of a 10 lb. weight difference between two different subjects (i.e. persons) on cholesterol is less than the slope for association of a 10 lb "Within Subject" weight change for the same person on cholesterol which we denote as β_{WS}. Again the reason β_{CS} is less than β_{WS} is; a) part a 10 lb. cross sectional weight difference between two subjects often reflects greater height in one of the persons b) but a 10 pound weight increase in the same subject is not influenced by height difference and thus the is more likely due to more body fat being in the heavier weight. Thus since greater body fat is what is directly associated with more cholesterol, the within person association of a 10 lb. increase with cholesterol is greater than the cross sectional repeated measures association of a 10 lb. difference. Common within person body height reflects a shared *within subject measurement bias* on weight as a predictor of cholesterol. For example 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 $X_{ij} = TX_{ij} * H_i$ contains this shared within subject measurement bias E_i. Section 4 describes more settings where $\beta_{WS} \neq \beta_{CS}$.

INSERT FIGURE 1 HERE

Therefore linear regression models fit to make causal inference often decompose the associations into "Within Subject" slopes (β_{WS}) described above and "Between Subject" slopes (β_{BS}) described below that capture associations of subjects' central tendencies of the exposure. To do this, subject means of the predictor variables $\bar{x}_i = \bar{x}_{1i}, \bar{x}_{2i}, ..., \bar{x}_{Ki}$ are calculated, where $\bar{x}_{ki} = \sum_{j=1}^{J_i} X_{kji} / J_i$. Then Y_{ij} is modeled as a combination of "Between Subject" slopes from $\bar{x}_{ki}$ and "Within Subject" slopes from deviations of X_{kij} about $\bar{x}_{ki}$.

$$\begin{aligned} Y_{ij} = \alpha_{BS/WS} + \beta_{1,BS}(\bar{x}_{1i}) + \beta_{2,BS}(\bar{x}_{2i}) + ... + \beta_{K,BS}(\bar{x}_{Ki}) + \\ \beta_{1,WS}(X_{1ij} - \bar{x}_{1i}) + \beta_{2,WS}(X_{2ij} - \bar{x}_{2i}) + ... + \beta_{K,WS}(X_{Kij} - \bar{x}_{Ki}) + \varepsilon_{ij} \end{aligned} \quad (2)$$

Or for K = 1, $Y_{ij} = \alpha_{WS/BS} + \beta_{BS}(\bar{x}_i) + \beta_{WS}(X_{ij} - \bar{x}_i) + \varepsilon_{ij}$. To illustrate this for our earlier example with Y_{ij}=Cholesterol and X_{ij}=Weight, let $\alpha_{WS/BS} = 30$, β_{BS}=0.9 and β_{WS}=3;

then $Y_{ij} = 30 + 0.9\bar{x}_i + 3(X_{ij} - \bar{x}_i) + \varepsilon_{ij}$. If person *i* had an average value of $\bar{x}_i = 210$ across all J_i measures with the jth measure being X_{ij}=200 then for that person-visit t_{ij},

$E\left[Y_{ij}\right] = 30 + 0.9(210) + 3(200 - 210) = 189.$

Now for some technical asides; First – the choice of the observed $\bar{x}_{ki}$ as the "central tendency" of X_{kij} for subject *i* is necessary as "μ_{ki}" a person's "true average weight" over the entire time period is unknown, but for J_i large enough, $\bar{x}_{ki}$ should be close to μ_{ki}. Thus while $\beta_{k,WS}$ only captures association with within subject change in in X_{kij}, $\beta_{k,BS}$ inherently contains some $\beta_{k,WS}$ from deviation of ($\bar{x}_{ki}$ - μ_{ki}); especially for small J_i. 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 weight gain of 10 lbs. in one month creates a shock that hyper-elevates cholesterol, but a 10 lb. weight gain over 12 months does not; in which case $\beta_{WS} \mid t_j - t_{j'} = 1 > \beta_{WS} \mid t_j - t_{j'} = 12$. But it is probably reasonable to assume that any such differences are minor.

In spite of these technical limitations, Between/Within-Subject decomposition is used including to test if $\beta_{k,BS} = \beta_{k,WS}$ in which case as shown in Section 2C they also equal $\beta_{k,CS}$ and the separated *BS*, *WS* decomposition can be collapsed. Due to orthogonal decomposition of X_{kij} about $\bar{x}_{ki}$, this previous test for collapsing the *BS*, *WS* decomposition is a two sample z-test of parameter estimates from fitted models comparing $|\hat{\beta}_{k,BS} - \hat{\beta}_{k,WS}| / \sqrt{\mathrm{var}(\hat{\beta}_{k,BS}) + \mathrm{var}(\hat{\beta}_{k,WS})}$ to $Z_{1-\alpha/2}$ [16]. Between/Within 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 X_{0ij} as such is often done when only one observation per subject is available; hence $X_{kij} \equiv \bar{x}_{ki}$.

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 correlations) is eliminated after making a Between – Within Subject decomposition. There were no new HIV infections after study entry; so $X_{HIV,ij} \equiv \bar{x}_{HIV,i}$ making Within Subject association of HIV unmeasurable. For Within Subject associations of BUN and Albumin, GEE-IND and GEE-E gave identical point estimates as centering about $\bar{x}_{ki}$ makes comparisons entirely within subject and invariant to these correlation choices (although within subject estimates would differ slightly if autoregressive (AR(1)) or other formulations for intrasubject correlation of residuals had been used). There were small GEE-IND, GEE-E differences on the Between Subject slopes as has been observed elsewhere[17]. 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 Vs. -1.57; 95% CI (-4.47, 1.33) with GEE-E.

INSERT TABLE 2 HERE

From now on, we only examine GEE-IND results for Between/Within subject decomposition models as GEE-E results are essentially similar (data not shown). For BUN, the GEE-IND the Between Subject $\hat{\beta}_{\mathrm{BUN,B}S}$ = -2.72, 95% CI (-3.10, -2.33) is qualitatively and statistically farther from 0 than the corresponding Within Subject slope; $\hat{\beta}_{\mathrm{BUN},WS}$ = -1.11, 95% CI (-1.34, -0.88). But serum albumin goes the other way; the Within subject slope $\hat{\beta}_{\mathrm{ALB,W}S}$=-10.70, 95% CI (-12.99, -8.40) is statistically further from 0 than the corresponding Between subject GEE-IND $\hat{\beta}_{\mathrm{ALB,B}S}$=-3.27 with a 95% CI (-7.88, 1.33) that overlaps 0.

So how can one interpret differences in the Between and Within Subject slopes in particular for causal inference? This depends on hypotheses of interest (and we do not have any for this illustrative example). But general rules also apply although we are unaware of any systematic exploration of reasons that between $\beta_{\mathrm{B}S}$ (or $\beta_{\mathrm{B}S}$ for K=1) could differ from within subject $\beta_{\mathrm{W}S}$ (or β_{WS}) slopes and implications for causal inference. But before going to this here it is important to first note the relationship between, Cross Sectional, Within Subject and Between Subject slopes.

2C Relationship Between β_{CS}, β_{WS} and β_{BS}. β_{CS} averages β_{WS} and β_{BS} according to relative variances of the within subject means Vs. the variance of the repeated measures about those sample means[16]. For example, with K=1, if $\sigma^2_{\bar{x}_{i.}}$ is the population variance of the within person mean $\bar{x}_i$ and $\sigma^2_{X-\bar{x}_{i.}}$ is the population variance of the deviations of differences of the repeat measures X_{ij} from their $\bar{x}_i$, then

$$\beta_{CS} = \beta_{BS}\sigma^2_{\bar{x}_{i.}} / (\sigma^2_{X-\bar{x}_{i.}} + \sigma^2_{\bar{x}_{i.}}) + \beta_{WS}\sigma^2_{X-\bar{x}_{i.}} / (\sigma^2_{X-\bar{x}_{i.}} + \sigma^2_{\bar{x}_{i.}}). \quad (3)$$

In the previous example of weight and cholesterol with $Y_{ij} = 30 + 0.9\,\bar{x}_i + 3(X_{ij} - \bar{x}_i) + \varepsilon_{ij}$, if $\sigma^2_{\bar{x}_{i.}} = 400$ and $\sigma^2_{X-\bar{x}_{i.}} = 100$, then $\beta_{CS} = 0.9*400/(100+400) + 3*100/(100+400)$ = 1.32. If the sample means are more homogeneous in weight with $\sigma^2_{\bar{x}_{i.}}$= 200 and $\sigma^2_{X-\bar{x}_{i.}}$= 100 still, then 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 involve specification of correlation (or working correlation) of ε_{ij} within the same subject i. We denote the working correlation structure by $\boldsymbol{V_i}$ which is a matrix. Typical choices for $\boldsymbol{V_i}$ are equicorrelation (E) with correlation of ε_{ij} and $\varepsilon_{ij'}$ for j ≠ j' always the same value ρ and independence (IND); with correlation of ε_{ij} and $\varepsilon_{ij'} \equiv 0$ as used in the illustrative examples of Table 1 and Table 2, and also AR(1) where correlation of ε_{ij} and $\varepsilon_{ij'}$ is $\rho^{|j-j'|}$ [3]. Again, current guidance articles[4-12] emphasize choosing the $\boldsymbol{V_i}$ 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, as doing so often improves precision of the model parameter estimates. However, this approach may be wrong for Cross Sectional Regression, because using any correlation structure other than independence can introduce structural bias into $\hat{\beta}_{CS}$ that the standard approaches used to chose the best working correlation structure such as intrasubject residual correlations, the AIC and QIC mentioned earlier do not account for[18, 19].

From Pepe & Anderson (1994)[18], for Cross Sectional Regression if X_{ij} varies within *i* and,

$$E[Y_{ij} \mid X_{ij}] \text{ depends on } X_{k,ij'} \text{ for any } k \text{ of a different replicate } j' \text{ in } i \qquad (4)$$

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 $\boldsymbol{V_i}$ = Independence, gives biased estimates of β_{CS}. Thus the only working correlation structure that should be used to estimate β_{CS} is $\boldsymbol{V_i}$ = IND. However, if (4) does not hold, then any working correlation structure obtains unbiased estimates for β_{CS} in which case, choosing the $\boldsymbol{V_i}$ that most accurately gives the correlation structure of ε_{ij} minimizes the variance of $\hat{\beta}_{CS}$.

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 "**Co**nditional **D**ependence **O**n **S**ibling **E**xposures" (i.e. on $X_{k,ij'}$). Thus the sibling exposure $X_{k,ij'}$ could be thought of as a "Co-DOSE" beyond the "dose" from the "self-exposure". Hence we use "Co-DOSE" to denote that (4) occurs.

Also while this point has not very well made, for Cross Sectional Regression, Co-DOSE or (4) largely occurs if and only if Between and Within Subject slopes differ. If Between and Within Subject slopes differ for any predictor (i.e. $\beta_{BS} \neq \beta_{WS}$) then Co-DOSE (4) happens. But if the Between and Within Subject slopes are equal for all predictors (i.e. $\beta_{BS} = \beta_{WS}$) then (4) does not occur. More details on this and an illustration are given in Appendix 1. But one trivial case is if the predictors are invariant within the same subject; (i.e. $X_{i1} \equiv X_{i2} \equiv \ldots \equiv X_{iJi} \equiv \bar{x}_i$ then the Within Subject slopes are not defined since $X_{ij} - \bar{x}_i \equiv 0$ and for the same reason Co-DOSE cannot occur. While the mathematical details are beyond this paper, if $\beta_{BS} \neq \beta_{WS}$, when $\mathbf{V_i} \neq IND$, the non-zero covariances $\rho_{ij} > 0$ besides adjusting for within *i* collinearity of ε_{ij} *also* overweight the β_{BS} relative to β_{WS} in (3) pushing "Cross Sectional Regression" parameter estimates away from β_{CS} towards β_{WS}[16]. Since robust covariance methods exist to adjust for

impact of misspecification of $\boldsymbol{V_i}$=IND from collinearity of the ε_{ij} on variance estimates in particular for GEE[3], $\mathbf{V_i} = IND$ can eliminate bias in estimating β_{0CS} and still provide a conservative variance for the estimate.

2E Implications for Clinical, Laboratory and Environmental Research. Much of what has been presented above is not commonly understood and implemented in clinical, laboratory and environmental research. Cross sectional models are typically fit, with β_{0CS} interpreted to also be β_{0BS} and β_{0WS} without checking if these slopes are equal. Non-independence $\boldsymbol{V_i}$ is often used for Cross Sectional Regression without checking if Co-DOSE (4) exists. Systematic epidemiological description of causal mechanisms for why Between and Within subject slopes can differ for clinical, laboratory and environmental measures have not been undertaken. We thus do this next in Section 3.

3. Classification of 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 given settings, (especially for clinical, laboratory and environmental) we classify major reasons this happens. For simplicity now let K=1 unless otherwise noted as the principles below 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 ($X_{i'j} - \bar{x}_{i'}$=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 $\beta_{WS} > \beta_{BS}$ and Co-DOSE (4) occurs. But it should be noted that as was mentioned in Section 2B, in this setting, β_{WS} would be undefined if say a 10 lb. gain in a shorter time period (i.e. 1 month) increases β_{WS} more than does a 10 lb. gain over a longer time period (i.e. 12 months).

3B Lag Causality of *X* on Future *Y*. The effect of prior 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. So as illustrated in Figure 2a, high HIV viral load at t_1 may affect CD4 loss from t_1 to t_2 (*lag causality of X at t_1 on Y at 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 that higher viral load at t_1 having created more CD4 destruction between t_1 and t_2. Thus $Y_{i,2}|X_{i2}$ at t_2 is not independent of $X_{i,1}$ at t_1; Co-DOSE (4) occurs and Within / Between Subject slopes differ; $\beta_{BS} \neq \beta_{WS}$. Lag causality is often considered when serial measures of X = long term environmental exposures (such as air pollution) that effect chronic conditions Y (such as lung function) are obtained[3,20].

INSERT FIGURE 2 HERE

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 ($Y_{i,1}$) at t_1 may have degraded the CD4 count from t_1 to t_2. Thus $Y_{i,1}|X_{i1}$ at t_1 is not independent of $X_{i,2}$ at t_2; Co-DOSE (4) occurs and Between / Within subject slopes differ ; $\beta_{BS} \neq \beta_{WS}$.

3D Spillover Causality of X on Adjacent Y. The setting of 3B 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 repeat measures are X_{ij} = average air pollution of neighborhood *j* in city *i* and Y_{jj} = average lung function of all residents living within neighborhood *j* of city *i*. A resident living in neighborhood *j* may work in different neighborhoods *j'* of the same city and thus have "*spillover exposure*" to air in the neighborhood they work in, for a given city i, $Y_{i,i}|X_{ij}$ is not independent of $X_{ij'}$ and hence Co-DOSE occurs.

3E Common within Subject Measurement Bias. Shared *within subject measurement bias* occurs if all repeat measures from the same subject have the same positively 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 people being heavier independently of body fat than shorter persons which lead to $\beta_{WS} > \beta_{CS}$ and Co-DOSE (4) when weight was a predictor of cholesterol. In this setting height is a *measurement bias* not as a confounder as height itself is not associated with cholesterol. We now present a similar setting where the unmeasured variable is a confounder.

3F Common within Subject Confounding. Figure 1b shows a diagrammatically similar phenomenon, that causes $\beta_{WS} \neq \beta_{CS}$ and Co-DOSE (4), *common within subject confounding* where now the extraneous factor shared by the repeated measures of the same subject is associated with Y. For example, let the confounder variable C_i = sex of subject i (which does not change with j) not be in the model and the outcome $Y_{i,j}$ be a linear score for male pattern baldness at time j with again X_{ij} being weight at time j. Now men are both heavier and independently of weight have greater male pattern baldness than women. So C_i is associated with the outcome. Here a 10 lb. weight difference in 2 persons, but not a within person increase of 10 lbs., could be informative of the heavier person more likely being male. Hence for this example, $\beta_{WS} = 0$ (assuming within person weight does not influence baldness), but $\beta_{BS} > 0$ and hence also $\beta_{CS} > 0$ reflecting unaddressed between subject confounding from heavier persons more likely to be men. Similarly, Mancl, Leroux & DeRouen proposed that unmeasured treatment compliance as a confounder could lead to Between/Within subject slope differences when j enumerated different teeth in the same subject[17].

In non-longitudinal settings 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"[21,22]. For example as Robinson (1950)[13] observed, when X was race of the student (White=0, Black=1) and Y was achievement-score, a higher $\bar{x}_i$ (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. But within the same school Black and White students performed equally well, thus β_{WS} was 0. Begg & Parides (2003) identify a similar setting in birthweight and IQ in families[23].

3G Measure Error in X_{ij} Makes E[Y_{ij}|X_{ij}] Dependent on $X_{ij'}$ It has been shown that, measure error on X that is independent of Y[24] or correlated with Y[25] biases estimates for association of X with Y. Now measure 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 measure error, then Co-DOSE (4) occurs and Within / Between subject slopes differ, $\beta_{BS} \neq \beta_{WS}$, because the biases being created from the measure error distribute differentially to different slopes. Figure 3 shows, if X_{i1} incompletely measures the true state TX_{i1} (i.e. true BUN) then $X_{i2,}$ is informative for TX_{i1} even after considering X_{i1}.

INSERT FIGURE 3 HERE

For example, Going back to the analysis of Table 1, let X_{ij} = BUN and Y_{ij} = EGFR. Let two persons have BUN of X_{i1}=10 mg/dL measured today but with Measure Error in BUN Assume the true BUN state changes slowly and after 6 months one of these persons measures X_{i2}=20 mg/dL while the other measures X_{i2}=5 mg/dL. Since BUN changes slowly, it is more likely that the True BUN today (TX_{i1}) of the former person is > 10 mg/dL and of the later 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}, thus 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 (4) occurs and Between/Within subject slopes differ. As Appendix 2 shows, measure error on 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 measure error has been proposed as a reason $|\beta_{WS}|<|\beta_{BS}|$ was observed in dental research[17].

But if M_{ij} is correlated with Y_{ij} (most likely correlated with measure error on Y_{ij}[25]) 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 creating negative β_{BS}, β_{WS} and β_{CS} for TY_{ij} on TX_{ij}; subjects with higher CD8 component of serum lymphocytes by converse having lower CD4 components. But the measured lymphocyte count (Z) is subject to a correlated measure error that equally spreads onto X and Y; for example if a person is dehydrated, the entire lymphocyte (meaning both CD8=X and CD4=Y) portions of blood become higher. If a person-

sample has a high (or low) measured lymphocyte count Z_{ij} = TZ_{ij}+M_{ij}, due to such sampling/measurement error, M_{ij} distributes onto both CD4 (X_{ij}) and CD8 (Y_{ij}) components making both simultaneously higher (or lower). Thus within person, higher measured CD4 count due to positive M_{ij} is associated with higher measured CD8 as the "M_{ij}" is shared making β_{WS} drawn towards being positive. But β_{BS} that tempers down M_{ij} on both X and Y through averaging as shown in Appendix 2 for X is less affected.

4. Predictors having Co-DOSE will bias adjusted parameter estimates of predictors not having Co-DOSE in cross sectional regression when $V_i \neq$ Independence is used

Going back to Table 1, it was shown earlier that the adjusted point estimate from GEE-IND $\hat{\beta}_{HIV,CS}$ for the adjusted cross sectional association of HIV with EGFR is still consistent for $\beta_{\mathrm{HIV},CS}$. However, HIV infection status was constant over all replicates within the same subject so cannot have Co-DOSE (4) as the entire effect of HIV is mediated between subject, not within subject so the question arises whether the adjusted estimate from an non-independence correlation structure $\hat{\beta}_{HIV,CS-E}$ can be biased for $\beta_{\mathrm{HIV},CS}$. 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. <u>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 biased</u>. The specific question addressed here is could having BUN and albumin that each have Co-DOSE (4) 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 $\beta_{\mathrm{HIV,CS}}$ in the multivariate model even though HIV itself is not Co-DOSE? This is important as $\hat{\beta}_{HIV,CS-E}$ = -2.04 95% CI (-5.07 0.98) qualitatively differs from $\hat{\beta}_{HIV,CS}$ 3.96 (-6.90, -1.03) in Table 1 with only $\hat{\beta}_{HIV,CS-E}$ statistically (P < 0.01) differing from 0.

We believe $\hat{\beta}_{HIV,CS-E}$ for HIV is biased away from $\beta_{HIV,CS}$ and to make this point refer to Table 3 which presents normative data broken down by HIV status of the subjects. A) From

Table 1, $\hat{\beta}_{BUN,CS-E}$ is biased higher for $\beta_{BUN,CS}$ (with GEE-E $\hat{\beta}_{BUN,CS-E}$ = -1.22 > $\hat{\beta}_{\mathrm{BUN},CS}$ = -1.87 for 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 "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. B) Similarly, from Table 1, $\hat{\beta}_{\mathrm{ALB},CS-E}$ is biased lower for $\beta_{ALB,CS}$ (with GEE-E $\hat{\beta}_{\mathrm{ALB},CS-E}$ = -9.84 < $\hat{\beta}_{\mathrm{ALB},CS}$ =-6.21) while from Table 3, HIV+ individuals have lower mean albumin; (3.97 Vs. 4.10, P < 0.0001 from GEE-IND). Thus the "positive effect" of the lower albumin in HIV+ subjects from $\beta_{\mathrm{ALB},CS}$ is overestimated by $\hat{\beta}_{\mathrm{ALB},CS-E}$ which pushes $\hat{\beta}_{HIV,CS-E}$ further down to compensate. C) Thus as Figure 4 shows, these deficits in A) and B) above are added to push $\hat{\beta}_{HIV,CS-E}$ downwards from the true adjusted $\beta_{HIV,CS}$. Therefore, non-independence $\boldsymbol{V_i}$ can bias multivariate cross sectional parameter estimates of variables that do not carry Co-DOSE (4) if other variables in the model carry Co-DOSE.

INSERT TABLE 3 AND FIGURE 4 HERE

5. Discussion

New GEE and MM cross sectional regression models from repeated measures of predictors that vary within subject that either use non-independence working correlations or do not state the correlation structure used continue to be published in the clinical, laboratory and environmental sciences literature. These papers do not show awareness of the points presented in Sections 1–4. Specifically, they either; a) do not specify if the coefficients of interest are β_{CS}, β_{WS} or β_{BS} nor check if $\beta_{BS} = \beta_{WS}$, b) make potentially invalid β_{CS} interpretations from MM and GEE using non-independence correlation $\boldsymbol{V_i}$**'s**, and/or c) do not justify the choice of non-independence working correlations $\boldsymbol{V_i}$ in light of potential differences between β_{WS}, β_{BS} and β_{CS}. We found >30 such papers from a quick literature search (including some authored by us before becoming aware of these issues); this is probably only a fraction of the total.

Yet papers published up to 65 years ago either warn against using non-independence working correlation structure in cross sectional regression with repeated measures[17,18], or to decompose the associations into Within Subject (β_{WS}) and "Between Subject" (β_{BS}) to make causal inference[13-16]. Numerous examples where $\beta_{CS} \neq \beta_{BS} \neq \beta_{WS}$ have been presented[13-18,20-23]. While it was not covered in our paper, this includes fitting GEE models of binary outcomes where the issues discussed here also apply[17,26]. But these points are still not well known or emphasized in statistical software documentation and papers providing guidance on GEE and MM analyses (i.e.[4-12]).

One problem that impedes acceptance of Within Subject (β_{WS}) and "Between Subject" (β_{BS}) decomposition is that it leads to much more complicated models that are very difficult to explain. Still some studies in environmental research have considered one mechanism described in Section 3. For example, lag causality has been considered in longitudinal analyses of association of air pollution on health measures[3]. Other air pollution / health studies have fit Within and Between Subject decompositions using cities as the subject and neighborhoods as the repeated measures within the city[27-30]. Most often in these studies the magnitude was greater for Within subject slope $|\beta_{WS}| > |\beta_{BS}|$, but sometimes $|\beta_{BS}| > |\beta_{WS}|$ meaning possibly multiple etiologies are involved. Those papers that did attempt to explain the reasons for the differences described common within subject confounding (Section 3E), such as unmodeled pollutants that were correlated between (but not within) cities with the modeled pollutants of interest as a potential reason.

We concur with others[17,18], that Cross Sectional Regression with repeated measures should use independence as the default working correlation unless justification is given to use other $\boldsymbol{V_i}$. While non-independence $\boldsymbol{V_i}$ can improve precision [20] 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 for was observed association with HIV in GEE-E compared to the more appropriate P=0.19 from GEE-IND.

While showing this is beyond the scope of our paper, when $\boldsymbol{V_i}$ is not independence, factors such as J_i and magnitude/structure of ε_{ij} strongly influence parameter estimates for β_{CS} from the misfitted Cross Sectional Models allowing the misfitted estimate to arbitrarily range from β_{WS} to β_{CS}. Standardization is important; and as such factors will arbitrarily vary between studies, parameter estimates of β_{CS} become harder to compare across studies when $\boldsymbol{V_i}$ differs at discretion of investigators. The working correlation structure used in Cross Sectional Regressions using repeated measures should thus always be reported.

We also concur with others [13-17,21-23] that in spite of the difficulties in identifying why "Within" and "Between subject slopes differ, causal inference analyses with repeated measures should initially make such decompositions and investigators be wary if there are qualitative differences between β_{BS} and β_{WS}. The example, Table 2 with 584 subjects and 10,782 measurements demonstrated need for β_{BS}, β_{WS} decomposition to make causal inference (as well as for using GEE-IND in cross sectional regression). But a smaller study could have been less clear-cut. If the same point estimates for β_{BS} and β_{WS} seen in Table 2 were observed but did not statistically differ, one would be tempted to merge β_{BS} and β_{WS} into a combined β_{CS} at least for some variables as 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 $\hat{\beta}_{BS}$ and $\hat{\beta}_{WS}$ for comparison to future studies and target potential mechanisms for Between/Within slope differences as described in Section 3.

Unfortunately, Between/Within Subject 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 $\bar{x}_{k,i}$ is ill-defined by averaging the $X_{k,ij}$ rather than being a true

mean for subject *i* creates confusion about interpretation of $\hat{\beta}_{BS}$ which can also be influenced by within subject slopes as 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 Table 2 with EGFR as the outcome, for BUN the Between Subject slope $\hat{\beta}_{BUN,BS} = -2.72$ (from GEE-IND) was statistically farther from 0 in the expected direction of association than was the Within Subject slope $\hat{\beta}_{BUN,WS} = -1.11$. But albumin went the other way; Within Subject slope $\hat{\beta}_{\mathrm{ALB,W}S}$ -10.70 was statistically farther from 0 than was Between Subject slope $\hat{\beta}_{BUN,\mathrm{W}S} = -3.27$. So what are the potential reasons for this?

While lag/reverse lag causality between serum BUN and creatinine (the main component of calculated EGFR) which could reduce magnitude of $\beta_{BUN,WS}$ Vs. $\beta_{BUN,BS}$ (Sections 3B and 3C), this was unlikely given separation of visits was 6 months and internal biochemistry operates over shorter time periods. But independent measure error on BUN (Section 3E) would temper $|\beta_{BUN,WS}|$ relative to $|\beta_{BUN,BS}|$. Several articles find greater; coefficient of variation [31,32], within person change[31,32], assay error[32], and sample degradation for serum BUN Vs. albumin measures[33], all of which could reflect BUN having larger independent measure error than does albumin that would selectively temper $\hat{\beta}_{BUN,\mathrm{W}S}$ (i.e., more than $\hat{\beta}_{\mathrm{ALB,W}S}$) to 0.

Conversely, serum creatinine and albumin are both constrained into the intravascular fluid compartment and will noninformatively increase together with greater hydration and decrease with less hydration of this compartment inducing positively correlated measurement error; much as was the case for measured CD4 and CD8 cells in the last paragraph of Section 3F. As creatinine factors inversely into EGFR calculation, this would constitute negative correlation of measure error between albumin and EGFR and selectively bias $\hat{\beta}_{\mathrm{ALB,W}S}$ to be more negative than $\hat{\beta}_{\mathrm{ALB},BS}$. However, BUN, which crosses across all body compartments, is less subject to such correlation in measure error with creatinine and thus with EGFR.

When Between and Within subject slopes differ, $(\beta_{BS} \neq \beta_{WS})$ it is unclear which is the "least confounded/biased", including the possibility that by "averaging" the different biases in each that β_{CS} could be least biased. There may be a heuristic perception to believe that by "matching within the same subject", β_{WS} is superior to $\underset{\sim}{\beta}_{BS}$ and β_{CS}, but this is not necessarily true as Measure Error on X (Section 3F) and lag / reverse lag and spillover causality (Sections 3B-3D) can in fact bias β_{WS} to a larger degree than happens for $\underset{\sim}{\beta}_{BS}$ and β_{CS}.

To conclude, for decades it has been known by some that when exposures vary within subjects in repeated measures regression then, i) Cross Sectional Regression using $\boldsymbol{V_i}$=Independence working correlation should be the default if building a model to estimate a future unknown Y is the goal, and ii) Between/Within 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 Between and Within Subject slopes differ and limited substantive study of the mechanisms that cause such differences. Clinical, laboratory and environmental studies should explore and quantify reasons for bias that occur in order to provide groundwork to improve future studies. To that end, analyses using repeated measures regression should investigate if Between and Within Subject slopes differ and when they do, try to identify the reasons for this including; change effect, lag/reverse lag and spillover causality, shared within subject measurement bias or confounding, and/or measure error on exposures.

References

1. 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(6),** 461-70. doi:10.7326/0003-4819-130-6-199903160-00002.

2. 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(2),**117-25.

3. Diggle P. J., Heagarty P. , Liang K. Y., Zeger SL. *Analysis of Longitudinal Data.* Oxford Press. New York; 2002.

4. Littell R. C., Pendergast J., Natarajan R. Modelling covariance structure in the analysis of repeated measures data. *Statist. Med.* 2000;**19**,1793-1819. doi:10.1002/sim.4073.

5. SPSS Inc. Linear Mixed-Effects Modeling in SPSS: Introduction to the MIXED Procedure) http://www.spss.ch/upload/1126184451_Linear%20Mixed%20Effects%20Modeling%20in%20SPSS.pdf ;2005; Accessed August 21, 2014.

6. Cui J. QIC program and model selection in GEE analyses *The Stata Journal* 2007;7(2): 209- 220.

7. Gardiner J. C., Luo Z., Roman L. A. Fixed effects, random effects and GEE: What are the differences? *Statist. Med.* 2009;**28**,221–239. DOI: 10.1002/sim.3478.

8. Shults J., Sun W., Tu X., Kim H., Amsterdam J., Hilbe J. M., Ten-Have T. A. (2009) Comparison of several approaches for choosing between working correlation structures in generalized estimating equation analysis of longitudinal binary data. *Statist. Med*. 2009;28,2338–2355**.** DOI: 10.1002/sim.3622.

9. 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. *Statist Med.* 2010;**29(4)**, 504–520. doi:10.1002/sim.3775.

10. IBM Inc. SPSS Manual http://publib.boulder.ibm.com/infocenter/spssstat/v20r0m0/index.jsp?topic=%2Fcom.ibm.spss.statistics.help%2Falg_genlin_gee_modeltest_goof.htm 2011; Accessed August 21, 2014.

11. Tiwar P., Shukla G. Approach of Linear Mixed Model in Longitudinal Data Analysis Using SAS *Journal of Reliability and Statistical Studies* 2011; **4(1)** (2011): 73-84.

12. SAS Institute Inc. (2011) SAS/STAT® 9.3 User's Guide. Cary NC http://support.sas.com/documentation/onlinedoc/stat/930/genmod.pdf ; 2011 http://support.sas.com/documentation/onlinedoc/stat/930/mixed.pdf ; 2011. Both accessed August 31, 2014.

13. Robinson W.S. Ecological correlations and the behavior of individuals. *American Sociological Review* 1950;15, 351-7.

14. Cronbach L.J. *Research on classifications and schools. Formulations of questions Designs and analyses. Occasional paper.* Stanford Ca. Stanford Evaluation Consortium; 1976.

15. Firebaugh G. A rule for inferring individual level relationships from aggregate data. *American Sociological Review* 1978; 43, 557-2.

16. Scott, A. J. & Holt, D. The effect of two-stage sampling on ordinary least square methods. *J. Amer. Statist. Assoc.* 1982;77, 848-854.

17. Mancl L.A., Leroux B.G., DeRouen T.A. Between-subject and within-subject statistical information in dental research. *Journal of Dental Research* 2000; **79**,1778-81.

18. Pepe MS., Anderson G. L. A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Comm Stat Simula* 1994;**23,** 939-51. doi:10.1080/03610919408813210.

19. Mancl L.A., Leroux B.G. Efficiency of regression estimates for clustered data. *Biometrics* 1996;**52**,500-511.

20. Schildcrout JS, Heagarty, PJ Regression analysis of longitudinal binary data with time-dependent environmental covariates: bias and efficiency *Biostat* 2005;**6(4)**,633-652.

21. Burstein L The analysis of multilevel data in educational research and evaluation *J Edu Stat* 1980;**3**:347-83.

22. Raudenbusch SW, Bryk AS *Heirarchical Linear Models: Applications and Data Analysis Methods 2nd Edition* Sage Publications London, 2002.

23. 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**,259—2602.

24. Fuller, W. A. *Measure Error Models*, New York: Wiley, 1987.

25. Rifkin R.D. Effects of Correlated and Uncorrelated Measurement Error on Linear Regression and Correlation in Medical Method Comparison Studies *Stat Med* 1995;**14(8)** 789-98.

26. Neuhaus, J. M. and Kalbfleisch, J. D. Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics* 1998;**54**, 638-645..

27. Zhang J, Hu, W, Wei F, Wu G, Korn LR, Chapman RS. Children's Respiratory morbidity prevalence in relation to air pollution in four Chinese cities. *Env Hlth Persp* 2002;**110(9)**, 961 – 967.

28. Miller KA, Siscovick DS, Sheppard L, Sheppard K, Sullivan JH, Anderson GL, Kaufman JD. Long term exposure to Air Pollution and Incidence of Cardiovascualr Event in Women. *New Eng J Med* 2007; **350(5)**, 447-458.

29. Pan G, Zhang S, Feng Y, Takahashi K, Kagawa J, Yu, L, Wang P, Liu M, Liu Q, Hou S, Pan B, Li J. Air pollution and children's respiratory symptoms in six cities of Northern China. 2010; **104**, 1903-11.

30. Crouse DL, Peters PA, Villenueve PJ, Proux MA, Shih HH, Goldberg MS, Johnson M, Wheeler AJ, Allen RW, Atari DO, Jerrett M, Brauer M, Brook JR, Cakmak S, Burnett RT. Within- and between-city contrasts in nitrogen dioxide and mortlaiy in 10 Canadian cities; a subset of the Canadian Census Health and Environment Cohort (CanCHEC). 2015; **25**, 482-89.

31. 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(10)**, 1799-1805.

32. 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* Hyattsville MD, 2010.

33. 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(2)**,202-14.

Table 1 – Cross Sectional Regression Using GEE[1] of EGFR = HIV infection, Serum Albumin and BUN in the Bronx WIHS

Variable	Working Correlation Structure[2]					
	Independence			Equicorrelation		
	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-Information Criteria (QIC)	10847.14			10836.27		

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

Table 2 – Between Subject and Within Subject Decomposition Regression Using GEE[1] for EGFR = HIV Infection, Albumin and BUN in the Bronx WIHS

Variable	Compartment	Independence Working Correlation			Equicorrelation Working Correlation		
		Point Estimate	95% CI	Z-Value (P)	Point Estimate	95% CI	Z-Value (P)
HIV Infection	Between Subject ($\beta_{HIV,BS}$)	-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 ($\beta_{ALB,BS}$)	-3.27	(-7.88, 1.33)	-1.39 (0.16)	-2.71	(-7.00 1.57)	-1.24 (0.21)
	Within Subject ($\beta_{ALB,WS}$)	-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 ($\beta_{BUN,BS}$)	-2.72	(-3.10, -2.33)	-13.89 (<0.0001)	-2.65	(-3.01 -2.08)	-14.21 (<0.0001)
	Within Subject ($\beta_{BUN,WS}$)	-1.11	(-1.34, -0.88)	-9.31 (<0.0001)	-1.11	(-1.34, -0.88)	-9.31 (<0.0001)

1. Mixed models gave essentially similar point estimates
2. There is no Within Subject Variation for HIV Infection Status

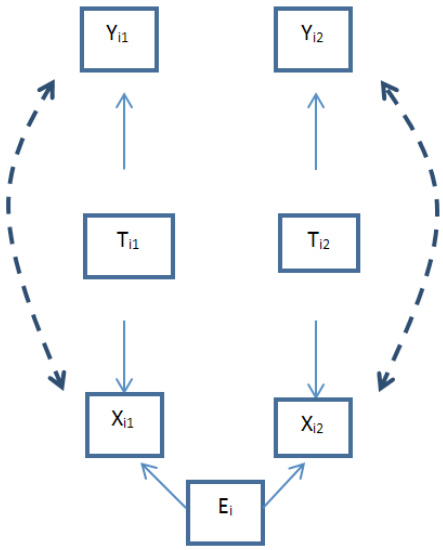
Table 3 – Means ± Standard Deviation of EGFR Serum Albumin and BUN Broken Down by HIV Status Across all Repeated Measures Used in Tables 1 and 2

Variable	For HIV+ Subjects (496 persons 7,326 Replicates)	For HIV- Subjects (178 persons 3,456 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

FIGURES

Figure 1 – Illustration of Common Within Person Measurement Bias and Confounding for K=1

1a. Common Within Person Measurement Bias



1b. Common Within Person Confounding

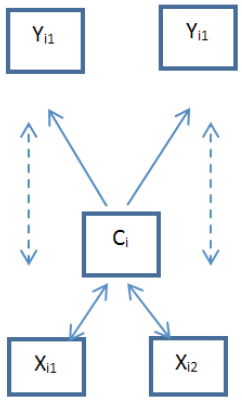
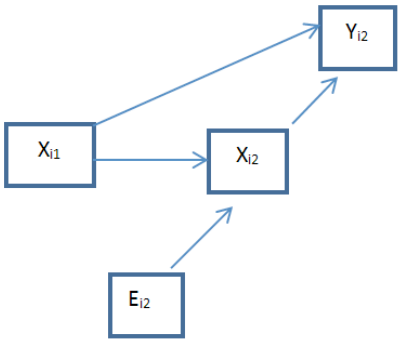


Figure 2 – Illustration of Lag Causality and Reverse Lag Causality for K=1

1a. Lag Causality



1b. Reverse Lag Causality

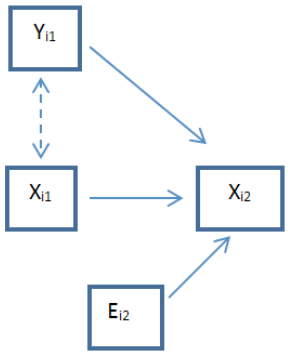


Figure 3 – Illustration of Residual Association With Independent Measure Error in X for K=1

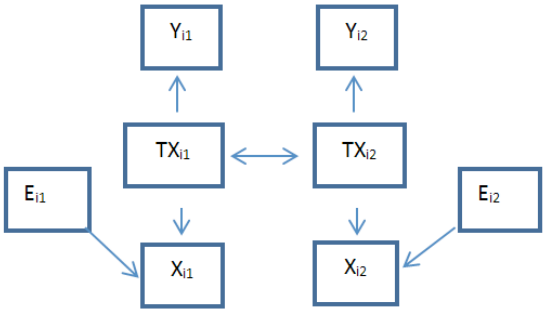


Figure 4 – Compensating Bias Pathways on Time Invariant HIV Estimate From Failure to Use An Independent Working Correlation Structure in Repeated Measures GEE

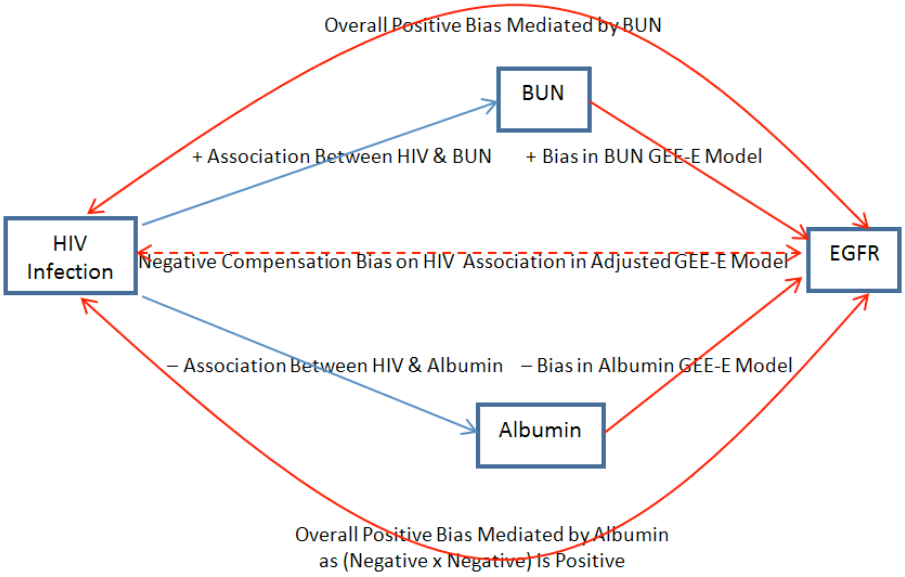
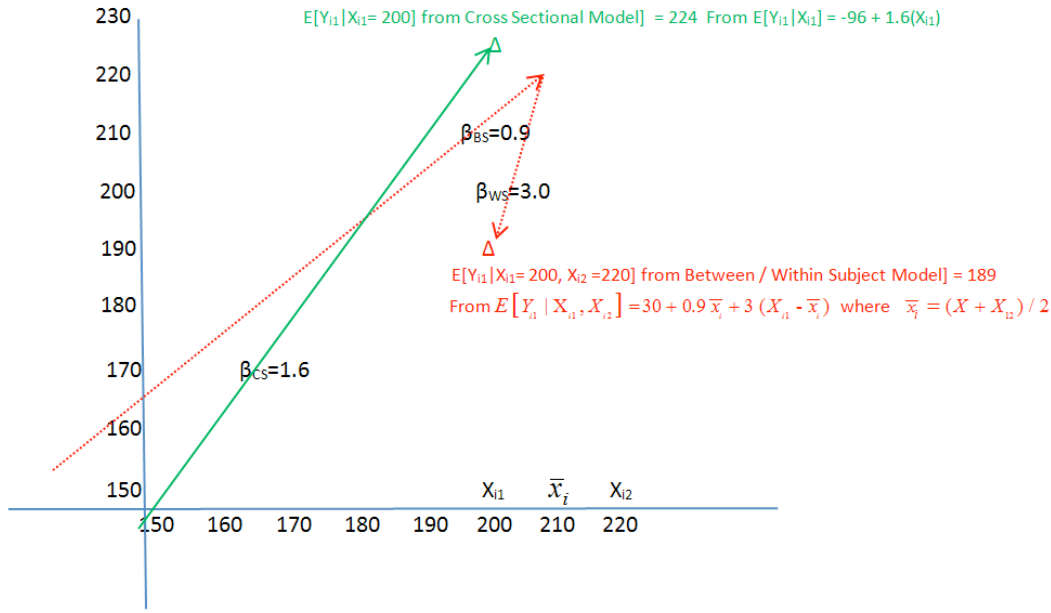


Figure 5- Illustration that $E[Y_{i1}\}X_{i1}]$ is not independent of X_{i2} in Cross Sectional Regression for K=1 when $B_{BS} \neq \beta_{WS}$ in a Between/Within Decomposition Model with $J_i = 2$, X_{i1}=200 and X_{i2}=220



APPENDIX 1– Homology Between Co-DOSE (4) Occurrence and Between and Within Subject Slopes being the Same or Differing

Figure 5 illustrates using the example of Section 2B (with $K = 1$) that Co-DOSE (4)" occurs if $\beta_{BS} \neq \beta_{WS}$. Remember that in this example, β_{BS}=0.9, β_{WS}=3, β_{CS}=1.60. Now let J=2. So for the Between / Within Subject decomposition model; $E\left[Y_{ij}\right]=30+0.9\bar{x}_i+3\,(X_{ij}-\bar{x}_i)$. 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 $E\left[Y_{ij}\right]=-96+1.60(X_{ij})$. 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 $E\left[Y_{ij}\mid X_{i1}\right]=-96\ +\ 1.60\ (200)=\ 224$. But since X_{i2}=220 and $\bar{x}_i$ = 210, as we saw earlier Between/Within Subject decomposition gives; $E\left[Y_{ij}\mid X_{i1},X_{i2}\right]=30+0.9(210)+3(200-210)=189$. Thus E[$Y_{1j}|X_{i1}$] 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 $\beta_{BS} \neq \beta_{WS}$. But if $\beta_{BS} = \beta_{WS}$, X_{i2} is non-informative on $Y_{i1}|X_{i1}$ as E[Y_{ij}] = α_{CS} + β_{CS} X_{i1} = α_{CS} + β_{CS} (X_{i1} - $\bar{x}_i$) + β_{CS} $\bar{x}_i$ = $\alpha_{WS/BS}$ + β_{WS} (X_{i1} - $\bar{x}_i$) + β_{BS} $\bar{x}_i$ since $\beta_{BS} = \beta_{WS}$.

INSERT FIGURE 5 HERE

As J_i≡2, in the prior example, the second observation was deterministic for $\bar{x}_i$. But for J_i > 2, additional $X_{ij'}$ go into computation of $\bar{x}_i$ are thus still informative on relative contributions of $\beta_{BS}\bar{x}_i$ and $+\beta_{WS}(X_{ij}-\bar{x}_i)$ on E[$Y_{ij}|X_{ij}$]. For β_{WS} not well defined, additional $X_{ij'}$ may still be informative on a $\beta_{BS}\bar{x}_i+\beta^{\&}_{WS}(X_{ij}-\bar{x}_i)$ decomposition with $\beta^{\&}_{WS}=E[\beta_{WS}]$.

Whether or not Co-DOSE (4) occurs also informs if $\beta_{BS}=\beta_{WS}$. If for a given j, $E[Y_{ij}\mid X_{ij}]$is independent of all other $X_{ij'}$, then $E[Y_{ij}\mid X_{ij}]$is independent of $\bar{x}_i=\sum_{j=1}^{J_i}X_{ij}/J_i$, which only happens if $\beta_{BS}=\beta_{WS}$. But if $E[Y_{ij}\mid X_{ij}]$is not independent of other $X_{ij'}$, then; i) if $Corr(Y_{ij},\bar{x}_i)\mid X_{ij}\neq 0$, then β_{WS}(if well defined) ≠ β_{BS}, ii) otherwise if $Corr(Y_{ij},\bar{x}_i)\mid X_{ij}=0$, then $\beta_{WS}$is not well defined.

***APPEND*IX 2– Illustration that 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 K=1, let there always be the same number of replicates, J, per subject and assume in the absence of Measure Error, $\beta_{WS} = \beta_{BS} = \beta_{CSR}$. For example, let $E\left[Y_{ij}\right]=5\ TX_{ij}\ =\ E\left[Y_{ij}\right]=5\ (TX_{ij} - \overline{tx_i}) + 5(\overline{tx_i})$, for simplicity the intercept is 0. But we observe $X_{ij} = TX_{ij} + M_{ij}$ where M_{ij} is Measure Error with variance σ_M^2 that is independent across all *i's* and *j's* and also from Y_{ij}. 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 across the *j's*. Let σ_C^2 and σ_R^2 be variances of TC_i and TR_{ij}, respectively.

Now Var(X_{ij}) = $(\sigma_C^2 + \sigma_R^2 + \sigma_M^2)$, Cov(X_{ij},Y_{ij}) = $5(\sigma_C^2 + \sigma_R^2)$ so $\beta_{CS} = \dfrac{5(\sigma_C^2 + \sigma_R^2)}{\sigma_C^2 + \sigma_R^2 + \sigma_M^2}$;

Var($\overline{x_i}$) = $(\sigma_C^2 + \sigma_R^2 / J + \sigma_M^2 / \mathrm{J})$, Cov($\overline{x_i}$,Y_{ij}) = $5(\sigma_C^2 + \sigma_R^2 / J)$ so $\beta_{BS} = \dfrac{5(\sigma_C^2 + \sigma_R^2 / J)}{\sigma_C^2 + \sigma_R^2 / J + \sigma_M^2 / J}$;

Var($X_{ij} - \overline{x_i}$)=$[(\mathrm{J}-1)/J](\sigma_R^2 + \sigma_M^2)$, Cov($(X_{ij} - \overline{x_i}), (\mathrm{Y}_{ij} - \overline{y_i})$) = $5\sigma_R^2[(\mathrm{J}-1)/J]$ so $\beta_{WS} = \dfrac{5(\sigma_R^2)}{\sigma_R^2 + \sigma_M^2}$

For example, let σ_C^2= 8, σ_R^2=2 and σ_M^2 = 10 and J=5. The entire variance of X_{ij} is 20 of which half, 10, is from Measure Error, 8 is variation of the central tendency of X between subjects and 2 is variation of X within subject. Then β_{CS} = 5(10)/20 = 2.5, β_{BS} = 5(8.4)/10.4 = 4.03 and β_{WS} = 5(2)/12 = 0.83. Considering that without Measure Error, true Between and Within person slopes both = 5, Measure Error has greatly tempered β_{WS}=0.83 towards 0 while β_{BS}=4.03 is the least tempered. This happens because β_{BS} most fully retains the signal in X, but tempers M through averaging, while β_{WS} more fully retains M while excluding the between subject signal in X.