A Mixed Model Approach with Control for Modeled Baseline in Longitudinal Analysis, with Applications to the Multi-Ethnic Study of Atherosclerosis

Amanda J. Gassett

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2013

Committee:

Lianne Sheppard

Robyn McClelland

Program Authorized to Offer Degree:

Biostatistics

for my family

Table of Contents

1. Introduction and Theoretical Background

Longitudinal data present particular opportunities as well as challenges. These data are clustered for the repeated measures within an experimental unit, thus we cannot rely on independence to give us some of the nice mathematical properties that we are used to in other contexts. Observations are often missing for at least some subjects, and this can complicate the analysis and inference. On the other hand, use of repeated measures usually gives us greater precision in estimating a parameter of interest by reducing some sources of noise, and some parameters are only estimable with repeated measures. The challenge is to find a method that can adequately account for the clustering, handle missing observations, and estimate an effect with precision and minimal bias. In this thesis, we present a mixed model approach to longitudinal analysis that we argue does just that.

This section provides an overview of some common methods of dealing with longitudinal data and lays some groundwork for our modeling approach. We will then present the form of our model and provide interpretations of its terms. We illustrate properties of the model using data from a large, prospective cohort study and use simulated data to refine our understanding of the results we find in the observed data.

There are several broad categories of longitudinal methods: simplified analysis approaches that use ordinary multivariate linear regression or generalized least squares, marginal models, transitional models, mixed models, and non-parametric methods. Non-parametric methods can be useful tools for description and prediction, but for this work we are interested in estimating associations between exposures or risk factors and health outcomes. Marginal models are useful for estimating population-relevant health effects, but require assumptions that are not met in our setting. Extensive discussion of these methods is outside of the scope of this document. The basic concepts relevant to the analytical method we are most interested in will be illustrated through simplified techniques. In this section, we present background on change models, rate of change models, mixed models, and then conclude with a brief overview of the specific model form that is central to this project.

1.1 General Examples

We will commonly refer to several examples of health outcomes to illustrate concepts. Two health outcomes are the primary motivators for the specific analytic method that is the focus of this work: intima-medial thickness (IMT) and coronary artery calcium (CAC). These two outcomes were measured in the Multi-Ethnic Study of Atherosclerosis (MESA), a large prospective cohort study. CAC is considered to be measured precisely, and it tends to accumulate over time rather than fluctuate. CAC will be the primary focus of the thesis after the introduction; the MESA cohort and the CAC data are described in Section 2. IMT is measured with error and tends to increase slightly over time. It will be mentioned in passing when we consider the effect of measurement error, but detailed examples and data descriptions are not provided. We will also refer to a hypothetical outcome for a limited number of illustrations, simply to clarify certain concepts. For our purposes, we consider the hypothetical outcome to increase over time for the average individual and to be modifiable with treatment. A cartoon of this idealized outcome is presented in Appendix A.

1.2 Notation and Terminology

Throughout, we will adhere to the following conventions:

i indexes subjects.
t indexes time or serial observations within a subject.
v_{it} is the quantity of time that has elapsed since baseline for subject i at time t.
y_i is a single outcome measurement or derived outcome for subject i.
x_i is a treatment indicator for subject i, coded as 0 for the reference group and 1 for the treatment group. This can be generalized to a measurement of an exposure or covariate.
Z, X, W, U are matrices of covariates or exposures; they may include constant terms.
β is a parameter of interest; usually a rate, but sometimes a difference.
α is a cross-sectional parameter or overall mean; often an intercept, sometimes the effect of a covariate.

ϵ_i is the error associated with a particular measurement, and usually includes both the error related to the instrument and the within-subject variability at any given time (e.g. an individual's blood pressure varies throughout the day).

a_i is an individual's random intercept, the amount his or her outcome deviates from the population mean.

b_i is an individual's random slope, the amount his or her rate of change over time naturally deviates from the population mean.

An individual-specific intercept, η, and an individual-specific slope, ϕ, could also be estimated from separate, individual-specific ordinary linear regressions:

$$y_{it} = \eta_i + v_{it}\phi_i + \epsilon_{it} \quad (1)$$

ρ is the within-subject correlation of measurements.

λ is the regression coefficient between measurements. That is, subjects measured at t = 0 and t = 1 have this relationship:

$$y_{i1} = \alpha_0 + y_{i0}\lambda + \epsilon_{i1} \quad (2)$$

If individuals do not change over time, or if baseline is not related to change, then the true population parameter, λ, is equal to 1 and the estimate, $\hat{\lambda}$, will be in the ballpark of the R^2 between the two measurements (ρ^2). This is the result of regression to the mean. In situations where the baseline is related to the change over time, the true relationship between the measurements is not 1. For example, if a higher baseline outcome is associated with a higher rate of change, then extra unit of outcome at baseline is associated with more than one unit of outcome at follow-up, and $\lambda > 1$. The estimate of this parameter will likely be higher than ρ^2, which must still be less than or equal to 1. This is the reason that it is important to distinguish the two parameters.

Unless otherwise stated, "efficiency" refers to the ratio of the variance of the parameter of interest in one method compared to the variance of a similar parameter in a competing method.

1.3 Simplified Analyses – Change Models

Change models examine differences in the outcome measured at different times. The background given in this section is based on clinical trials literature. In clinical trials, the study design usually calls for a baseline measurement, followed by randomization to treatment groups, with a follow-up measurement collected after the treatment has taken effect. Referring again to the cartoon in Figure 7, the exposure begins after the baseline visit, and Individual B shows a constant treatment difference over time compared to Individual A. This is the ideal situation for a change model. Change models may also be used in cases where the progression in two groups diverges over time but individuals are only measured twice. In the cartoon, compare Individual C's trajectory to Individual A's trajectory. A difference in slopes will not be well-estimated if the data support is not sufficient[*], so a change analysis might be an appropriate alternative to a method that estimates slopes. There are three main models used to assess change: 1) cross-sectional or "post-treatment", 2) delta, and 3) control for baseline or ANCOVA.

One weakness of all three models is that individuals that were only measured at baseline are completely dropped from all of these analyses. If these observations are missing completely at random (MCAR), then there is no problem, but this is rare in applications. Observations are missing completely at random only if the probability of an observation being missing is completely unrelated to the outcome. Unfortunately, missingness is often related to the health status. Options for handling missing observations in change models include re-weighting the observed individuals' information through inverse probability weighting (IPW) [1] or taking sensitivity analysis approach and/or imputing the missing observations using a variety of assumptions[2]. Any of these approaches could be applied to any one of the three models considered.

1.3.1 Cross-Sectional Analysis

Consider briefly a cross-sectional analysis. In introductions to longitudinal analysis, this is sometimes referred to as a 'post' analysis, as it assumes that the outcome of interest was measured at follow-up in two groups randomized to different treatments. In a randomized,

[*] Slopes can either be estimated through repeated measures of each individual or by leveraging the between-person information from participants with sufficient heterogeneity. Either method requires an assumption of a linear relationship between the outcome and time.

controlled trial (RCT), simple linear regression or a T-test can be used to compare group means. Assuming the treatment and control groups have the same average outcome at baseline, the mean difference between the two groups at follow-up will give us the treatment effect. The model is straightforward:

$$y_i = \alpha + x_i\beta + \epsilon_i \quad (3)$$

where y_i is the outcome measurement for the i^{th} subject at the end of treatment, α is the average outcome in the untreated or unexposed population, x_i is an indicator for the treatment group, and β is the mean difference in outcome between the treated population and the untreated population. In a more general model, X_i might be a matrix that includes additional adjustment variables.

At baseline, we have just one source of variation in an RCT: the individual-to-individual variability. Randomization should give us the same average outcome in both the treatment and control group at baseline, so at follow-up we attribute all the systematic difference to treatment. In observational studies or in unlucky randomization, group means can be very different at baseline. This is one of the major limitations of cross-sectional observation studies.

1.3.2 Delta Analysis

Taking the difference between the follow-up and baseline measures is the most intuitive approach when we have two observations for each subject. This is an obvious way to isolate the treatment effect and improve power by reducing the main source of variability from between participants to within participants.

We use a very similar model to Model (3):

$$y_{i1} - y_{io} = \alpha + x_i\beta + \epsilon_i \quad (4)$$

where y_{i0} is the outcome measurement for the i^{th} subject at baseline, y_{i1} is the outcome measurement at follow-up, x_i is the treatment indicator, α is the mean change in the control (or reference) population, and β is the mean additional change in the treatment (or comparison)

population. If y_{i1} and y_{i0} are fairly correlated, the variance of the ε_i will be less in Model (4) compared to Model (3).

The estimate for the treatment effect, β, is equivalent in expectation for Model (3) and Model (4) for an RCT, but we expect the standard error to be smaller in the latter unless the variability in the measurement error approaches the between-individual variability. The approach is scientifically attractive because the group difference is much less likely to be due to differences at baseline and because the interpretation of β is obvious. Unfortunately, this is not a perfect method. Measurement error, differences at baseline, and departures from linearity in the treatment response can all affect the estimate. In the presence of measurement error, the estimate of β from Model 3 may be slightly biased because the relationship between participants' measurements is attenuated. According to Frison and Pocock,the bias is[3]:

$$-(1-\lambda_{true})(\overline{\sum_{j\ \epsilon\ x=1} y_{j0}} - \overline{\sum_{k\ \epsilon\ x=0} y_{k0}}) \quad (5)$$

where λ_{true} is the association between the subjects' true outcome values. Usually λ_{true} is approximately ρ, the observed correlation between measurements. Highly correlated measurements within subject will lead to small bias[3].

If the treatment effect is stable over time and follow-up is short, focusing directly on the difference between the two measurements makes sense. For many outcomes, individuals can be expected to change slowly in the absence of treatment or exposure, so over a short time frame in an RCT, average within-person differences will be attributable to the treatment. If follow-up times are long, or if length of follow-up is widely variable, or if the exposure effect is cumulative, differences between two measurements could be related to the time between measurements. Suppose a hypothetical outcome tends to increase with age and poor health status. In observational studies, younger, healthier individuals could be followed longer than older, sicker individuals, since the latter are more likely to die or become too frail to continue in the study. The average difference between a healthy individual's first measurement and a measurement taken after ten years of follow-up could be numerically similar to a frail

individual's first measurement and a measurement taken one year later, but the interpretation would be very different. This brings us back to one of our primary questions, "are we interested in change or rate of change?" If we are in a setting where the treatment effect is likely to be stable even though individuals are changing over time, it is reasonable to control for the amount of time that has passed:

$$y_{i1} - y_{io} = \alpha_0 + v_i\beta_1 + x_i\beta + \epsilon_i \quad (6)$$

where v_i is the amount of time between baseline and the observation at t = 1 for individual i. The true value of α_0 should be 0, since this is the parameter for the average difference between two measurements of the same untreated individual taken at t = 0[*]. We attribute β_1 units of the change to each unit of time that has passed, and some consistent amount of the change to the treatment difference. That is, the two populations differ by β at every time point after baseline, assuming the treatment takes effect quickly. The alternative setting where rate of change is of interest is discussed later in Section 1.4.

1.3.3. Analysis of Covariance (ANCOVA), or Control for Baseline

An alternative to calculating the changes is to control for the baseline measurement. This approach is often called an analysis of covariance (ANCOVA) for historical reasons[†]. The ANCOVA model is given by:

$$y_{i1} = \alpha + y_{i0}\lambda + x_i\beta + \epsilon_i \quad (7)$$

where α is the average outcome at follow-up in the reference group, λ is the amount of increase at follow-up associated with a one-unit increase in outcome at baseline, and β is the additional difference in an individual in the treatment group compared to an individual in the reference

[*] This statement assumes neither time nor the outcome has been centered. Therefore, the estimate of $\hat{\alpha}_0$ is likely an extrapolation and may differ from 0 by a surprising amount. If all predictors have been centered at their mean, then $\hat{\alpha}_0$ is the mean outcome in the untreated group after an average amount of follow-up time.

[†] ANCOVA refers to a model that is like an ANOVA but that incorporates some continuous covariate. It is not specific to models that control for baseline, even though it seems to be understood as such in the literature. We use the term "ANCOVA" as convenient shorthand.

group with the same baseline outcome. In this equation, x_i is a treatment indicator, but this could be generalized to a matrix of predictors, X_i.

In a study where $\lambda = 1$, we could subtract y_{i0} from both sides of this equation and end up back with Model (4). As in Model (4), the ANCOVA will account for group differences at baseline, but the standard error for the estimate of β will be lower because the variance of the ε_i will be smaller. Of all the most common, simple models, this is the most efficient, where efficiency is defined as var($\hat{\beta}$ from Model (7))/var($\hat{\beta}$ from Model (4)) [3]. At a high level, the reason for the efficiency gain is that the baseline measurement explains some of the variability in the outcome, but the relationship between the two measurements is less constrained than the relationship imposed by Model (4), making the variance of the ε_i smaller. Assuming a data structure similar to the observational data that will be described in Section 2, the efficiency of Model (7) over Model (4) is in the range of 0.65 – 0.8. More generally, the improvement depends on the within-individual correlation and the total number of measurements[3*]. The more correlated the follow-up measures are, the less the gain in efficiency is for Model (7) over Model (4). This implies that the more measurement error there is, the more efficiency is gained by controlling for baseline rather than analyzing the differences. The downside is that in the presence of measurement error, ANCOVA will accumulate bias. This is presented as the expected bias in $\hat{\beta}$ from ANCOVA[3]:

$$(\lambda_{true} - \lambda_{obs})(\overline{\sum_{j \,\epsilon\, x=1} y_{j0}} - \overline{\sum_{k \,\epsilon\, x=0} y_{k0}}) \qquad (8)$$

where λ_{obs} is the observed association between the baseline measurements and the follow-up measurements and the expression on the right is the observed difference between the treatment groups at baseline. Essentially this occurs because λ_{obs} is attenuated compared to λ_{true}. The

[*] In the cited article, Frison and Pocock consider taking multiple measurements before and after treatment and averaging them. This is another way to reduce the variance of the ε_i. We will consider multiple measurements in later sections, but will incorporate them differently. This specific result is cited to "bound our expectations" (figuratively speaking) on efficiency gained by controlling for baseline, so that the reader can refer back to it when reviewing the results.

more measurements per subject one uses to calculate λ_{obs}, the closer the estimate should be to λ_{true} and the less biased our estimate of β will be.

We make the comment here that these statements about bias are based on RCT literature, where the interpretation of coefficients in Models 4 and 7 could be more comparable than the same coefficients from models applied to observational data. β in Model (7) compares two individuals with the same measured baseline, and in some cases the baseline measurement could either be a confounder or a mediator. Then the interpretability of the result is an additional concern, alongside bias and variance.

Yanez et. al.[4] present a more general form of the expected bias in $\hat{\beta}$ from Model (7), which considers that the same covariates, X_i, might predict both baseline and change:

$$(\lambda_{true} + 1)\left(\frac{var(\epsilon_{it})}{var(y_{i0}|X_{i0})}\right)\beta_{baseline} \tag{9}$$

where λ_{true} is the true association between the baseline and follow-up outcome, $var(\epsilon_{it})$ is the measurement error or within-person error, $var(y_{i0}|X_{i0})$ is the total residual variance[*], and $\beta_{baseline}$[†] is the true association between the covariates, X_{i0}, and the baseline outcome:

$$y_{i0} = \alpha_0 + X_{i0}\beta_{baseline} + \epsilon_{i0} \tag{10}$$

So the bias in $\hat{\beta}$ from Model (7) depends on the ratio of the within-person variability to the total residual variance, and also on the extent to which the same covariate is correlated with both the baseline measurement and the follow-up measurement. If measurement error is small, then the ratio of the variance of the errors to the total residual variance will be small and the bias will be small. Note also that if $\beta_{baseline}$ is 0, then the bias in $\hat{\beta}$ from Model (7) is also 0, no matter what the ratio of the error variances is.

[*] I understood this model framework to imply that the error was partitioned into person-specific effects, i.e. random intercepts, and independent errors. Then $var(y_{i0}|X_{i0})$ = var(a_i) + $var(\epsilon_{it})$.

[†] The notation used for $\beta_{baseline}$ in Yanez et. al. [4] is ρ, for readers trying to connect the notation between this thesis and Yanez et. al. [4].

Comments made in Section 1.3.2 regarding variable follow-up times also apply here, and would be addressed with similar modifications to the model.

1.4 Simplified Analyses – Rate of Change Models

Potential reasons that one might be interested in the rate of change rather than the absolute change were mentioned in the previous section. Extensions of analytical methods discussed thus far are covered in Section 1.4.1; here we still assume participants were measured twice. In study designs where the rate of change is of interest, often participants will be measured more than twice. This leads to cases where variable follow-up time can result in a varying number of observations by participant, which will be discussed further in Section 1.4.2.

1.4.1. Change Model Extensions

Both difference models and control for baseline models can contain rate-of-change parameters, rather than just change parameters, and these take the form of interactions between factors or exposures and time. Because these models typically contain main effects for both time and exposure, study designs where participants are only measured twice are not ideally suited for estimating this parameter. If participants are measured more than twice, we will either need a method to account for clustering of participants' observations, which will be introduced in Section 1.5, or we will need to calculate a different summary statistic, which will be covered in Section 1.4.2. Caveats stated in Section 1.3 that were related to missing observations still apply to this section.

A possible parameterization for a difference model that includes a rate parameter is

$$y_{i1} - y_{io} = \alpha_0 + x_i\alpha_1 + v_i\beta_0 + v_i x_i\beta + \epsilon_i \quad (11)$$

Here, β_0 is interpreted as the growth rate in the reference population, and β is the difference between the rate in the reference population and the growth rate in the treated population. The parameters α_0 and α_1 are intercepts without a clear scientific interpretation. Consider $v_i = 0$. The model implies that the difference between two measurements collected at $v_i = 0$ is non-zero

on average. Thus, we might interpret α_0 and α_1 as the expected value of a re-measurement of the baseline outcome in the reference and treatment groups respectively[*].

We note that there are two specific sets of coefficients that cannot both be well estimated if participants are only measured twice with similar follow-up times: α_1 and β. Although this is true, either of these could be estimated as long as we remain aware of the limitations of the data. The choice depends on the preferred interpretation. If the variability in the follow-up times is small, v_i essentially acts as a scaling constant. If the the $x_i\alpha_1$ term is dropped, the resulting β from this version of Model (11) will be consistent with estimating β in Model (4) and dividing the result by the average follow-up time. We will use this fact to produce comparable results between models later on.

Another possible parameterization of the difference model is to use the scaled change as the outcome of interest;

$$\frac{y_{i1} - y_{i0}}{v_i} = \beta_0 + x_i\beta + \epsilon_i \quad (12)$$

where β_0 is the average scaled change in the reference population, and β is the average difference between the scaled change in the treated population and reference population. This model effectively assumes that α_0 and α_1 in Model (11) are both zero.

A possible parameterization for a control for baseline model that includes a rate parameter:

$$y_{i1} = \alpha_0 + y_{i0}\lambda + x_i\alpha_1 + v_i\beta_0 + v_i x_i\beta + \epsilon_i \quad (13)$$

The interpretation for β in Model (13) is as in Model (11), but comparing two individuals with the same measured baseline and thus invoking the same caveats about interpretation. The intercepts α_1 and α_0 are similarly meaningless in Model (13) as in Model (11).

[*] Although these parameters are not meaningful, we estimate them to improve the estimation of the slope parameter of interest,β.

1.4.2. Individual Slopes

The most obvious solution when the outcome of interest is the rate of change is to calculate and model just that for each participant. For two points, the quantity calculated is the scaled change as in Model (12). For participants with more than two observations, we can calculate individual trajectories using Model (1):

$$y_{it} = \eta_i + v_{it}\phi_i + \epsilon_{it} \quad (1)$$

where v_{it} is the quantity of time that has passed from baseline to time t for individual i, η_i is individual i's "true" baseline, and ϕ_i is that individual's average rate of change.

The outcome of interest then becomes the individuals' estimated slopes, the $\hat{\phi}_i$:

$$\hat{\phi}_i = \beta_0 + x_i\beta + \epsilon_i \quad (14)$$

where β_0 is interpreted as the average slope in the control or reference group, x_i is still our treatment indicator, and β is the difference in slope between the reference population and the comparison population.

Fitzmaurice cites this as an 'historical' approach that was popular for a time at the National Institutes of Health and was referred to as the "NIH method"[5]. One of its limitations is that all of the additional time-varying information about the individuals is discarded. If participants were followed over a long period of time, then the values of adjustment factors may change. This information is not included in Model (14). Time-varying information could be incorporated into Model (1) in order to produced adjusted rates of change. An analysis of adjusted slopes could become difficult to interpret, so we would have to be very careful to select the "right" adjusted slopes. It is also important to keep in mind that the slopes can only be adjusted for time-varying covariates, not for covariates that only vary between people. There may also not be enough within-person contrast to estimate many (or any) additional parameters, depending on the variability of the within-person measurements of time-varying covariates and the number of observations per individual.

Depending on how the treatment was applied and what temporal relationship is expected among the observations, the baseline measurement may or may not be included in the estimation of the slopes by Model (1). Baseline measurements could be included in observational settings where individuals do not move between groups or where covariates are relatively stable over time, but might be excluded if the treatment or exposure is expected to move the "intercept" of individuals' trajectories (as in Figure 30 for Individual B, in Appendix A). Whether the measurement is included depends on whether the slope of interest is the slope that was observed over all measurements, or whether the slope of interest is the adjusted slope that is relevant to the rate of change at the time of the measurement. The former may have time-varying information embedded in it that becomes a source of noise. The latter may be difficult to estimate and interpret.

Frison and Pocock describe a few ways to improve the efficiency of analyses based on this outcome, which follow from the ways efficiency was improved from the difference model to the ANCOVA model. The first possibility is to control for the measured baseline, which they refer to as a "SLope-based ANCOVA" or SLANC[6]:

$$\hat{\phi}_i = \beta_0 + y_{i0}\beta_1 + x_i\beta + \epsilon_i \quad (15)$$

In the RCT setting, y_{i0} is a pre-randomization measurement, and $\hat{\phi}_i$ is derived from a number of post-randomization measurements. β_0 is interpreted as the average slope in individuals in the reference group ($x_i = 0$) with an outcome measurement = 0 at baseline.

The rationale here is similar to that given for the ANCOVA outlined in Section 1.3.3: some of the variability in the slopes can be explained by the variability in the baseline measurement, which makes our estimate of β more precise. The asymptotic relative efficiency (ARE) for Model (15) compared to the slopes analysis (14) depends primarily on the correlation between the measurements, and depends on the number of measurements per subject to a lesser extent[6]. The reason for the latter is that both models leverage the same number of measurements, so the precision in the $\hat{\phi}_i$ is consistent between the models.

More importantly, the interpretation of the parameter β has changed to be "the treatment difference between two individuals with the same baseline measurement". If baseline and rate of change are related, then β_1 in Model (15) is not 0 and β in Model (14) is a different parameter from β in Model (15). This consideration may not be important in a clinical trial, where the comparison groups are likely to be similar, but could be important in an observational study where group differences at baseline may be related to the exposure or covariate of interest.

A variation on SLANC is a slope-based ANCOVA that controls for an estimated intercept rather than a measured baseline ("SLope Adjusted for INtercept" or SLAIN). This can be advantageous in cases where the outcome is measured with error because the intercept can be estimated more precisely and accurately than the baseline is measured. Then the relationship between the slopes and the baseline is itself estimated more precisely. Because the estimated intercept should be closer to the 'true' baseline than the measured baseline under these conditions, the estimate for β should also be less biased. Frison and Pocock[6] recommend this method "when a divergence between mean curves seems plausible," rather than a shift to a parallel curve. This makes it ideal when the parameter of interest is a difference in the rate of change, rather than an absolute change[*].

$$\hat{\phi}_i = \beta_0 + \hat{\eta}_i\beta_1 + x_i\beta + \epsilon_i \qquad (16)$$

Here, β_0 is the average slope for an individual with an estimated intercept of zero; this may not be scientifically meaningful if the outcome was not centered and 0 is not a reasonable baseline outcome. β_1 is the difference in slopes between two individuals with a difference in baseline outcome measurements of one unit. If β_1 is not close to zero, this model implies that the rate of change is related to baseline.

As proposed, $\hat{\phi}_i$ is based on a number of post-treatment measurements. Model (16) will be consistent with Model (15) if the reference and treatment groups diverge from a common intercept, where the measured, pre-randomization baseline would be consistent with an intercept

[*] Assuming we are comfortable with the scientific interpretation of controlling for the measured baseline.

estimated from post-randomization measurements. This would not be the case for Individual B in Figure 7, so under that kind of treatment effect, β in Model (15) and β in Model (16) are different parameters. In either case, these two β's are different parameters from β in Model (14) if baseline (either measured or estimated) and rate of change are related. The question of which parameter to estimate may be separate from the question of bias, particularly if there is reason to believe that either the measured or estimated baseline could be treated as a confounder.

That being said, if the "true" baseline is consistent with this estimated baseline, then Model (16) should be less biased than Model (15) to the extent that bias is attributable to measurement error. Note that all of (14) – (15) use an outcome that is based on multiple measurements, and any of these approaches will counter attenuation due to measurement error better than any of the change models presented in Section 1.3. As in the change models, the variance of ϵ_i can be reduced by adding explanatory terms to and removing constraints from the model. Of all of the methods described, SLAIN is the most efficient in terms of asymptotic variance. With correlations in the "plausible range" between 0.5 and 0.7, SLAIN is expected to be about 20 percent more efficient than an analysis of the slopes[6], although more generally the efficiency gains are primarily related to the correlation between measurements. As above, since the $\hat{\phi}_i$ are based on the same number of observations per individual in both Model (14) and Model (16), the relative efficiency depends less on the number of those observations.

There are several limitations of SLAIN (16) to keep in mind. If few measurements are collected per individual, relatively many degrees of freedom are required to estimate both a slope and an intercept for each person. If a varying number of measurements is collected from each person, some individual slopes will be more precise than others. It may then be desirable to upweight some individuals by the number of measurements. This could lead to a weighting that does not reflect the underlying population if the number of measurements is informative. However, of the derived analyses, SLAIN is the most efficient in terms of the standard error on the parameter of interest. Bias is minimized by the estimation of the intercept, making it an effective method in the presence of measurement error. Another limitation of both the slopes (14) and SLAIN (16) analysis as proposed is that neither accounts for confounding in the cross-sectional relationship between the outcome and factors at baseline. These factors could be incorporated by adjusting

the slopes for baseline characteristics, but then selecting the 'correct' slope for each participant could be problematic. Although SLAIN controls for baseline, this adjustment is not decomposed into its constituents. The implications of this simplification likely depend on features of the data, but were not discussed in our review of literature or addressed by the current work.

1.5 Mixed Models

The general form of a mixed model is[1,7]:

$$y_{it} = X_{it}\beta + Z_{it}b_i + \epsilon_{it} \tag{17}$$

where the random effects, b_i, are assumed not to depend on the fixed effects, β[1]. The subject-specific random effects are usually modeled as having a multivariate normal distribution, $b_i \sim MVN(0, G)$ and the within-subject errors are assumed to be independent and identically distributed, $\epsilon_i \sim MVN(0, \sigma^2 \mathrm{I_n})$. The observations conditional on the fixed effects have this distribution[1]:

$$Y_i | X_i \sim MVN(X_i\beta, \Sigma_i) \tag{18}$$

where the covariance structure of the ith subject's observations can be specified to depend on a small number of parameters, such as subject-specific intercepts and slopes.

$$\Sigma_i = Z_i G Z_i' + \sigma^2 \mathrm{I_n} \tag{19}$$

In a random intercept model, when $Z_{it} = 1$, the variance of the outcome Y_{it} partitions into between-subject and within-subject components:

$$Var(Y_{it}) = \sigma_b^2 + \sigma^2 \tag{20}$$

where σ_b^2 is the variance of the random intercepts and σ^2 is still the variance of the within-subject errors. This type of model can be complicated and computationally-intensive to fit, but has several big advantages over other methods reviewed so far. The subject-specific random

effect structure permits more imbalance in the data without loss of interpretability. That is, we no longer have to assume that missing observations are missing completely at random. Instead, we can assume that they are missing at random (MAR), conditional on the fixed effects[1]. Observations are missing at random if the probability of an observation being missing depends only on adjustment variables that are included in the model. For example, if patients that are sicker are more likely to drop out, but the probability of drop out might be the same among all individuals with the same health status. As long as health status is included in the model, the MAR assumption will be met. Varying follow-up time is less of an issue for this method, both because the assumptions about missingness are relaxed and because the random effects specification (or "dependence model", so called because a distribution is assumed for the random effects) handles the issues of precision associated with varying numbers of observations[1]. As long as drop-out is related to the outcome primarily through covariates included as fixed effects, the conclusions from modeling will generalize to the original population from which the sample was drawn. Another advantage of this method is that, because each observation is modeled separately, time-varying covariates can be used as adjustment variables. On the downside, random effects are assumed to be drawn from a specified distribution and this does impose constraints.

Another feature of the ability to incorporate a complex dependence model is that standard errors can be adjusted by group-specific random effects as well as subject-specific random effects[1]. Often, just subject-specific intercepts are required but slopes may also be included. The subset of mixed models that include group effects as well as individual effects is referred to as "hierarchical" or "multi-level". It can be tricky to specify the right random effects, and special care must be taken where subjects are clustered in space. Because observations that are spatially-clustered may not be independent, it may be necessary to assess the impact of that clustering on fixed effect standard errors. Residual spatial confounding may also be present in these cases, but if the main exposure of interest is also spatial in nature it may be difficult or impossible to control for spatial confounding without attenuating the effect of interest. Even trying to incorporate spatial clustering into the dependence model can affect the estimated fixed effect of interest, and there is discussion in the literature regarding the best method of accounting for both the residual confounding and the clustering[8].

A dependence model can be applied to any of the simplified methods that we have discussed so far. Subject-specific random effects can be used in ANCOVA or difference models to handle repeated measures of the subjects. Group-specific effects could be layered onto these models as well, and could also be applicable to models of individuals' slopes.

1.6 A Mixed Model Approach with Adjustment for Estimated Baseline

In the previous sections, we showed that simpler modeling methods that controlled for baseline showed efficiency gains compared to other simplified methods. Increased precision is very important when the expected effect sizes are very small. On the downside, using the measured baseline was shown to be prone to bias. Our hypothesis is that control for the modeled, rather than measured, baseline outcome will be precise and have a small bias compared to other methods. In Section 3, we will present a specific form of a mixed model that adjusts for the modeled baseline outcome, incorporates time-varying information, and easily accommodates varying numbers of observations per subject, even leveraging participants that are only observed at baseline.

2. Data

The model under investigation was originally developed to investigate the health effects of environmental exposures on noisy outcomes. Our scientific goal is to establish the effects of increased environmental exposure on rate of change in outcomes over time. The effects on these outcomes are expected to be small, and the ideal method would be able to account for the kinds of data issues that frequently arise in long-term studies: drop-out and the resulting selection bias, variable follow-up times, variable numbers of measurements, multiple levels of spatial clustering, and individuals' changes over time. For the current analysis, we use an outcome that is actually measured reasonably well, and illustrate the model's properties under favorable circumstances. This outcome is coronary artery calcium (CAC). Through simulation, we illustrate its properties under less favorable circumstances, to approximate properties we might observe in an outcome that is less well-measured. Important examples of outcomes measured with error are intima-medial thickness (IMT) and blood pressure. We will refer to these as examples, but extensive discussions or presentations of IMT or blood pressure data are outside the scope of this project.

2.1. Cohort and Study Description

The Multi-Ethnic Study of Atherosclerosis is a large, multi-center prospective cohort study of individuals free of clinical heart disease at recruitment, with over-sampling for Black, Hispanic, and Chinese participants. Six centers (Wake Forest University, Winston-Salem, NC; Columbia University, NYC, NY; Johns Hopkins University, Baltimore, MD; University of Minnesota, St. Paul, MN; Northwestern University, Chicago, IL; and the University of California Los Angeles) each enrolled a gender-balanced, age-stratified, multi-ethnic group of people to follow over ten years. This included white participants and participants from at least one other racial/ethnic group. The demographics of this group have been fully described elsewhere[9]. Participants in MESA completed a host of clinical visits and questionnaires. Visits occurred five times over approximately 10 years: the baseline visit is referred to as Exam 1, Exam 2 occurred after an average of 1.6 years, Exam 3 after an average of 2.8 years, Exam 4 after an average of 4.8 years, and Exam 5 after an average of 9.4 years. They were tracked for health outcomes via annual phone calls, and most participants also returned for in-person examinations after approximately one, three, five, and ten years. Of particular interest were the detection of subclinical

atherosclerosis and the development of clinical heart disease. Markers of subclinical atherosclerosis included CAC and IMT. All participants were measured for both outcomes at Exam 1. About half of participants were measured for both outcomes at Exam 2, and the other half were measured at Exam 3. Concerns over cost and participant burden prohibited examination for all participants at Exams 4 and 5. At Exam 4, a subset of 1,349 participants was selected for CT scans; participants without scans at Exam 3 were prioritized for measurement. At Exam 5, a subset of 3,611 participants was selected for CT scans. Participants with scans at Exams 3 and/or 4 were prioritized for measurement at Exam 5. In both cases, other participants were included as necessary to meet recruitment goals and to accommodate clinics' radiology departments' scheduling constraints.

2.2. Coronary Artery Calcium

One useful marker of subclinical cardiovascular disease is coronary artery calcium. This term refers to the amount or extent of calcification in the coronary arteries. CAC can be quantified using computed tomography (CT). A calcium score is usually calculated from the CT data according to the methodology proposed by Agatston[10], and the use of this metric to indicate subclinical disease is well-established[11,12,13,14,15,16]. Progression of coronary artery calcium is itself a predictor of clinical events[17-19], and is predictive above and beyond the cross-sectional information available from a single scan[20,21].

Analysis of risk factors for CAC progression builds on an earlier analysis of CAC and risk factors in the MESA cohort by Kronmal et. al.[22] That analysis of CAC in the MESA cohort covered an average of 2.4 years of follow-up and included two scans per participant. Kronmal et. al. found that increasing age, male sex, white race/ethnicity, hypertension, BMI, diabetes mellitus, glucose, and family history of heart attack were associated with both "incident CAC", defined as an Agatston score > 0 in participants previously free of CAC, and increases in existing CAC in participants with non-zero scores at baseline. Low-density lipoprotein cholesterol, high-density lipoprotein cholesterol and creatinine were associated with incident CAC only [22]. This analysis extends the follow-up period to an average of ten years, with up to four total temporally-separated measurements per participant.

One limitation of studies that focus exclusively on the Agatston score is that this method assigns a score of zero to many people with relatively low levels of calcium, obscuring any information that may exist at the low end of the distribution. Liang and colleagues[23] proposed the use of a Spatially-Weighted Calcium Score (SWCS) for epidemiological analyses, which calculates a positive score for all participants, including those with Agatston scores of zero. This method has the advantage of being able to quantify the progression of subclinical disease in low-risk individuals; it also simplifies the analysis so that it can be conducted using a single model. For two reasons we will not use the SWCS for all participants: 1) inputs required to calculate a valid score are missing for over one-third of the scans at Exams 3 and 5, and 2) other analyses of the SWCS indicate that this method is more sensitive than the Agatston score to changes in scanner technology.

In this analysis, we use an Agatston-based approach that incorporates the SWCS to recover variability in individuals with levels of CAC that would be assigned 0 by the Agatston method. The SWCS was developed to be linearly associated with the Agatston score on the log scale, so the "limit of detection" (LOD) was determined to be 2.5[*].

Specific score decision logic:

1) We use the Agatston score whenever this score is greater than 0.
2) When the Agatston score is 0:
    a. If SWCS is below the Agatston LOD, we use the SWCS scaled to the Agatston scale[†]. This scales SWCS so that the units are comparable to Agatston units.
    b. If the SWCS is greater than the Agatston LOD, we use the LOD. The reason for this is that if Agatston is 0 we have strong evidence that there is little calcification, yet if SWCS is high, we have evidence that the individual has relatively high calcification compared to other individuals with a 0 Agatston score. The large discrepancies between the Agatston score and SWCS for a small number of scans is not yet well-understood.

[*] Log(SWCS+1) = 2.5; this is approximately the 90th percentile SWCS in scores where Agatston is 0. This was chosen to maximize the number of participants for whom we can use the information in SWCS but minimize the number of people who have unreasonably high values of SWCS.
[†] This is based on linear regression between log(swcs + 1) and log(agatston + 1), and is exactly: exp(log(swcs+1)*.8473+.5010)-1.

c. If SWCS is missing, we use LOD/$\sqrt{2}$. Dividing by $\sqrt{2}$ is common in exposure assessment literature in cases where a measurement is below an instrument's threshold. If Agatston is 0, we have evidence that the amount of calcification is low. Since 90% of SWCS associated with Agatson 0 scores are below the cut-off chosen, it is likely that the true amount of calcification would be between 0 and the LOD. For an average score in this range, using the cut-off value would over-estimate the score, and using 0 would under-estimate the score. LOD/$\sqrt{2}$ is a reasonable approximation of the expected score.

Mixed model analyses (not presented) for Agatston score alone without SWCS have been consistent with results from this specification of the score. Thus we use this outcome definition without question throughout the remainder of this document. Publications may report Agatston-based results as primary, or additional reader markings and phantom data will be provided by the CT reading center so that SWCS can be used to score scans for which the necessary inputs are currently missing.

CAC score varies between participants over a large range and the distribution of measurements is extremely skewed. Many individuals present with non-detectable CAC on the Agatston scale, and a few other individuals have CAC scores in the thousands. Measurements are sometimes log-transformed so that the mathematical properties of the model are better. However, log-transformation tends to over-emphasize small changes in low-risk individuals. Changes on the native scale are more clinically relevant than percent changes, so our analyses will use the untransformed measurements as a continuous outcome.

2.3. Data Description

In this section, we present descriptive statistics of the cohort demographics and outcome data in subsets of the data defined by the number of measurements per subject, focusing on a subset with more than 2 repeated measures. As mentioned in earlier sections, some parameters are only well-estimated when based on more than two temporally-separated measurements for each individual. Real datasets, including the dataset chosen for the current project, often contain participants that were measured once or twice even if the study design called for several repeated

measures. In MESA, participants were followed over 10 years and a number died or became too sick to continue follow-up. However, there was also a large number of participants that participated in the study for the full 10 years, but were not selected for CT scans at Exam 5.

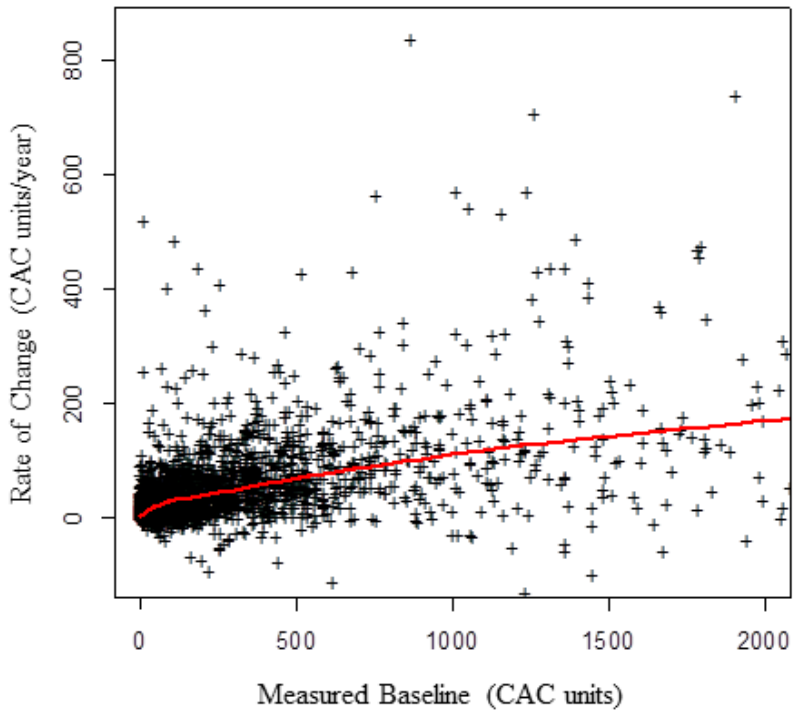
Table 1. Participant demographics, risk factors, and CAC at baseline by number of repeated measures. Continuous variables are shown as mean (+/- standard deviation); categorical variables are shown as count (%). Income is presented as median [IQR] due to its skewness.

	Have 1 CAC Score	Have > 1 CAC Score	Subset, 3-4 Scores
N	**795**	**6019**	**3191**
Time to Last CT Exam (years)	--	6.5 (+/- 3.5)	9.6 (+/- 0.6)
Age	65 (+/-11)	62 (+/-10)	60 (+/-9)
Male	350 (44%)	2863 (48%)	1517 (48%)
Race/Ethnicity			
White	245 (31%)	2378 (40%)	1253 (39%)
Chinese	93 (12%)	711 (12%)	377 (12%)
Black	265 (33%)	1626 (27%)	847 (27%)
Hispanic	192 (24%)	1304 (22%)	714 (22%)
Site			
Winston-Salem	107 (13%)	970 (16%)	543 (17%)
New York City	111 (14%)	991 (16%)	596 (19%)
Baltimore	166 (21%)	920 (15%)	384 (12%)
St. Paul	109 (14%)	957 (16%)	546 (17%)
Chicago	110 (14%)	1054 (18%)	627 (20%)
Los Angeles	192 (24%)	1127 (19%)	495 (16%)
Education & Income			
< High School	234 (29%)	1014 (17%)	443 (14%)
High School	151 (19%)	1085 (18%)	563 (18%)
Some College/Technical	217 (27%)	1720 (29%)	945 (30%)
College or Graduate School	193 (24%)	2200 (37%)	1240 (39%)
Income ($10k)	27 [14-58]	42 [23-74]	48 [26-78]
Health Status			
Total Cholesterol (mg/dl)	194 (+/-39)	194 (+/-35)	195 (+/-35)
Triglycerides (mg/dl)	139 (+/-108)	131 (+/-86)	131 (+/-83)
HDL (mg/dl)	51 (+/-16)	51 (+/-15)	51 (+/-15)
BMI (kg/m^2)	28 (+/-6)	28 (+/-5)	28 (+/-5)
Systolic Blood Pressure (mmHg)	132 (+/-24)	126 (+/-21)	124 (+/-20)
Diastolic Blood Pressure (mmHg)	72 (+/-11)	72 (+/-10)	72 (+/-10)
Statin Use at Baseline	129 (16%)	1065 (18%)	562 (18%)
Hypertensive Medication at Baseline	349 (44%)	2188 (36%)	1110 (35%)
Hypertension Diagnosis at Baseline	421 (53%)	2637 (44%)	1315 (41%)
Family History of Premature CVD	91 (27%)	1586 (27%)	862 (27%)
CAC Score at Baseline	204 (+/-546)	134 (+/-376)	99 (+/-289)

Overall, we observe that the participants with the fewest measurements had higher blood pressure on average, were more likely to be hypertensive, had lower incomes and less education, were older, and had much higher, more variable CAC scores. These participants were most likely in the worst health, although there is also differential retention by site and race that may be partially related to factors other than health status.

One relevant feature of CAC is that the amount of CAC present at baseline is related to the change in CAC over time. Typically, calcification is not removed from the arteries once it forms. Individuals who have high levels of CAC at baseline are those that tend to accumulate additional calcium at a faster rate than those who have low levels of CAC at baseline. Because calcification generally increases, we often refer to the rate of change as the progression rate throughout this document. Figure 1 shows the relationship between CAC at baseline and the rate of change over follow-up.

Figure 1. Association between progression rate of CAC and the amount of CAC present at baseline. Individual slopes or scaled changes were calculated by individual over all available measurements. Smoothing curve was calculated using R's supsmu function, with default parameters.



Since we observe in Figure 1 that individuals with higher levels of calcification tend to increase more than those with low levels, it is possible that individual trajectories over time would look exponential rather than linear. All of the statistical methods that we consider assume that progression is linear over time, so departures from linearity are a cause for concern. Samples of individual trajectories are presented in Figures 2 and 3.

Figure 2. Sample of participant-specific trajectories over time for participants with CAC scores of more than 100 units at baseline.

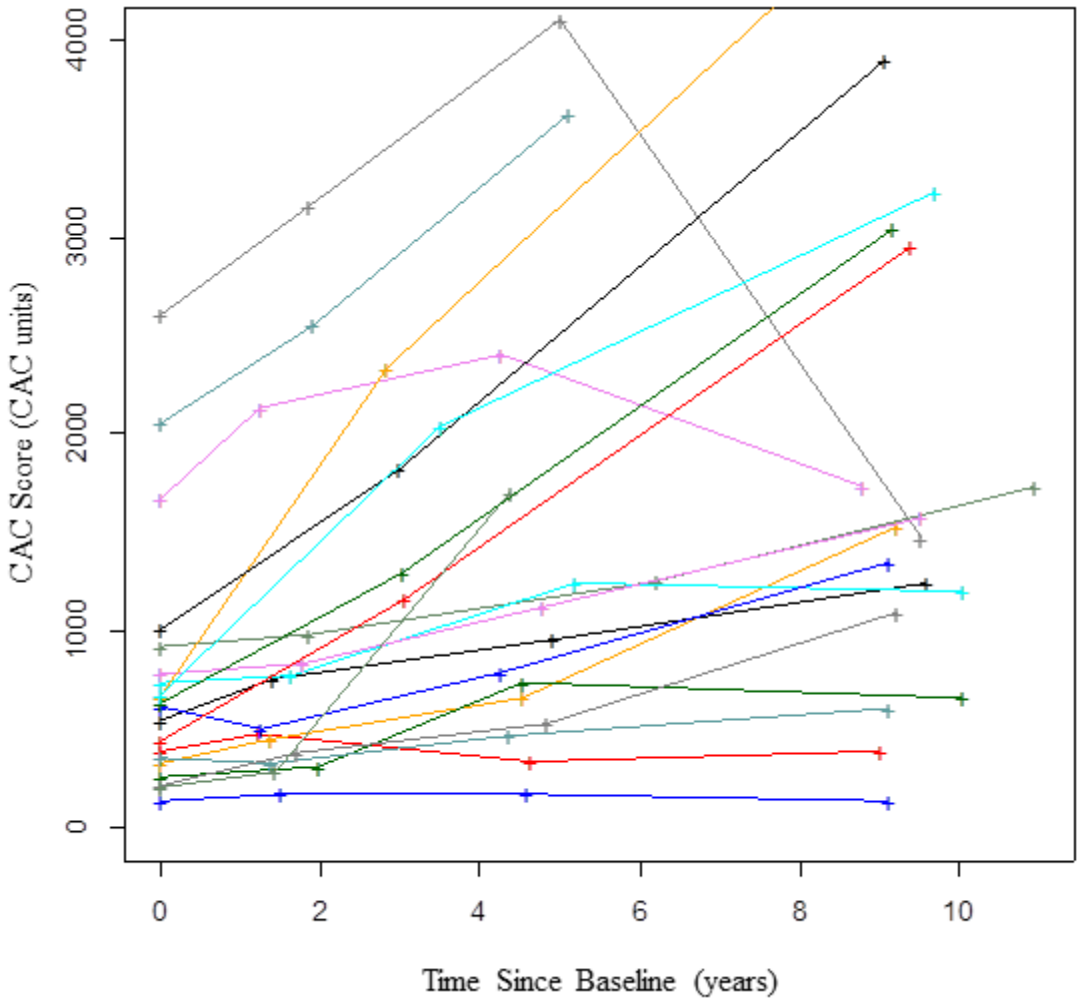
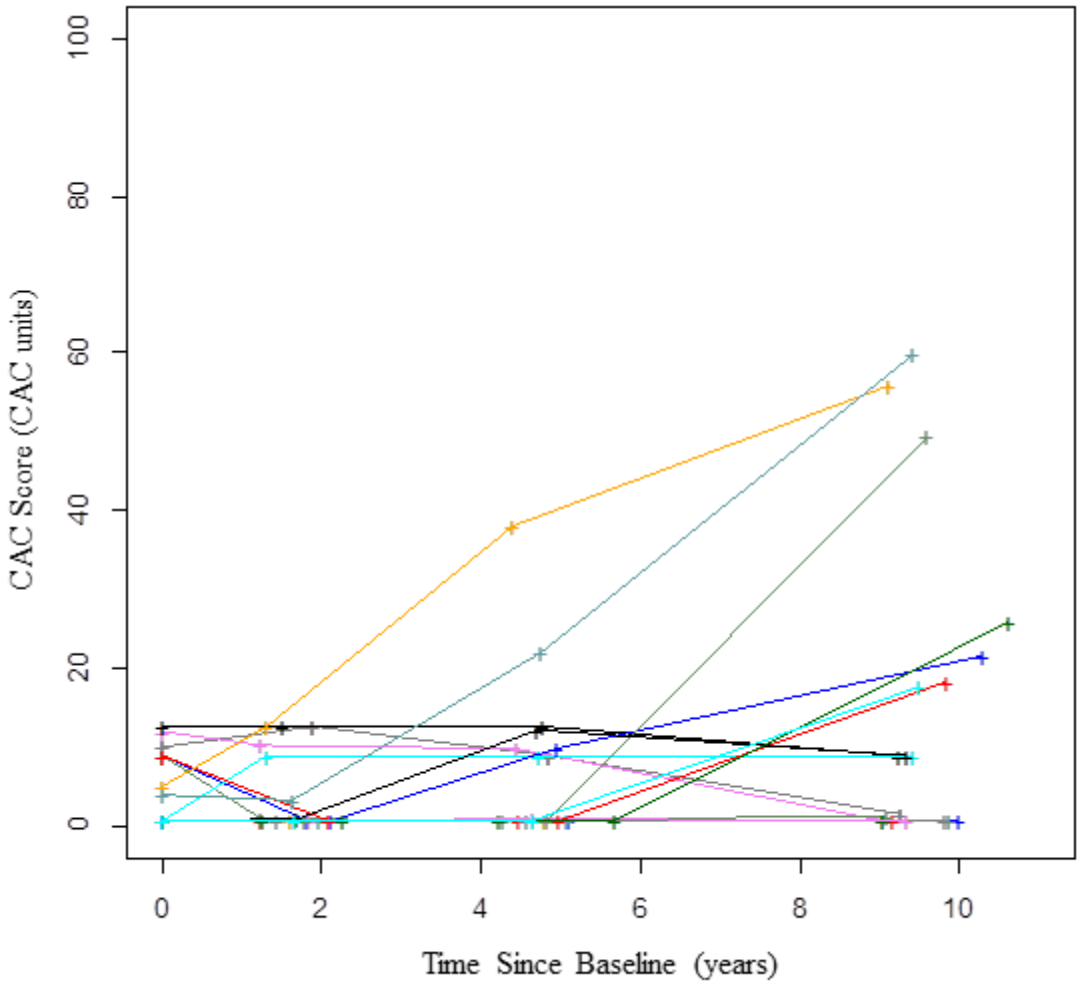


Figure 3. Sample trajectories of individuals with CAC scores less than 100 at baseline.



In Figure 2, we observe that there are some individuals with extremely high progression rates and some individuals that regressed. Individuals that regressed more than 10 units per year are unusual (n = 68), and may include participants whose scans have a data cleaning issue. These could also be participants who had a procedure that removed the part of the artery that showed calcification in a prior scan. In Figure 3, we observe that individuals with lower levels of CAC often keep a consistent CAC score for some time and then jump up. A few participants decrease slightly. In both figures, increases are consistent with linear progression, but with so few observations per individual it would be difficult to estimate non-linear trajectories.

In addition to changes in CAC score, we observe changes in many of the health-related covariates. For example, many individuals were prescribed statins or anti-hypertensive therapy during the course of follow-up. The following plots and figures characterize some of these

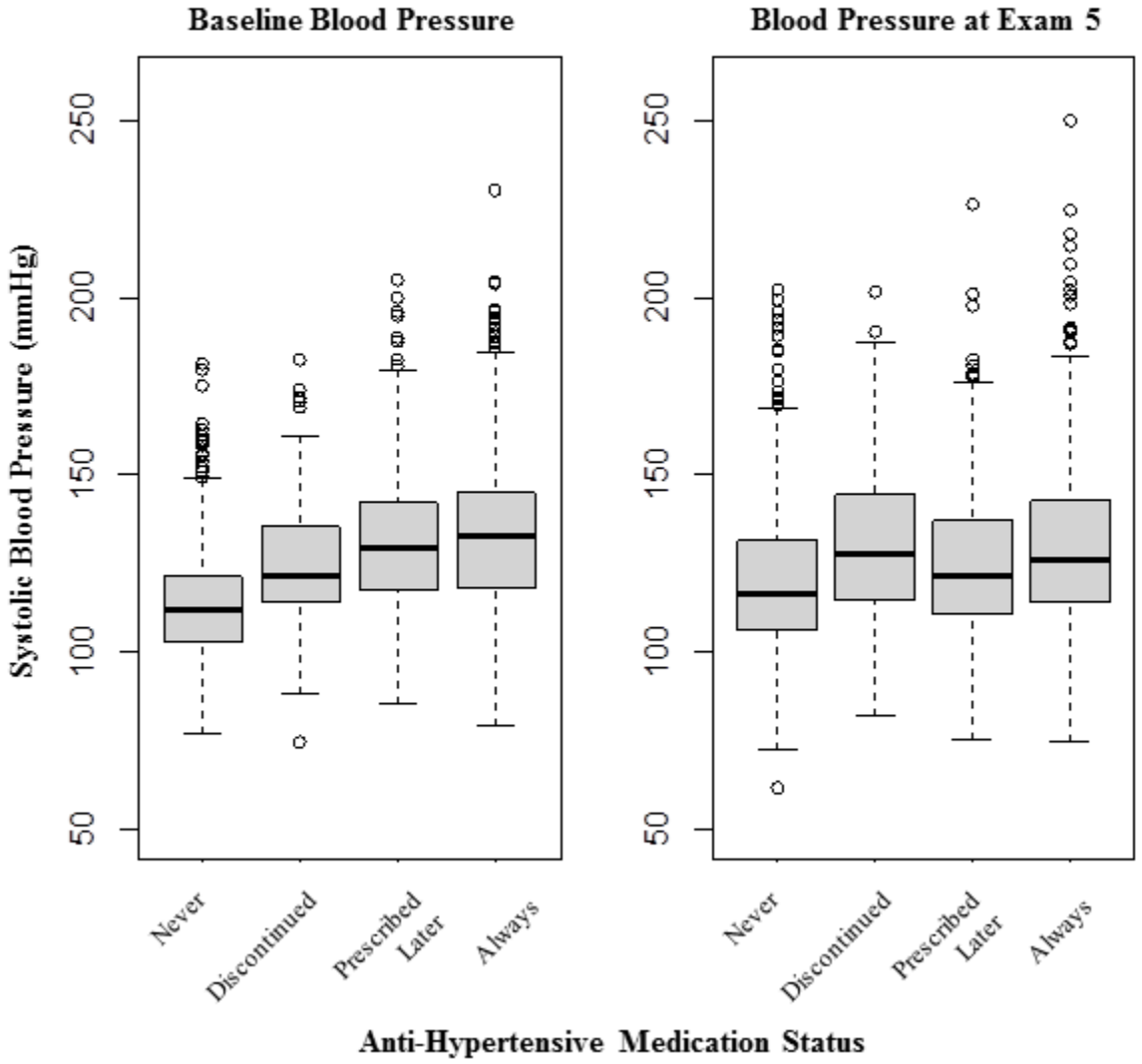
changes. Throughout the rest of this section, we focus on the 3191 participants described in the last column of Table 1 since analyses in later sections will be based on this subset.

Table 2. Participant counts by medication status. Once participants started taking lipid-lowering medications, they were classified as "ever-users". Otherwise they were considered "never users". For anti-hypertensive medication, participants were classified as "current users" or "not current users".

History of Use	Lipid-Lowering Medication	Anti-Hypertensive Medication
Never took medication	1641	1248
Not taking medication at baseline but prescribed later	988	833
Taking medication at baseline, but discontinued	--	144
Always took medication	562	966

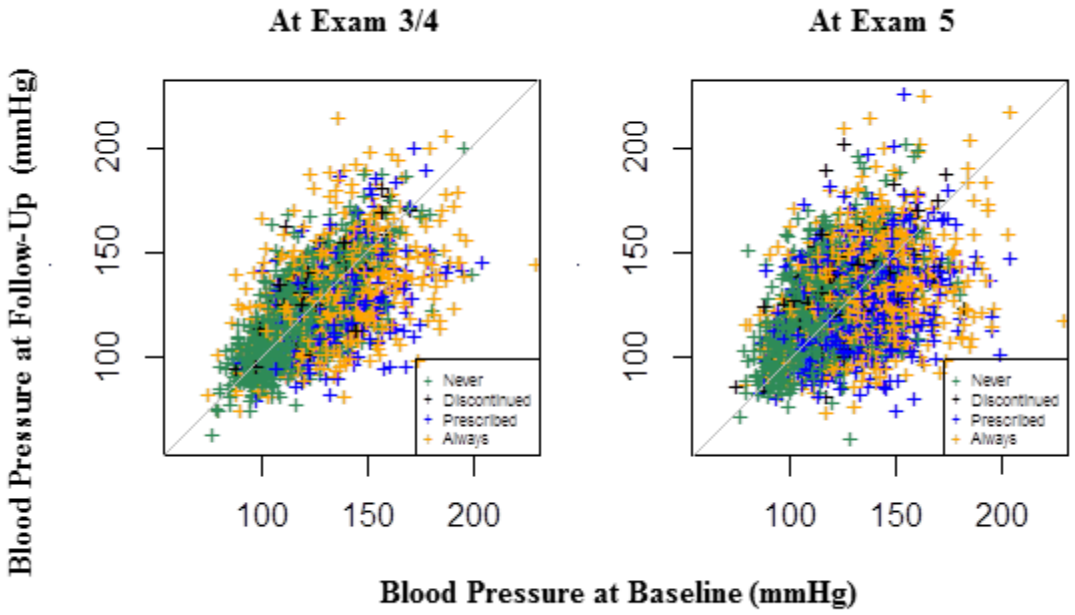
A substantial proportion of the population was prescribed lipid-lowering medications, anti-hypertensive medication, or both during follow-up. Once individuals were prescribed anti-hypertensive medication, they rarely discontinued treatment.

Figure 4. Boxplots of systolic blood pressure at baseline and at Exam 5 by medication status for the subset of 3191 participants described in Table 1, last column. Counts by category are provided in Table 2.



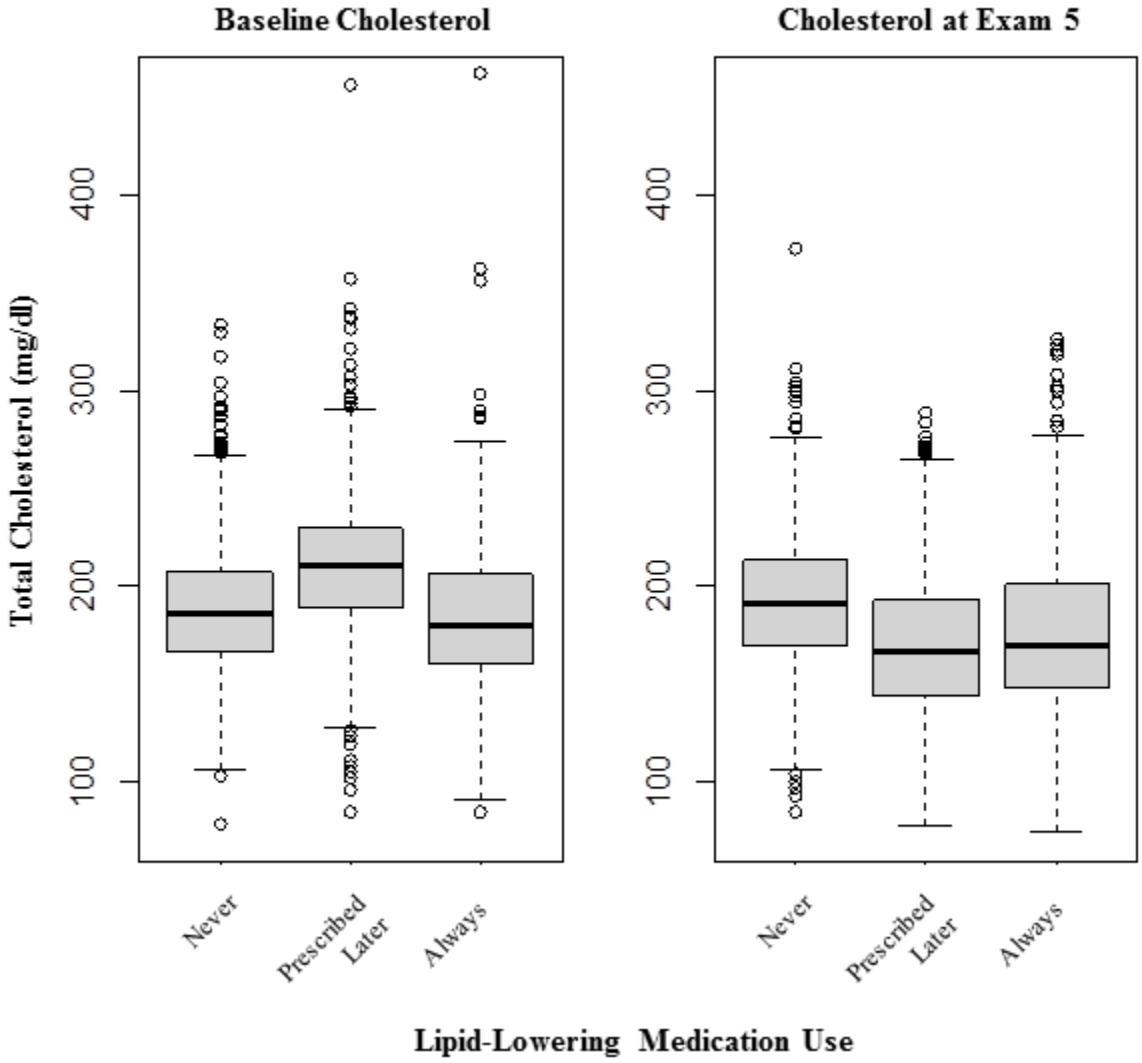
In Figure 4, we observe differences at baseline and at follow-up in measured blood pressure among individuals with different histories of medication use. Participants that started anti-hypertensive therapy prior to baseline had the highest blood pressure at both time points. Participants that were never prescribed therapy had the lowest blood pressure at both time points. Participants that discontinued therapy had higher blood pressure at follow-up than at baseline, consistent with the discontinuation of anti-hypertensive medication. Participants that would be prescribed medication later had high blood pressure at baseline; blood pressure was lower at follow-up.

Figure 5. Correlations between blood pressure at baseline and blood pressure at two follow-up times. The number of participants measured at both Exam 3 and Exam 4 is 29. The correlation between blood pressure at baseline and Exam 3/4 is 0.62. The correlation between blood pressure at baseline and Exam 5 is 0.50. 1-1 line provided in gray.



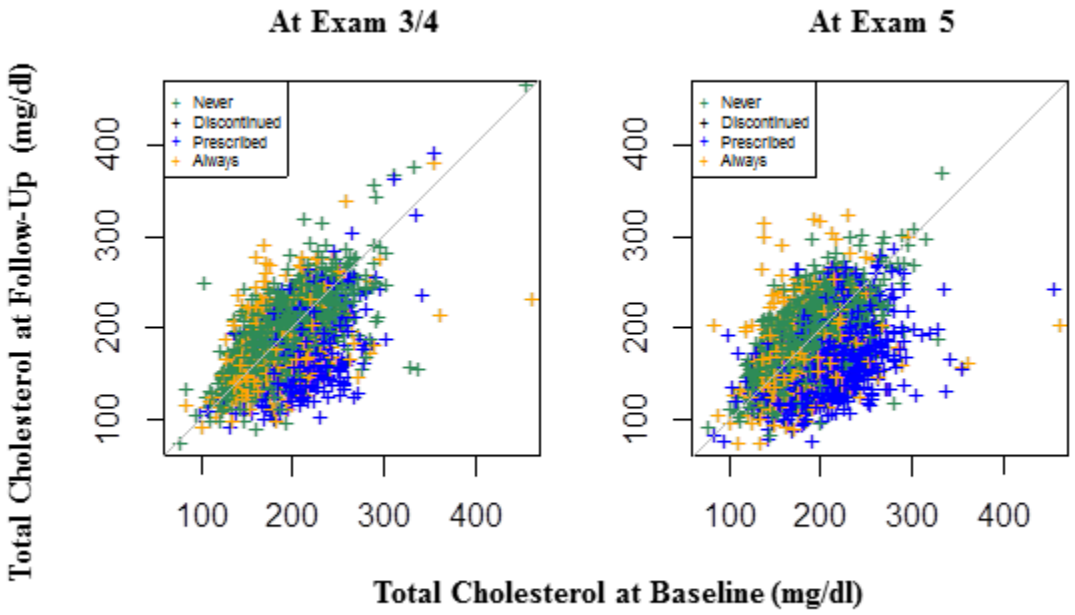
In Figure 5, we observe that correlations between more temporally-distant measurements are lower. The green cluster, which contains participants that were not taking anti-hypertensive medication at either time point, is more concentrated at Exam 5 than at Exam 3/4. Blue points are more predominant at Exam 5 than at Exam 3/4, possibly either because participants at Exam 5 had more opportunities to be prescribed medication by Exam 5 than by Exam 3/4, or because participants were prescribed medication more aggressively during the later time period. Correlations to baseline blood pressure were higher than overall in the never-treated group at both times: 0.72 at Exam 3/4 and 0.59 at Exam 5. Correlations were lowest for participants on medication at follow-up. In the always-treated group, the correlation was 0.49 at Exam 3/4 and 0.33 at Exam 5. In those prescribed medication after baseline, the correlation was 0.45 at Exam 3/4 and 0.33 at Exam 5. In the group that discontinued treatment, the correlation was 0.70 at Exam 3/4 and 0.54 at Exam 5.

Figure 6. Boxplots of total cholesterol at baseline and at follow-up by medication status, among the 3191 participants described in the last column of Table 1.



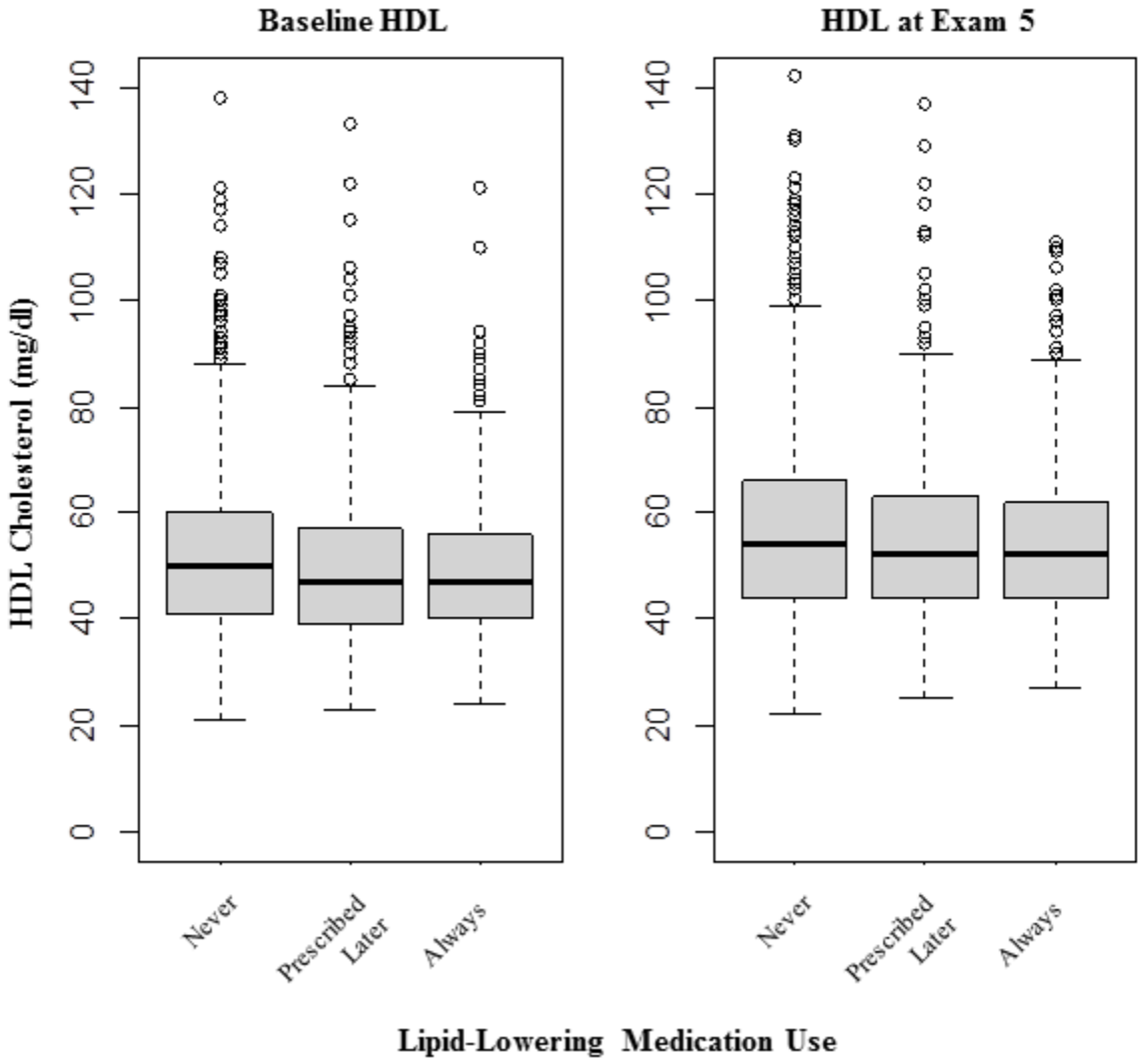
In Figure 6, we observe that medication use affects cholesterol levels. Participants with high cholesterol at baseline were prescribed lipid-lowering medications. These participants had lower cholesterol at follow-up. Participants that were never prescribed lipid-lowering medications had cholesterol levels that increased slightly over time, and had higher cholesterol at follow-up than participants not on medication.

Figure 7. Correlation between total cholesterol at baseline and cholesterol at two follow-up times, among 3191 participants described in the last column of Table 1. The number of participants measured at both Exam 3 and Exam 4 is 29. The correlation between total cholesterol at baseline and cholesterol at Exam 3/4 is 0.59; between total cholesterol at baseline and cholesterol at Exam 5 the correlation is 0.39. 1-1 line provided in gray.



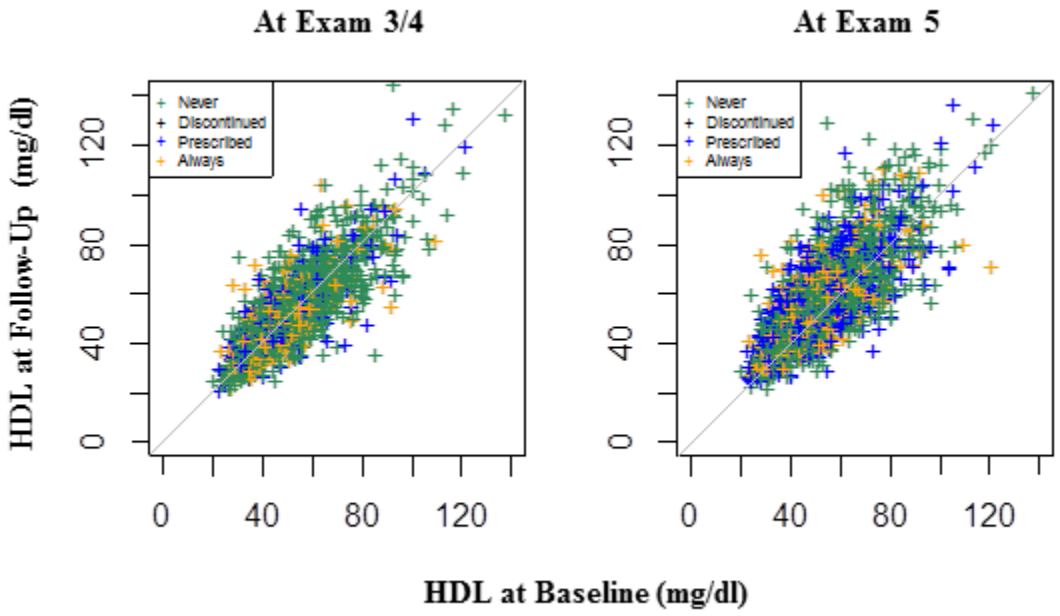
In Figure 7, we observe that cholesterol measurements that are more temporally-distant are less correlated. Blue points, representing participants that were prescribed lipid-lowering medications during follow-up, are more predominant at Exam 5. Many more blue points lie below the 1-1 line than do points of other colors, indicating that treatment was effective for these individuals.

Figure 8. Boxplots of HDL cholesterol at baseline and at Exam 5 by medication use among 3191 participants described in the last column of Table 1. Participants that were prescribed lipid-lowering medications at any time were classified "ever-users"; thus, any participants that may have discontinued treatment are not shown. Counts of participants by classification are provided in Table 2.



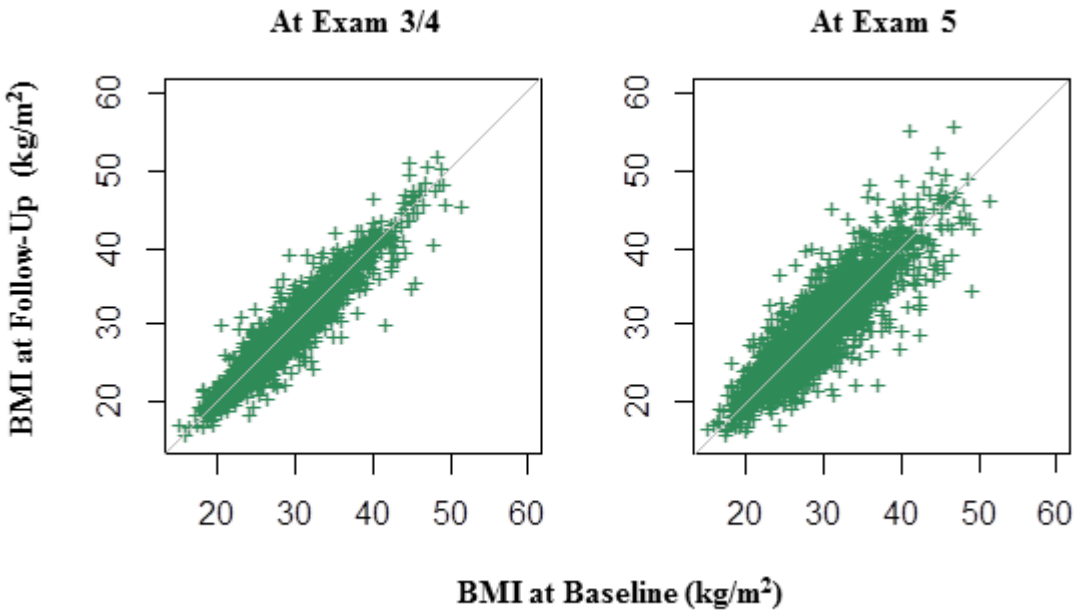
In Figure 8, we observe that the distribution of HDL levels is similar across participants with different histories of medication use. Levels increased slightly in all groups between baseline and Exam 5.

Figure 9. Correlation between HDL at baseline and HDL at two follow-up times, among the 3191 participants described in the last column of Table 1. The number of participants measured at both Exam 3 and Exam 4 is 29. The correlation between baseline HDL and HDL at Exam 3/4 is 0.83; between HDL at baseline and HDL at Exam 5 the correlation is 0.78. 1-1 line provided in gray.



In Figure 9, we observe that measurements of HDL that are more temporally-distant are less correlated. Blue points, indicating participants that were prescribed lipid-lowering medications during the course of follow-up, are more predominant at Exam 5, but no strong clustering is observed.

Figure 10. Correlation of BMI at baseline and BMI at two other time points, among the 3191 participants described in the last column of Table 1. The number of participants measured at both Exam 3 and Exam 4 is 29. The correlation between BMI at baseline and BMI at Exam 3/4 is 0.95; between BMI at baseline and BMI at Exam 5 the correlation is 0.89. 1-1 line provided in gray.



In Figure 10, we observe that measurements that are more temporally-distant are less-correlated, although BMI is a reasonably stable metric over time.

3. Statistical Methods

3.1. Mixed Model Description and Parameter Interpretations

This method controls for the modeled baseline, hence we will refer to it as a Mixed Model that Controls for the Modeled Baseline (MMCMB). A continuous outcome is modeled with time-varying exposures or risk factors in the following model for subject *i*, and time index *t*:

$$Y_{it} = [\alpha_0 + X_{i0}\alpha_1 + a_i] + [v_{it}\beta_0 + W_{it}v_{it}\beta_2 + v_{it}b_i] + [U_{it}\gamma_1 + \epsilon_{it}] \quad (21)$$

where

Y_{it} measurement of CAC for subject *i* in at t^{th} follow-up visit

ε_{it} = measurement error associated with Y_{it}

v_{it} = time of the t^{th} follow-up visit for subject i

X_{i0} = time-invariant cross-sectional covariates assigned at baseline for subject *i*

W_{it} = possibly time-varying longitudinal covariates assigned at visit *t* for subject *i*

U_{it} = time-varying covariates assigned to adjust for factors measured at visit *t* for subject *i*

β = coefficient for covariates in longitudinal relationship

b_i = subject-specific random effect for longitudinal effect (mean 0)

α_1 = coefficient for covariates in cross-sectional relationship

γ_1 = coefficients for covariates in transient relationship

a_i = subject-specific random intercept, i.e., error in implied model

$Y_{i0} = [\alpha_0 + X_{i0}\alpha_1 + a_i]$, where Y_{i0} is the outcome at baseline for subject *i* without measurement error (mean 0)

The terms are divided into three components, distinguished by the square brackets: 1) cross-sectional, 2) longitudinal, and 3) transient.

The cross-sectional terms model the subject's baseline, using both fixed and random effects. This allows for control for confounding in the cross-sectional relationship (through the terms in X_{i0}) and incorporates additional (non-confounding) individual-specific information in the random effects. Furthermore, where the baseline is a mediator, the association between the

outcome and the exposure prior to baseline is modeled jointly with the longitudinal relationship. This is advantageous for the scientific interpretation of the results. As described in Section 1.5, the relative weighting of the estimated fixed effects compared to the within-person information depends on the relative variance of the within-person errors (ϵ_{it}) compared to the between-person variance of the random intercepts (a_i). For example, observations of CAC are more variable between participants than within participants because it is well measured and varies over a large scale. Measurements of an outcome like IMT are not as reproducible, and the within-person variability is close to the variability between individuals in the sample.

We can isolate the cross-sectional portion of the model and show that the fixed effect estimates are consistent with a cross-sectional analysis at baseline:

$$Y_{i0} = \alpha_0 + X_{i0}\alpha_1 + a_i + \epsilon_{i0} \quad (22)$$

If each individual only has one observation at baseline, then we can use an ordinary linear model because the person-specific intercept, a_i, is not necessary or separately estimable. This is the case for CAC, since we have just one score at Exam 1. We could still use a mixed model for the cross-sectional relationship in other settings. For example, the IMT of the common carotid artery was measured on the right and the left in MESA, and blood pressure was measured three times in the Sister Study. When there are reasons to include these measurements separately, a_i would be included to account for the person-level clustering of the observations.

The longitudinal piece models the rate of change in the outcome over time. The exposure or covariate of interest will be included in W, and the parameter of interest is an interaction between the exposure and time. The person-specific slope, b_i, models the individual variation in change over time that is not explained by the fixed effects. Predicted slopes ($\hat{\beta}_0 + W_{it}\hat{\beta}_2 + \hat{b}_i$) from the overall mixed model (21) should be consistent with the slopes that would be estimated by linear regression for each individual's set of observations (Model [1]), with a few differences: 1) they are shrunken towards the fixed information estimated from the overall population, including time-varying information; 2) constraints on the overall distribution of the random effects are imposed by the covariance structure chosen.

The transient part of the model contains parameters for time-varying covariates whose values apply uniquely to the current measurement. For example, blood pressure shows some seasonal fluctuation. A participant with a baseline measurement in the summer and a follow-up measurement in the winter should have the seasonal effect adjusted out. Thus it would be reasonable to include a season variable as part of the matrix U_{it}. For CAC, there is some concern about scanner differences. We may want to adjust our analysis for these differences by including a scanner variable in U_{it}.

Most covariates are centered for this analysis to produce desirable statistical properties; mean-centered design matrices are less affected by model-misspecification. There is also less collinearity and variance inflation in the regression coefficients. However, to preserve the interpretation of the cross-sectional fixed effects as contributing to the predicted intercept, time is left uncentered. In the MESA dataset this baseline time occurs at Exam 1.

3.2. Models for Comparative Analyses of Observed Data

In analyses of the MESA data, we will present results that we observe from more familiar methods that were presented in Section 1 to provide context for the results from the MMCMB. These include delta (similar to Models (11) and (12)), ANCOVA (similar to (13)), and slope-based models (similar to (14) and (15)). We will present "comparable" rate of change coefficients[*] and standard errors for four predictors: male sex, age, HDL, and systolic blood pressure, as estimated in minimally-adjusted models. We focus on an analysis that excludes transient terms. We think of HDL and blood pressure as analogous to a continuous, time-varying exposure. Sex and age are strongly associated with CAC score at baseline, so they are expected to be associated with change or rate of change as well.

There are a few aspects of the MMCMB that we will explore through the data analysis. We have stated that the MMCMB has the ability to incorporate time-varying information, to leverage varying numbers of observations, and to control for confounding in the cross-sectional relationship. In this section, we propose several iterations of model types that have been

[*] Rates of change are not quite comparable in all cases, given the earlier comments in Sections 1.3.3 and 1.4.2 regarding the interpretation of parameters when the analysis adjusts for the measured baseline.

reviewed. We will look at versions of MMCMB using the subset of data described in Table 1, last column: 1) Using all of the observations for each participant and using baseline covariates as the adjustment variable in both the cross-sectional and longitudinal relationship, 2) Restricted to these same participants' Exam 1 and Exam 5 observations only and using the baseline covariates as adjustment variables, 3) Using all observation for each participant and adjusting the longitudinal relationship for the time-varying information, and 4) Restricted to these same participants' Exam 1 and Exam 5 observations only and adjusting the longitudinal relationship for the Exam 5 covariate. We select parameterizations of the simpler models that include parameters similar to those in the MMCMB, and look at results from those models under similar iterations. Models are summarized in Table 3. We will consider staged sets of adjustment variables in an effort to characterize variability in results that may be due to confounding.

Table 3. Specification of competing models for illustrations involving risk factors for CAC progression.

Model Name	Model Form	Model Number
MMCMB (Excluding Transient Terms)	$y_{it} = [\alpha_0 + X_{i0}\alpha_1 + a_i] + [v_{it}\beta_1 + W_{it}v_{it}\beta_2 + b_i v_{it}] + \epsilon_{it}$	(23)
Repeated Difference	$y_{it} - y_{io} = [\alpha_o + a_i] + [v_{it}\beta_1 + W_{it}v_{it}\beta_2 + b_i v_{it}] + \epsilon_{it}$	(24)
Repeated ANCOVA	$y_{it} = [\alpha_o + a_i + y_{io}\lambda] + [v_{it}\beta_1 + W_{it}v_{it}\beta_2 + b_i v_{it}] + \epsilon_{it}$	(25)
Repeated Scaled Change	$\frac{y_{it} - y_{i0}}{v_{it}} = \beta_1 + W_{it}\beta_2 + b_i + \epsilon_{it}$	(26)
Slopes	$\hat{\phi}_i = \beta_1 + W_{i0}\beta_2 + \epsilon_i$	(27)
SLAIN	$\hat{\phi}_i = \beta_1 + \hat{\eta}_i\beta_3 + W_{i0}\beta_2 + \epsilon_i$	(28)

Confounder sets for staged models:

I) Minimal adjustment
   a. Age, sex, race, site
   b. Age, sex, race, site, HDL

c. Age, sex, race, site, systolic blood pressure

II) Moderate adjustment to be used for simulation: age, sex, race, site, BMI, systolic blood pressure, HDL, cholesterol

III) Moderate adjustment, scientifically-motivated

a. HDL, cholesterol, lipid-lowering medication use, BMI, exercise category at baseline, smoking status, second-hand smoke exposure, diabetes

b. Systolic blood pressure, Anti-hypertensive medication, diabetes, BMI, smoking status, second-hand smoke exposure, exercise category at baseline

IV) Full adjustment including socioeconomic status, scientifically-motivated: HDL, cholesterol, lipid-lowering medication use, BMI, exercise category at baseline, smoking status, second-hand smoke exposure, diabetes, systolic blood pressure, anti-hypertensive medication use, income averaged over all times ("permanent" income), education

We define this MMCMB (23) to specifically refer to versions of Model (21) without any transient (U_{it}) terms. The first comparison model (24) is a change model. We use the repeated measurements from all participants so that variance comparisons between models are based on the same underlying dataset and so that time-varying information can be included for some iterations. Next, we include a repeated version of the ANCOVA (25), since controlling for measured baseline and modeled baseline are similar conceptually. Again, we will not be able to control for confounding in the cross-sectional relationship, but assume that including the measured baseline accounts for this relationship. It may also account for unmeasured confounding not explicitly incorporated in MMCMB. For both the difference and ANCOVA models, interactions between time and factors are always included, without cross-sectional terms. For analyses with only two time points, the data do not support estimation of both cross-sectional and longitudinal parameters. For models fit to this dataset, the inclusion of time as given in (24) and (25) will effectively scale the change to be comparable to the estimates from the other parameterizations. For analyses fit to the dataset with all of the time points included, we assume the cross-sectional terms are not necessary because the measured baseline appears in the model. The scaled change (26), slopes (27), and SLAIN (28) analyses incorporate time into the outcome. The repeated scaled change (26) can incorporate the time-varying information into the analysis,

but the slopes (27) and SLAIN (28) analyses cannot do so without adjusting for these covariates during estimation of the outcome variable $\hat{\phi}_i$.

Overall, we hypothesize that results from MMCMB (23) will be more precise than results from simpler models (24 – 28), but that the point estimates will be consistent among the models because CAC is precisely measured.

3.3. Simulation Study

A main motivation for MMCMB is that estimating progression from a modeled baseline should be more precise and less biased than estimating progression from a poorly-measured baseline. The parameter of interest (β_2) also has a different interpretation and is a more appropriate approach to confounder adjustment than directly adjusting for y_{i0}; this direct adjustment is inappropriate when y_{i0} is in the causal pathway. Using simulated data, we can assess the properties of this model in terms of the bias and variance of the estimation of a parameter of interest under different assumptions about the variability in the errors. We will compare results obtained from MMBMB to results from several competing models. Model specifics are provided in Table 4.

Table 4. Specification of competing models for the simulation study.

Model Name	Model Form	Model Number
MMCMB	$y_{it} = [\alpha_0 + X_{i0}\alpha_1 + a_i] + [v_{it}\beta_1 + W_{it}v_{it}\beta_2 + b_i v_{it}] + \epsilon_{it}$	(23 II)
Repeated ANCOVA	$y_{it} = [\alpha_o + a_i + y_{io}\lambda] + [v_{it}\beta_1 + W_{it}v_{it}\beta_2 + b_i v_{it}] + \epsilon_{it}$	(25 II)
Repeated Scaled Change	$\frac{y_{it} - y_{i0}}{v_{it}} = \beta_1 + W_{it}\beta_2 + b_i + \epsilon_{it}$	(26 II)

The confounders from confounder set II are included in the cross-sectional and longitudinal pieces of the above models as follows:

Constant (i.e., not time-varying) covariates included in X_{i0}: age at baseline, sex, race/ethnicity, site, HDL at baseline, systolic blood pressure at baseline, body-mass index at baseline, total cholesterol at baseline

Constant covariates included in W_{it}: age at baseline, sex, race/ethnicity, site

Time-varying covariates included in W_{it}: HDL, BMI, total cholesterol, systolic blood pressure

MMCMB (23 II) specifically refers to the model and confounder set that were used to generate the data for the simulation. The repeated ANCOVA (25 II) mimics the MMCMB (23 II) as closely as possible by replacing the modeled baseline with the measured baseline, although we still estimate α_o and a_i rather than to constrain them to be 0. As mentioned in Section 1, this should produce a better estimate of β_2. Cross-sectional effects of covariates are assumed to be incorporated into the model through control for the measured baseline, so we do not estimate them. We estimate only covariate by time interactions. The repeated scaled change (26 II) only controls for confounding in the longitudinal relationship except to the extent that we have adjusted for cross-sectional effects by subtracting the baseline measurement. Furthermore, in (26 II) all of the parameters estimated are related to slopes. The repeated scaled change is the most plausible simplified alternative that we would consider for a data analysis, which makes it an important competing method.

Data generation and model fitting process:

1) Condition on the data subset described in Table 1, with n = 3191 participants, of whom 2449 have 3 measurements and 742 have 4 measurements.
2) Fit a moderately complex MMCMB (23 II), and take the resulting fixed effect model and parameter estimates to be true. We select the parameter of interest to be the interaction between a time-varying continuous covariate that is relatively stable over time (we chose HDL) and time.
3) Generate a random intercept and slope for each individual, using the multivariate normal distribution indicated by the estimated variance-covariance matrix for random effects from the mixed model fit. These are the "true" subject-specific intercepts and slopes for each individual. Condition on these values for the main simulation study, but check

sensitivity to the random effect model realization by repeating the simulation from this step forward twice more.

4) Add four levels of independent, normal errors to each measurement as a measure of noise. The variance of the error distribution is selected to bring the correlation of within-person measurements (ρ) down to 0.9, 0.7, 0.5, and 0.3.
5) Fit the true MMCMB (23 II) model to these data. Fit simplified models (24 II – 25 II) that are as similar as possible to the MMCMB model.
6) Repeat steps 4-5 1,000 times. Compare the mean effect estimate and 95% confidence interval coverage to the underlying "true" fixed effect of interest. Compare this to the performance of the simplified models.

We hypothesize that the result from MMCMB (23 II) will be more precise and less biased than the results from the simpler models (24 II – 25 II). We hypothesize further that the difference between models will become more pronounced as more noise is added, i.e. results from all models will be more similar when $\rho = 0.9$ than when $\rho = 0.3$.

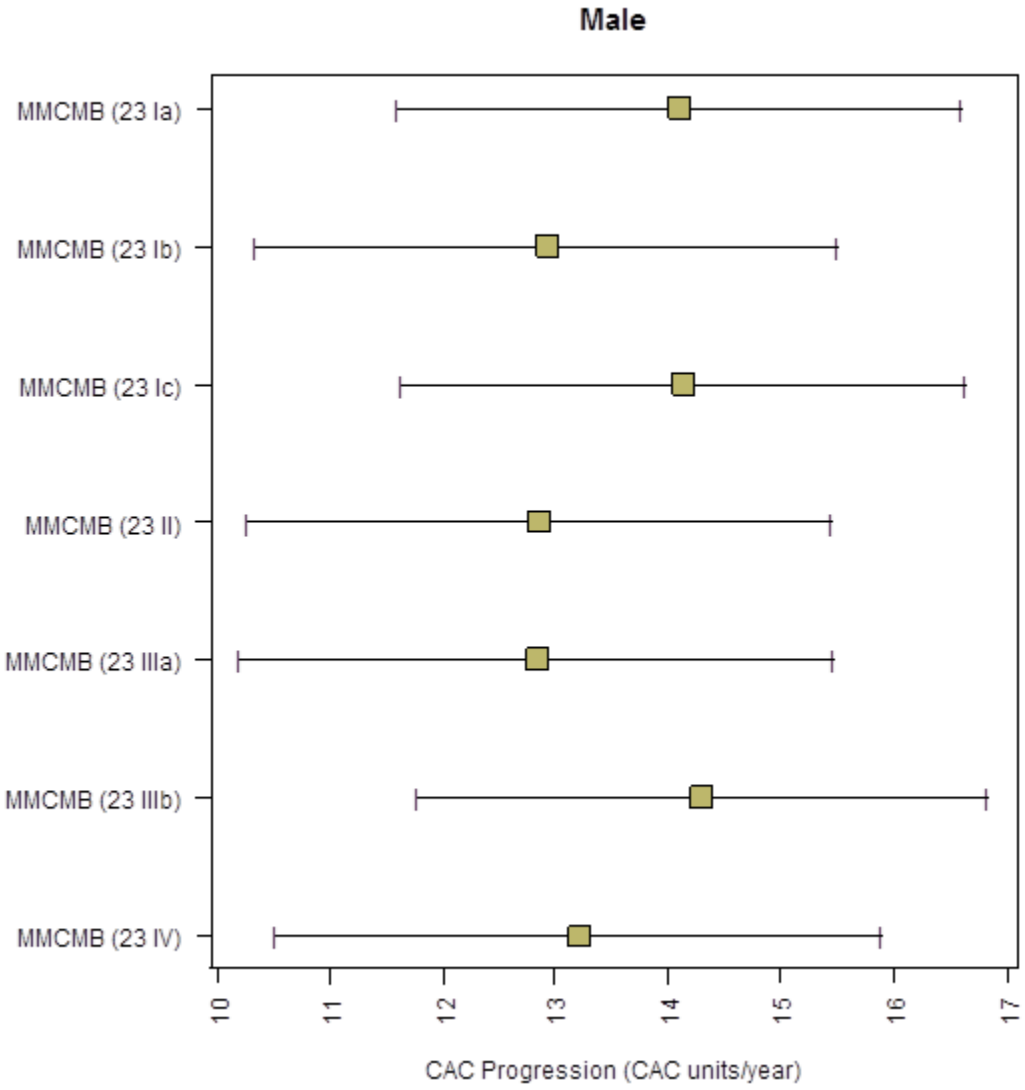
4. Results

The first five sections of this chapter are primarily descriptive, focusing on the properties of parameter estimates as observed in the limited dataset (Table 1, last column). First, we present results for parameters of interest from staged confounder models (Section 4.1). Next, in Section 4.2, we present crude rates of change. The purpose of Section 4.2 is to illustrate that the same average rate of change is estimated by most of the models outlined in Table 3, except for those that control for measured baseline. Then, Section 4.3 presents results for a few parameters of interest, using the various models. Models were described in Section 1 and were specifically summarized in Table 3. In Section 4.4 and 4.5, we present additional illustrations of the mixed model's properties and assessments of model fit. The modeled baseline is illustrated in Section 4.4. Sensitivity to control for confounding in the longitudinal relationship is presented in Section 4.5. Finally, in Section 4.7, we present the results of the simulation study to illustrate the properties of the MMCMB under several error scenarios.

4.1. Staged Confounder Model Results

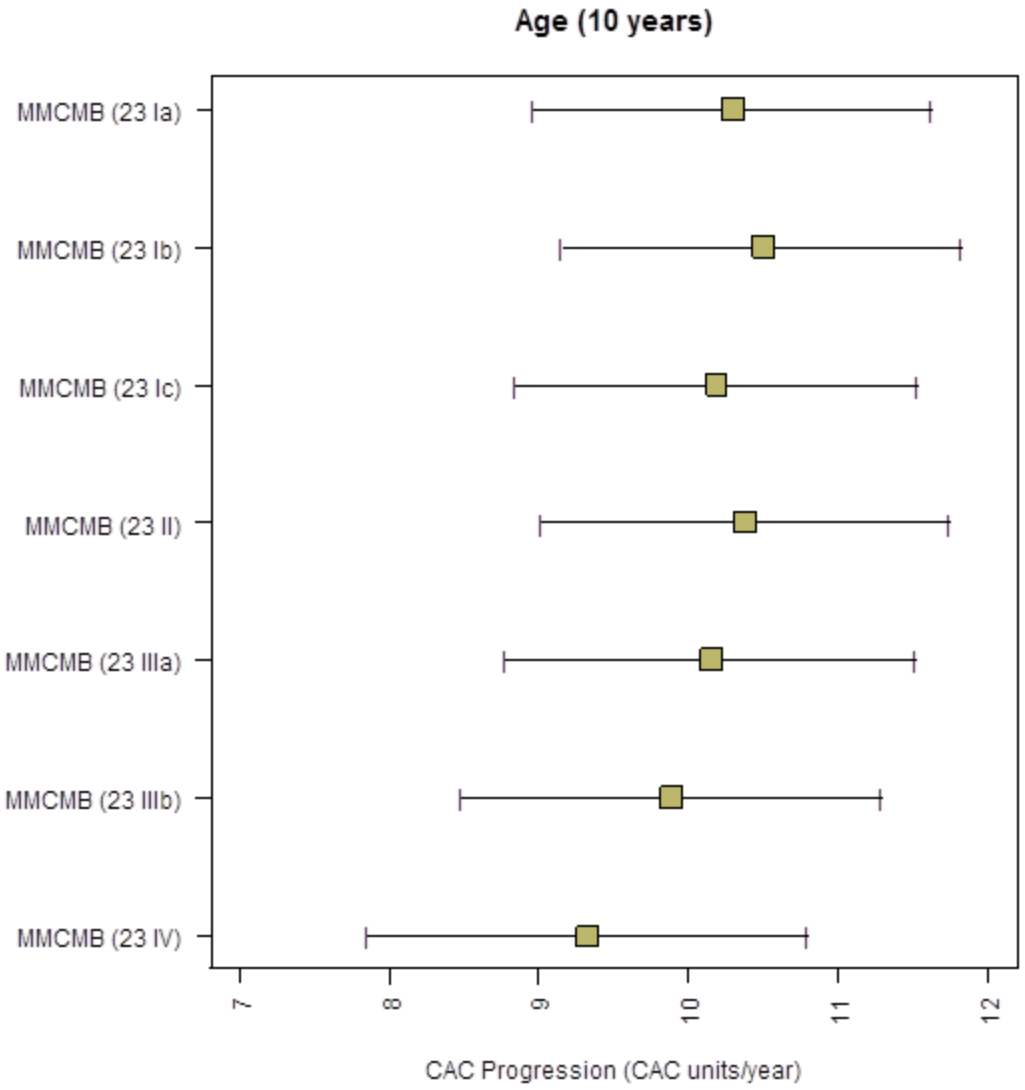
In this section, we present results for parameters of interest from staged confounder models, to orient the reader to the effect sizes in both the longitudinal and cross-sectional relationships (Figures 11 - 18). Confounder staging was presented in Section 3.2. In each of these figures, the baseline value of each confounder was used in X_{i0}. Each confounder was also included in W_{it} with time-varying information included where possible.

Figure 11. Observed longitudinal coefficients (β_2) for the difference between rate of change in males compared to females in MMCMB (23) models with confounder staging. Alpha-numerals alongside the model number correspond to the outline of confounder sets described in Section 3.2. Arms represent 95% confidence intervals.



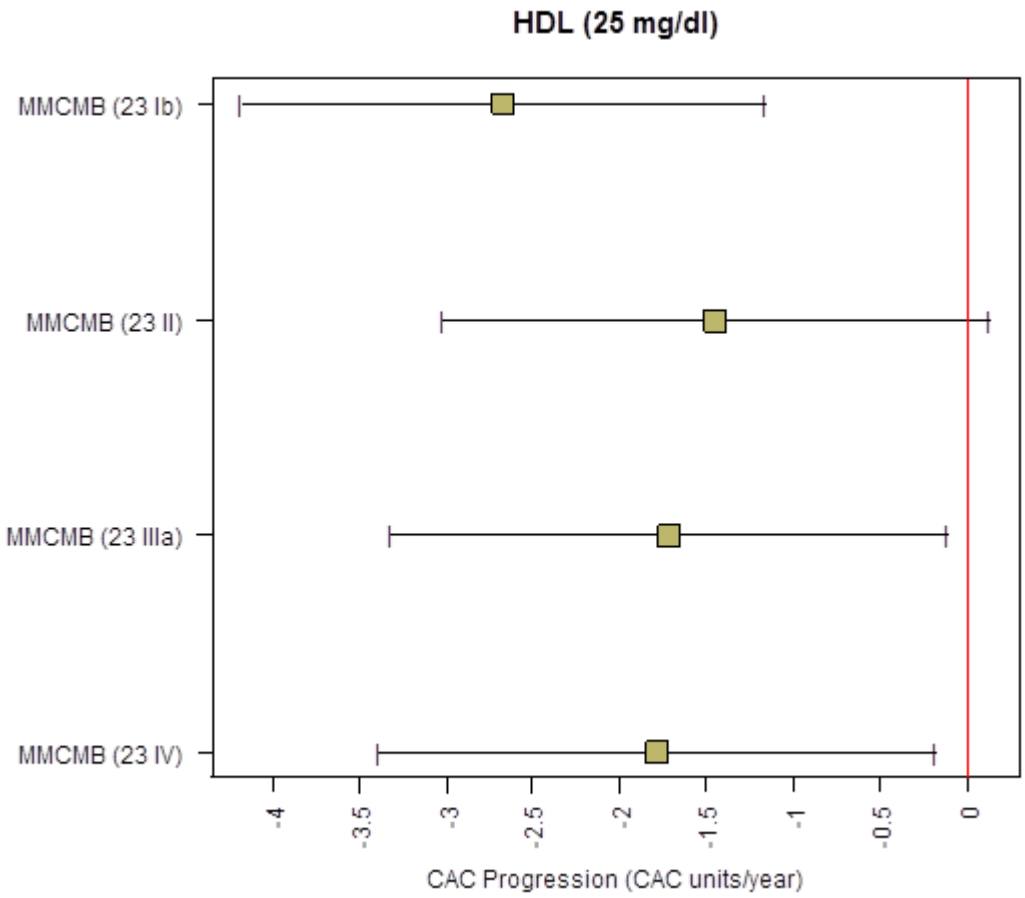
In Figure 11, we observe that males progress faster than females, with some sensitivity to the confounder set used. Confidence intervals are wide in all cases, and all confidence intervals cover coefficients from all models.

Figure 12. Observed longitudinal coefficients (β_2) for the difference between rate of change for individuals 10 years older than other individuals in MMCMB (23) models. Alpha-numerals alongside the model number correspond to the outline of confounder sets described in Section 3.2. Arms represent 95% confidence intervals.



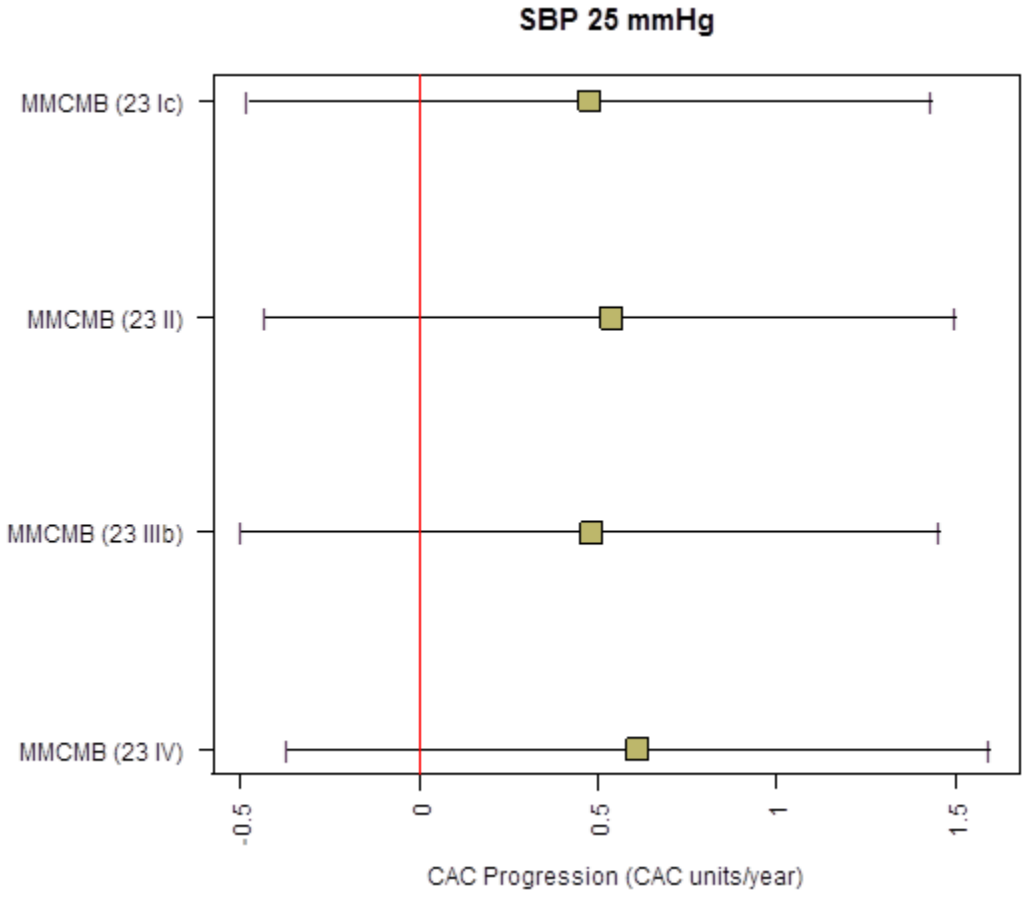
In Figure 12, we observe that older individuals progress faster than younger individuals, with some sensitivity to confounder adjustment. Confidence intervals are wide in all cases, and all confidence intervals cover coefficients from all models.

Figure 13. Observed longitudinal coefficients (β_2) for the difference between rate of change for an increase of 25 mg/dl HDL in MMCMB (23) models with confounder staging. Alpha-numerals alongside the model number correspond to the outline of confounder sets described in Section 3.2. Arms represent 95% confidence intervals.



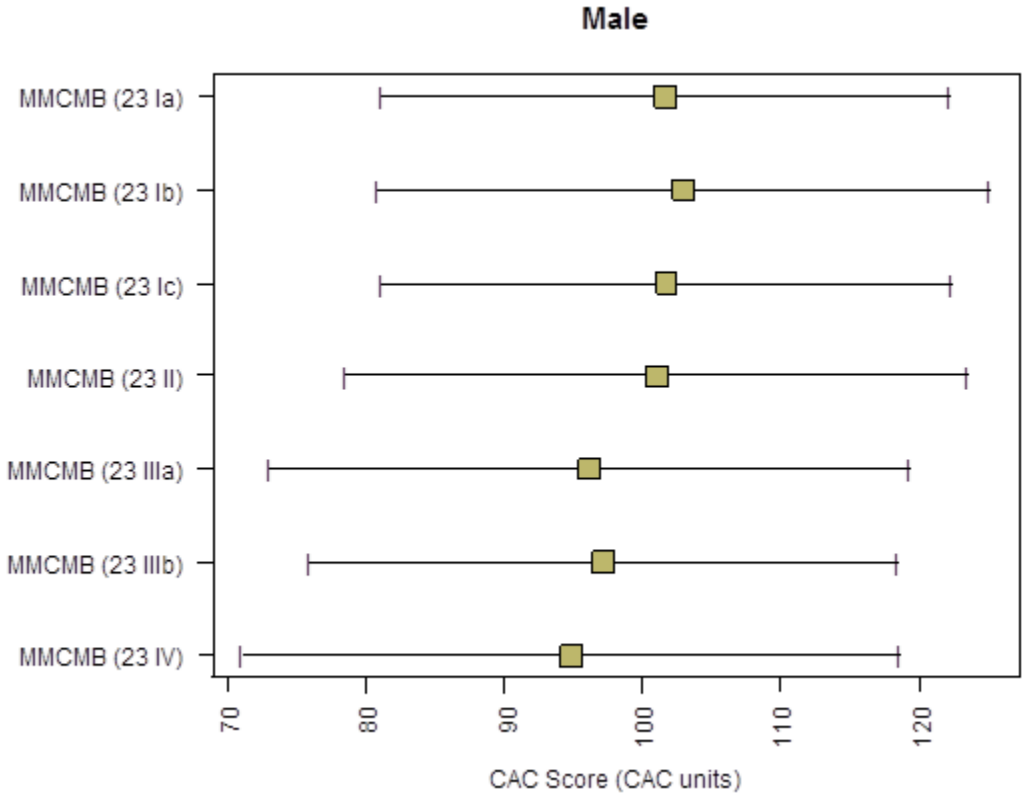
In Figure 13, we observe that participants with higher levels of HDL progress more slowly. The estimated effect is most extreme in the least-adjusted model, but all coefficients are consistent with all 95% confidence intervals.

Figure 14. Observed longitudinal coefficients (β_2) for the difference between rate of change for an increase of 25 mmHg systolic blood pressure in MMCMB (23) models with confounder staging. Alpha-numerals alongside the model number correspond to the outline of confounder sets described in Section 3.2. Arms represent 95% confidence intervals.



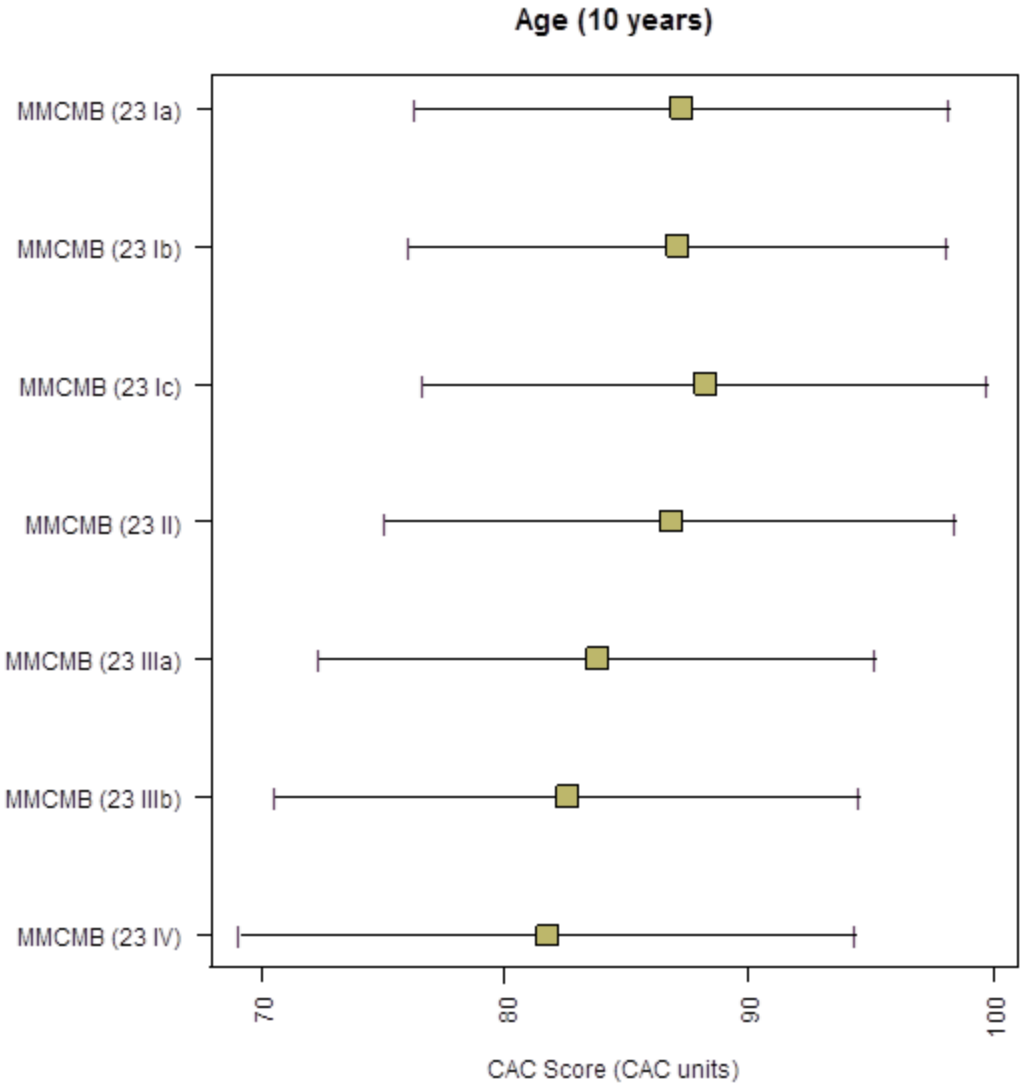
In Figure 14, we observe no evidence that participants with higher blood pressure progress more quickly than individuals with lower blood pressure, even in the minimally-adjusted model.

Figure 15. Observed cross-sectional coefficients (α_1) for the difference at baseline between males and females in MMCMB (23) models with confounder. Alpha-numerals alongside the model number correspond to the outline of confounder sets described in Section 3.2. Arms represent 95% confidence intervals.



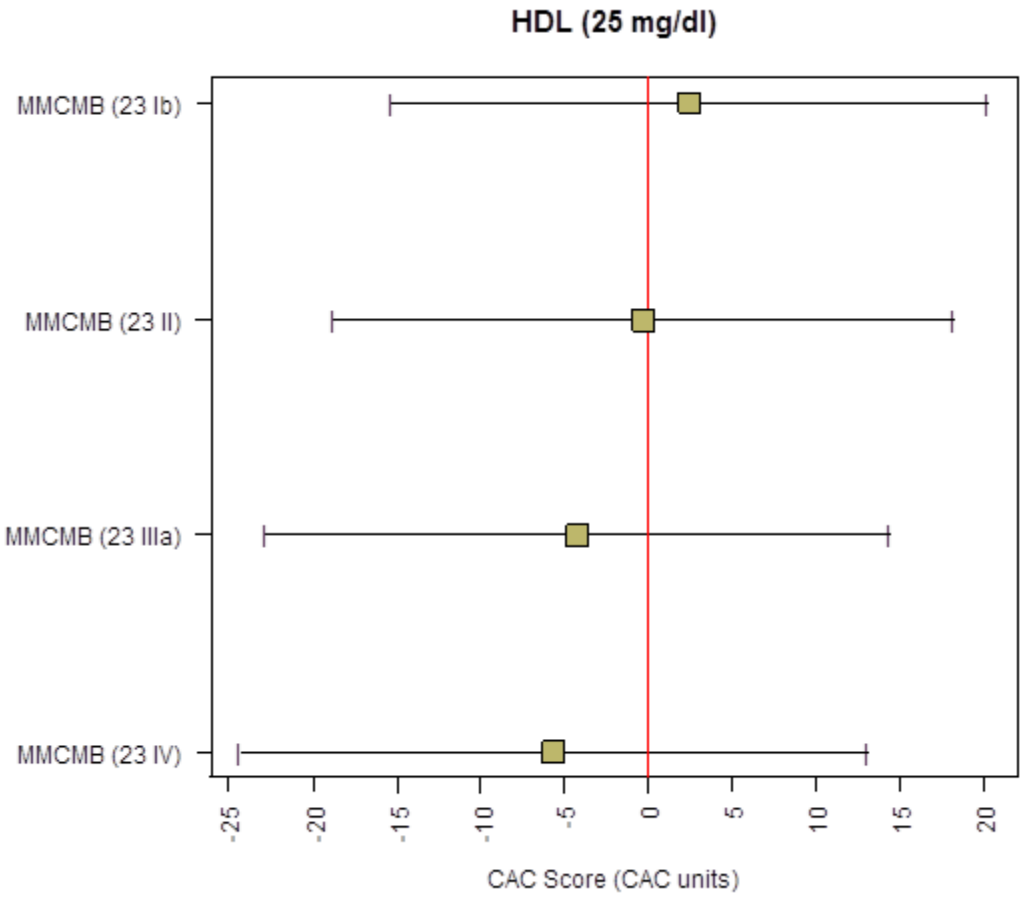
In Figure 15, we observe that sex is strongly related to CAC score at baseline. Coefficients are comparable across models.

Figure 16. Observed cross-sectional coefficients (α_1) for the difference at baseline between individuals 10 years apart in age in MMCMB (23) models with confounder staging. Alpha-numerals alongside the model number correspond to the outline of confounder sets described in Section 3.2. Arms represent 95% confidence intervals.



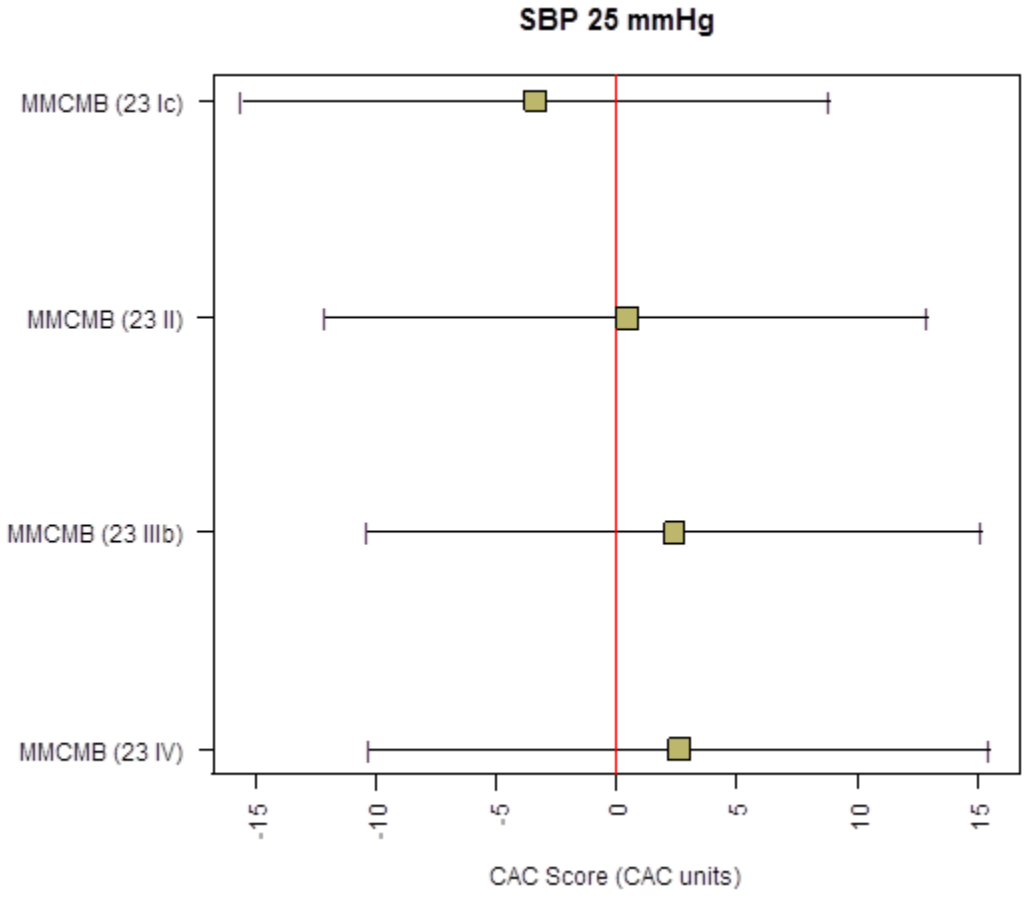
In Figure 16, we observe that age is strongly related to CAC score at baseline. Coefficients are comparable across models.

Figure 17. Observed cross-sectional coefficients (α_1) for the difference at baseline between individuals with 25 mg/dl more HDL in MMCMB (23) models with confounder staging. Alpha-numerals alongside the model number correspond to the outline of confounder sets described in Section 3.2. Arms represent 95% confidence intervals.



In Figure 17, we do not observe evidence for a cross-sectional relationship between HDL and CAC score in any model.

Figure 18. Observed cross-sectional coefficients (α_1) for the difference at baseline between individuals with 25 mmHg higher systolic blood pressure in MMCMB (23) models with confounder staging as outlined in Section 3.2. Arms represent 95% confidence intervals.



In Figure 18, we do not observe a cross-sectional effect of blood pressure at baseline.

In summary, in Figures 11 – 18, we observe that age and sex are associated with both CAC score at baseline and rate of change in CAC score over time. In contrast, we observe that HDL is weakly associated with rate of change of CAC score over time, but not associated with the baseline CAC score. We do not observe evidence for an association between blood pressure and either baseline CAC score or change in CAC. None of these relationships were strongly affected by adjustment for additional confounders.

4.2. Comparisons of Crude Progression Rates

In Table 5, we show that it is possible to estimate the average rate of change using each of the proposed models and that the estimated rate is comparable between models. Referring back to Table 3, we drop X_{i0} and W_{it} to estimate crude rates of change using each method.

Table 5. Crude progression rates as calculated from the models presented in Table 3 without the W_{it} and X_{i0} terms. The coefficient for the overall rate is indicated by β_1 in each of the models.

Model	Crude Rate (CAC Score Units/year)	Standard Error	Variance	Efficiency
MMCMB (23)	19.48	0.67	0.45	Reference
Repeated Difference (24)	19.16	0.71	0.50	1.12
Repeated ANCOVA (25)	19.20	0.71	0.50	1.12
Repeated Scaled Change (26)	19.86	0.77	0.59	1.32
Slopes (27)	19.48	0.67	0.45	1.00
SLAIN[a] (28)	12.96	0.60	0.36	0.80
SLAIN[b] (28)	19.53	--	--	--

[a]Crude rate for an individual with CAC = 0 at baseline.
[b]Crude rate for an average individual, with CAC = 101 at baseline.

Differences between the estimates are minor between the models, except for the parameter estimate from the SLAIN model. The slope identified in the SLAIN model is the average slope for individuals with an estimated intercept of 0. That is, participants with an estimated baseline CAC score of 0, which is within a plausible range of scores but is below average. The average individual has an intercept of 101 units of CAC at baseline. We estimate β_3 from SLAIN Model (28) as 0.063, meaning that for each additional unit of CAC score at baseline, an individual's progression is accelerated by 0.063 units per year. To calculate the average rate of change that is equivalent to that estimated in the other models, we calculate a linear combination of the appropriate parameter estimates. We observe that the estimate from MMCMB (23) is similar to the overall average (27), but not to the estimate from SLAIN (28) for individuals with no CAC at baseline. Although MMCMB (23) models baseline through random effects, these random effects are centered around the sample average, not 0. In other words, although we control for a modeled baseline, the random effects are constrained to have mean 0 around the overall sample

mean rather than 0, and the difference leads to a discrepancy between $\beta_1$estimated from MMCMB vs $\beta_1$estimated from SLAIN.

In Table 6, we illustrate which slopes are estimated when we adjust for sex. This binary indicator separates the sample into two groups that have quite different average baselines, which is an important point when we consider the scientific interpretation of β_2 in the models that control for the measured baseline (ANCOVA [25] and SLAIN [28]). We estimate the average rates for women and men, β_1 and $\beta_1 + \beta_2$ respectively, from the models in Table 3 with sex as the only term in both X_{i0} and W_{it}.

Table 6. Progression rates adjusted only for sex in both the longitudinal relationships (Models 23-28) and in the cross-sectional relationship for MMCMB (23). Average rates in females are estimated as the coefficient β_1 and for males as the linear combination of $\beta_1 + \beta_2$. The mean presented is a weighted average of these two coefficients. Efficiencies are ratios of variances, with MMCMB (23) in the denominator. An efficiency ratio greater than 1 indicates that the estimate is less efficient than MMCMB (23).

Model	β_2 (CAC units/year)	Standard Error	$\beta_1 + \beta_2$ Male	β_1 Female	Weighted Average[a]	Efficiency
MMCB (23)	14.76	1.29	27.23	12.47	19.55	Reference
Repeated Difference (24)	14.76	1.32	26.92	12.16	19.24	1.05
Repeated ANCOVA (25)	10.25	1.33	22.59	12.34	--	1.06
Repeated Scaled Change (26)	15.64	1.48	28.06	12.42	19.93	1.32
Slopes (27)	14.70	1.30	26.15	12.50	19.25	1.02
SLAIN[b] (28)	8.00	1.15	16.78	9.37	--	0.77
SLAIN[c] (28)	--	--	26.82	12.52	19.38	--

[a] Sample is 48% male, 52% female.
[b] Average slopes in individuals with estimated CAC score = 0 at baseline.
[c] Average male and female rates recovered from linear combinations of the effect estimates from Model (28).

Again, we observe that overall rates in females and males separately are similar between models that do not control for a measured baseline (MMCMB [23], repeated difference [24], scaled change [26], and slopes [27]). The weighted average presented in Table 6 demonstrates the similarity between these overall average rates to those presented in Table 5. It is instructive to examine the calculation we make using the SLAIN (28) parameter estimates to come up with an overall average comparable to what we observe as the overall average rate (~19.5 CAC score units/year) in Table 5. Males and females have very different average values of the estimated intercept: about 50 CAC score units for females and about 150 CAC score units for males. Since measured baseline is positively associated with change, the estimated within-person baseline is also positively associated with change. In the SLAIN model (28), we estimate β_3 as 0.063. Therefore, for the male average our linear combination is: 9.37+ 8.00 + 150*0.063. For females, this is 9.37 + 50*0.063. What this means is that even if we center the design matrix for SLAIN (28) so that the parameter for the sex difference in rate is mean-centered, we cannot actually

recover the parameter from this model that is comparable to the type of overall average (as shown in Table 5) that we observe using the other models (MMCMB [23], repeated difference [24], scaled change [26], and slopes [27]). The interpretation is always: the difference in rate for male compared to females with the same baseline. MMCMB (23) does not have the same issue. The conclusion from this illustration is that MMCMB (23) parameter estimates do not have quite the same interpretation as ANCOVA (25) and SLAIN (28), which condition on the measured baseline. One reason for discrepancies between MMCMB (23) and SLAIN (28) is that the MMCMB (23) specified for Table 6 estimates four fixed effects whereas the SLAIN model estimates three. In the SLAIN model, β_3 pools information from men and women, whereas MMCMB (23) allows more flexibility between the groups. Regardless, the result from MMCMB (23) seems more generalizable than either ANCOVA (25) or SLAIN (28).

We observe that MMCMB (23) exhibits comparable precision compared to most of the simpler models (24 – 28) in either Table 5 or Table 6. It is an improvement over the scaled change model (26), but not an improvement over either slope-based model (27 – 28). The slopes (27) and SLAIN (28) models may be overly optimistic since they do not incorporate the uncertainties specific to each intercept and slope, whereas the MMCMB (23) does include the variance contributed by each observation.

4.3. Cross-Sectional and Longitudinal Fixed Effects

This section considers the fixed effects as modeled by the MMCMB (23) compared to competing models described earlier (all specified in Table 3). Table 7 compares estimates for longitudinal parameters as estimated by the various methods. Table 8 compares cross-sectional parameters.

Table 7. Estimates for longitudinal parameters (β_2) with standard errors. Models are as specified in Table 3 and adjusted for covariates in both X_{i0} and W_{it}.

Model	Male[c]	Male, Standard Error	Age[d] (10 years)	Age, Standard Error	HDL[e] (25 mg/dl)	HDL, Standard Error	Systolic BP[f] (25 mmHg)	SBP, Standard Error
Baseline Covariates Only								
MMCMB (23), All Exams	14.10	1.27	10.30	0.68	-4.77	1.20	6.29	0.85
MMCMB (23), Exam 1 and 5	14.19	1.28	10.42	0.68	-4.71	1.21	6.35	0.86
Repeated Difference (24)	14.13	1.28	10.43	0.68	-4.78	1.20	6.29	0.85
Repeated ANCOVA (25)	10.17	1.12	7.54	0.60	-4.51	1.05	5.27	0.75
Repeated Scaled Difference (26)	14.98	1.47	12.00	0.78	-4.85	1.39	6.17	0.98
Difference, Exam 1 and 5 (24)	13.92	1.28	10.24	0.70	-4.62	1.19	6.33	0.85
ANCOVA, Exam 1 and 5 (25)	7.16	1.09	4.56	0.61	-4.16	1.00	4.85	0.72
Scaled Difference (26), Exam 1 and 5	14.07	1.28	10.47	0.68	-4.76	1.21	6.29	0.86
Slopes (27)	14.04	1.28	10.33	0.68	-4.77	1.20	6.29	0.85
SLAIN (28)	8.02	1.13	5.18	0.62	-4.32	1.05	5.05	0.75
Time-Varying Covariates								
MMCMB (23), All Exams	[b]	[b]	[b]	[b]	-2.67	0.77	0.48	0.49
MMCMB (23), Exam 1 and 5	[b]	[b]	[b]	[b]	-3.26	0.96	0.96	0.69
Repeated Difference (24)	[b]	[b]	[b]	[b]	-2.52	0.85	0.73	0.54
Repeated ANCOVA (25)	[b]	[b]	[b]	[b]	-2.91	0.80	0.62	0.53
Repeated Scaled Difference (26)	[b]	[b]	[b]	[b]	-3.50	1.01	1.86	0.64
Difference (24), Exam 1 and 5 [a]	[b]	[b]	[b]	[b]	-3.02	1.04	1.40	0.80
ANCOVA (25), Exam 1 and 5 [a]	[b]	[b]	[b]	[b]	-3.22	0.87	1.10	0.68
Scaled Difference (26), Exam 1 and 5[b]	[b]	[b]	[b]	[b]	-3.18	1.05	1.36	0.81

[a] Exam 5 covariates used
[b] Baseline covariates always used
[c] Adjusted for age, race, and site (confounder set Ia)
[d] Adjusted for sex, race, and site (confounder set Ia)
[e] Adjusted for age, sex, race, and site (confounder set Ib)
[f] Adjusted for age, sex, race, and site (confounder set Ic)

Table 8. Estimates for cross-sectional parameters (α_1) with standard errors. Models are as specified in Table 3 and for covariates in both X_{i0} and W_{it}.

Model	Male[c]	Male, Standard Error	Age[d] (10 years)	Age, Standard Error	HDL[e] (25 mg/dl)	HDL, Standard Error	Systolic BP[f] (25 mmHg)	SBP, Standard Error
From Models Containing Baseline Covariates Only								
MMCMB (23), All Exams	101.71	10.50	87.25	5.60	-7.62	9.92	21.36	7.07
MMCMB (23), Exam 1 and 5	98.99	9.72	82.83	5.18	-7.37	9.18	21.44	6.54
Cross-Sectional Exam 1 (22)	99.15	9.72	82.69	5.18	-7.27	9.19	21.53	6.54
From Models Containing Time-Varying Covariates								
MMCMB (23), All Exams	[a]	[a]	[a]	[a]	2.47	9.05	-3.36	6.21
MMCMB (23), Exam 1 and 5	[a]	[a]	[a]	[a]	0.87	8.37	-3.67	5.58

[a] Baseline covariates always used
[c] Adjusted for age, race, and site (confounder set Ia)
[d] Adjusted for sex, race, and site (confounder set Ia)
[e] Adjusted for age, sex, race, and site (confounder set Ib)
[f] Adjusted for age, sex, race, and site (confounder set Ic)

Our primary interest is in longitudinal relationships. Table 7 presents comparable coefficient estimates from the models presented in Table 3. In the top section, we control for the baseline covariates at all time points. In the bottom section, we control for time-varying covariates where applicable. Coefficients are presented for sex, age, HDL and systolic blood pressure. Estimates are similar between methods where we adjust for baseline covariates. Estimates are also similar between methods where we adjust for time-varying covariates. However, estimates are different when we adjust for baseline covariates compared to when we adjust for time-varying covariates. We explore the possibility that this is due to confounding in Section 4.5. In terms of variance, the repeated scaled difference has the worst performance of the models that provide similar results, although its two-time point counterpart has similar efficiency to the other two-time point models.

The inclusion of the time-varying covariates impacts both the estimate and standard error of these analyses for those factors that do vary over time. In particular, the association observed between CAC and blood pressure depends on which measurements of blood pressure are included. In Section 4.5, we show results from staged models, where we observe attenuation of the association even after control for medication use. We also observe that the precision of MMCMB (23) is similar to the simplified methods if baseline covariates are used for all time points, but that MMCMB (23) is the most precise by a fair margin if the time-varying information is included.

In Table 8, we present cross-sectional associations from MMCMB and from data restricted to Exam 1. Other models presented in Table 7 do not contain cross-sectional parameters and are not included in the table. We observe again that the model specification does not make very much difference in terms of the effect estimate or precision of the covariates that are not time-varying. That is, results from the cross-sectional model (22) are consistent with the results from MMCMB (23). This makes sense; if the baseline covariate is used, then these models leverage the same information to determine the cross-sectional relationship. Results are quite different for the time-varying factors, and the effects are in the unexpected direction, opposite of the cross-sectional (22) results. For HDL, the coefficients are consistent with the result from the cross-

sectional (22) model since the confidence intervals are so wide, but the blood pressure results are very different.

To investigate the possibility that the difference is caused by confounding, we present cross-sectional results in staged model in Section 4.5. We do not see strong evidence that the discrepancy between the baseline covariate-adjusted models and the time-varying-adjusted models is due to confounding. Blood pressure is measured with more error and has a lower within-person correlation than HDL, which may partially explain this inconsistency. As we will show in Section 4.5, we suspect that these inconsistencies are in some way attributable to the time-varying covariates. Possible avenues for further research related to time-varying covariates will be discussed in Section 5.

4.4. Individual Level Intercepts

In this section we compare the intercepts obtained via several methods: the measured baseline value of the outcome, the intercept estimated from only within-person information (that is, the intercepts that would be used for the SLAIN analysis, calculated for individuals from Model [1]), and the intercept predicted by the mixed model, i.e. the "conditional" predictions from Model (23 II). For this comparison, we use results from the fairly rich model that was fit to generate the fixed effect coefficients (with terms as specified for Model [23 II]). We also present the intercepts predicted by the fixed effects ("marginal") portion of the mixed model, to illustrate the population-mean contribution to the intercept variability.

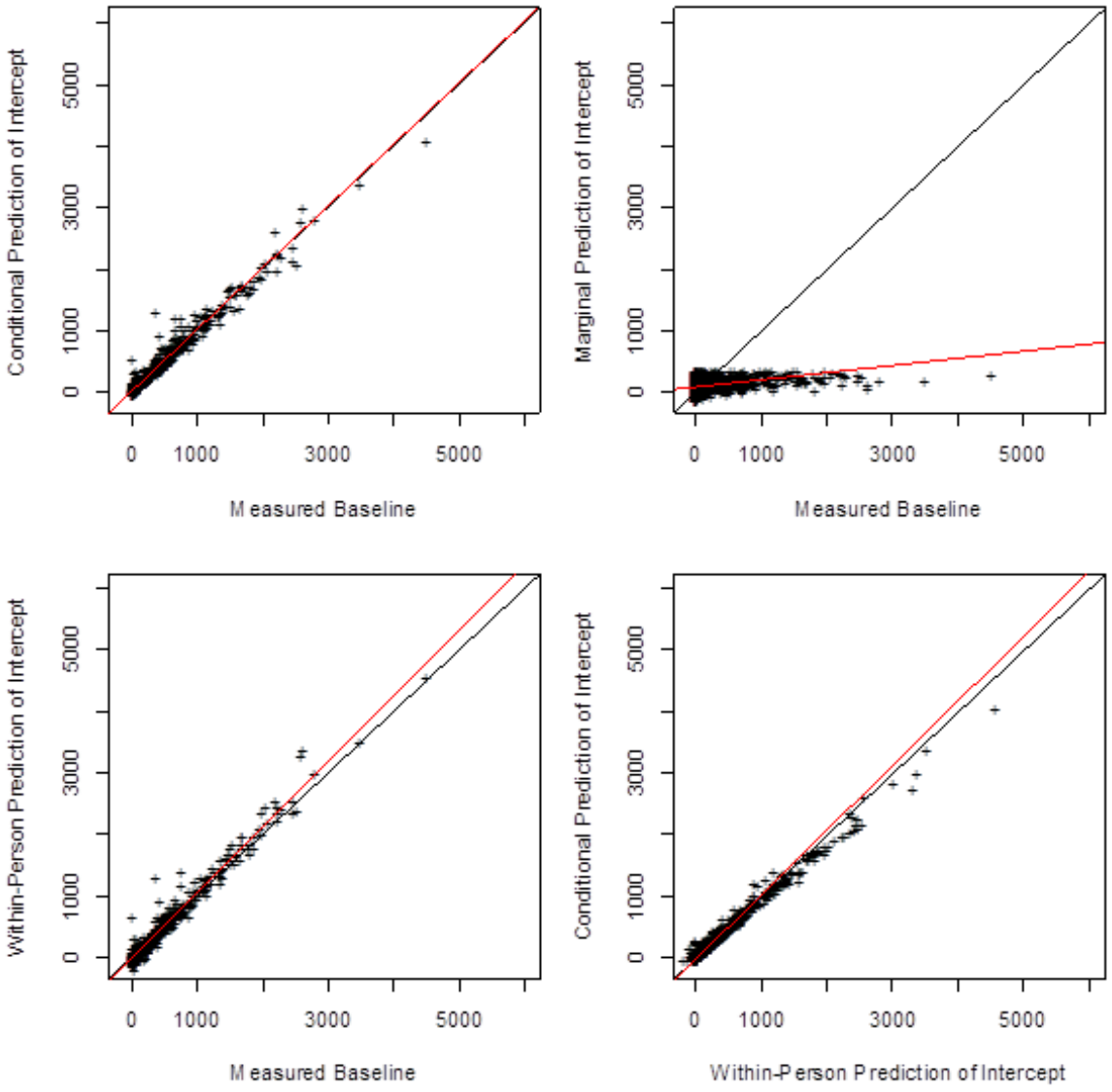
Referring back to the notation used in Section 1.5, marginal predictions are calculated as:

$$\hat{y}_{i0} = X'_{i0}\hat{\beta} \tag{29}$$

and conditional predictions are calculated as:

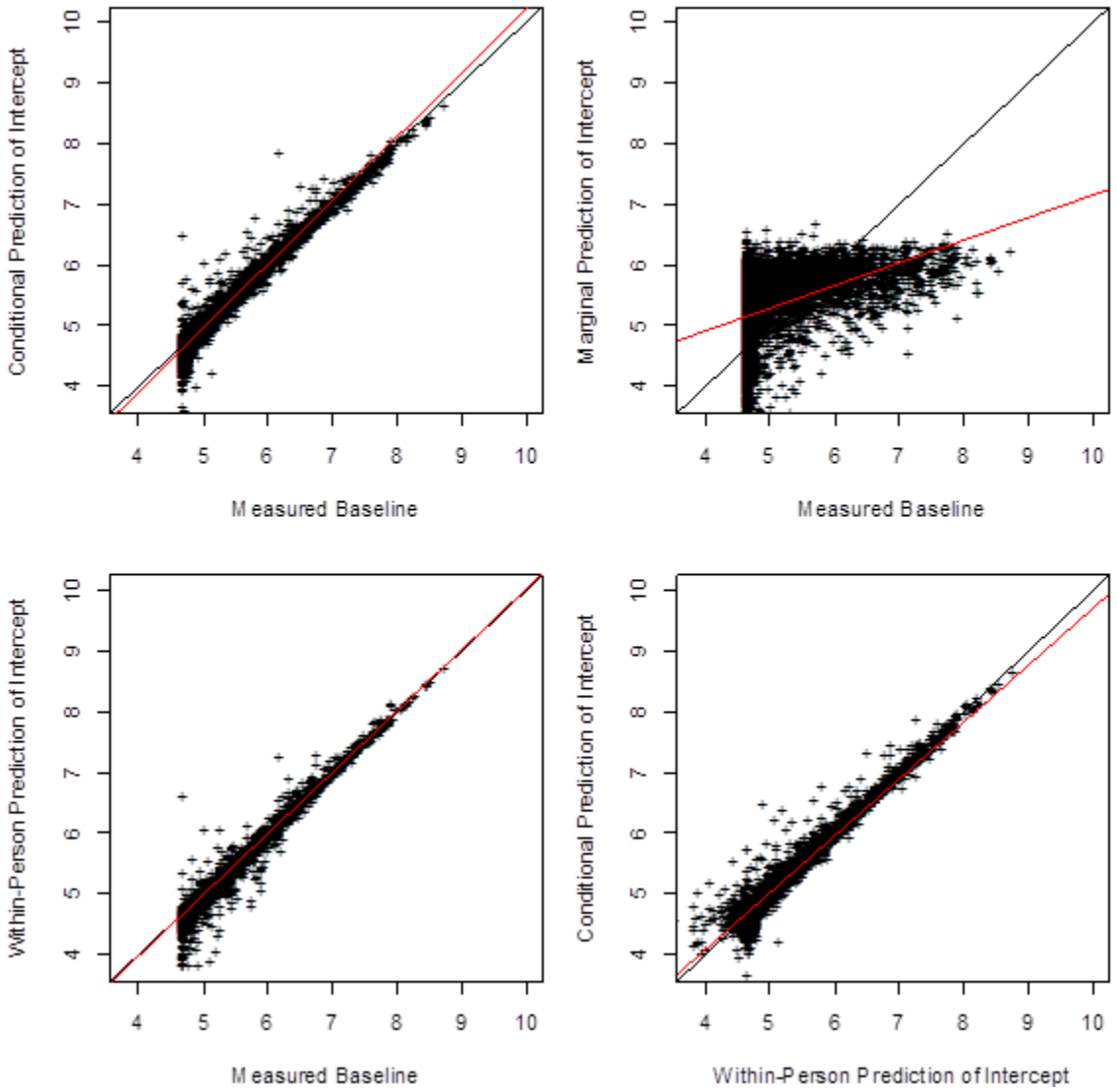
$$\hat{y}_{i0} = X'_{i0}\hat{\beta} + Z'_{i0}\hat{b} \tag{30}$$

Figure 19. Comparisons of intercepts as modeled or measured for each individual. 1-1 line shown in black on each panel, best fit line shown in red. Upper left: comparison between measured baseline CAC and modeled baseline in the MMCMB (23 II), including predicted random effects. Upper right: comparison between measured baseline CAC and predicted CAC at baseline using only the fixed effect portion of the MMCMB (23 II). Lower left: association between intercepts estimated from individual trajectories (Model (1)) compared to measured baselines. Lower right: association between intercepts from individual trajectories and from MMCMB (23 II) including predicted random effects.



The purpose of this figure is to support the claim that MMCMB (23 II) models the individual baselines well. Although the marginal (29) model underpredicts baseline measurements for participants with high CAC scores, incorporating random effects in the full model (30) predicts baseline about as well as the individual trajectories (Model (1)).

Figure 20. Comparisons of log-transformed intercepts as modeled or measured for each individual. A transformation of log(x+105) was used; the constant was added because some predicted intercepts were negative. 1-1 line shown in black on each panel, best fit line shown in red. Upper left: comparison between measured baseline CAC and modeled baseline in the MMCMB (23 II), including predicted random effects. Upper right: comparison between measured baseline CAC and predicted CAC at baseline using only the fixed effect portion of the MMCMB (23 II). Lower left: association between intercepts estimated from individual trajectories (Model (1)) compared to measured baselines. Lower right: association between intercepts from individual trajectories and from MMCMB (23 II) including predicted random effects.



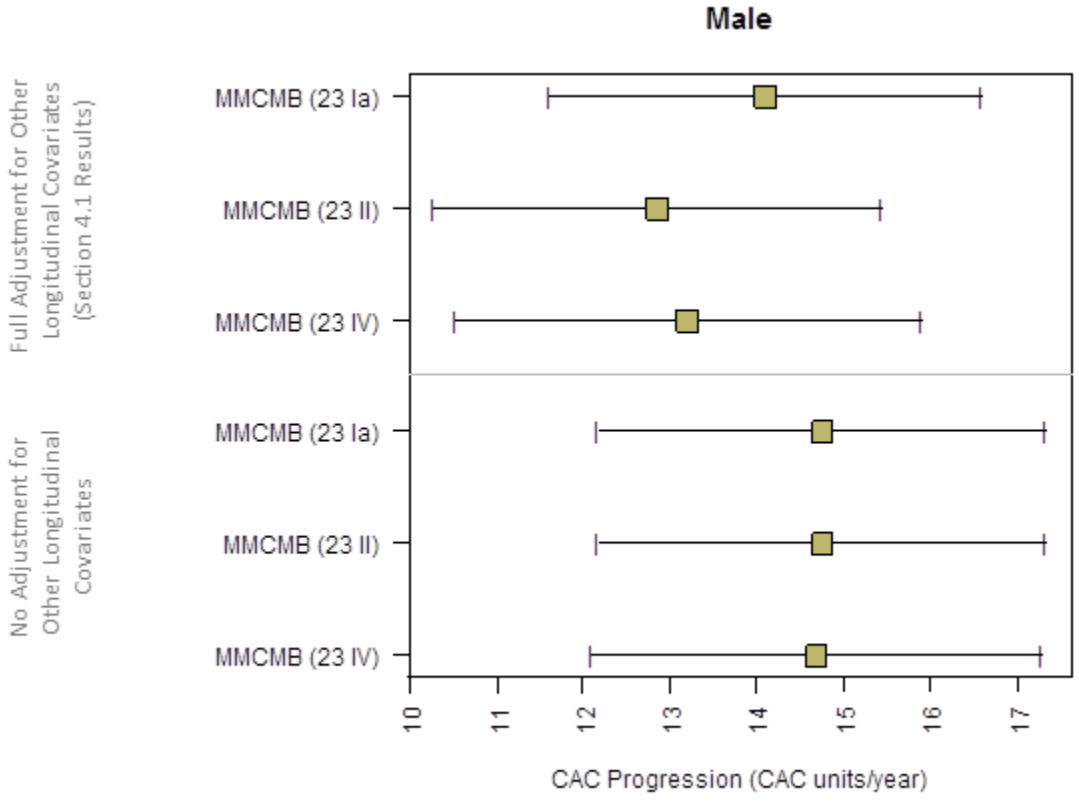
Plots of log-transformed intercepts are presented to show the lower end of the distribution more clearly. All methods of calculating baselines underpredict the baseline measurement for participants with the lowest CAC scores because scores are truncated at 0. In both figures, the

Model (1)-based intercepts seem to underpredict baseline more frequently than the conditional predictions, perhaps because the Model (1)-based intercepts are more sensitive to outliers that lead to steep slopes,

4.5. Variations on the Adjustment Model

In this section we present results from MMCMB (21 and 23) with everything but the covariate of interest dropped from W_{it}, so that we model the baseline with a rich model but do not control for confounding in the longitudinal relationship. For some of the comparisons in this section, we refer to MMCMB (21) rather than MMCMB (23) because we included a transient term for medication use.

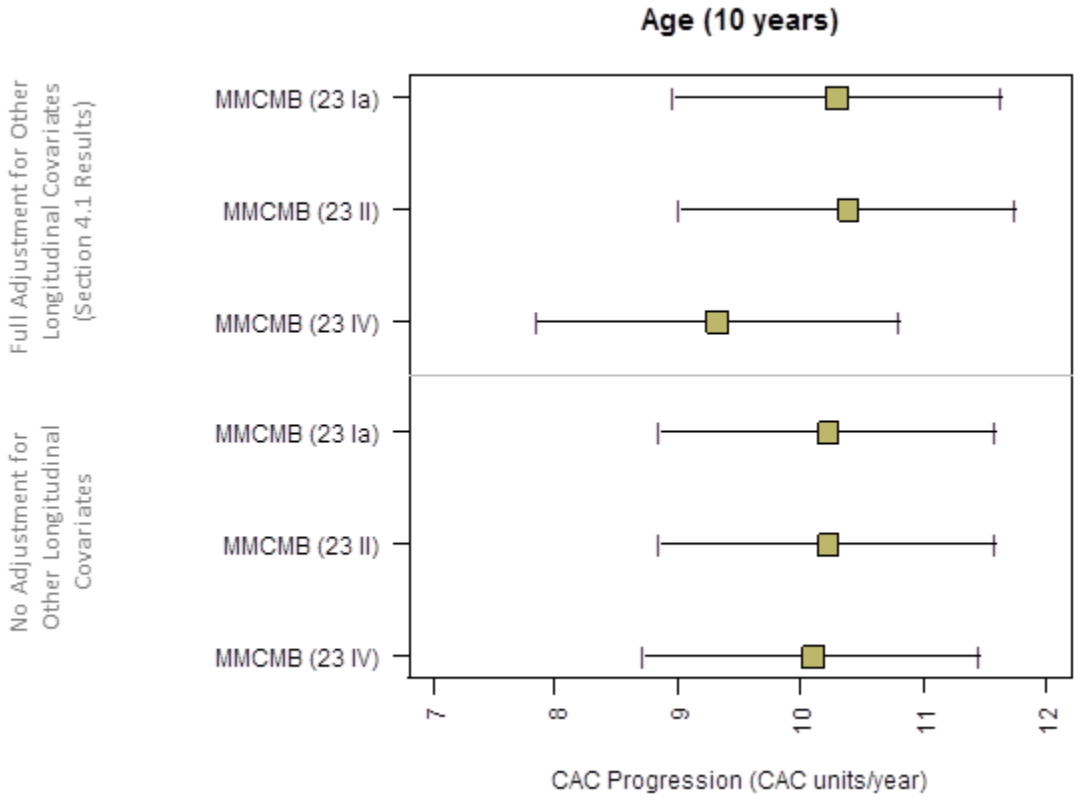
Figure 21. Longitudinal parameters (β_2) from staged models MMCMB (21) as outlined in Section 3.2. In the top section, W_{it} and X_{i0} contain all confounders as previously specified. In the bottom section, W_{it} contains only an indicator for "male". X_{i0} contains baseline values of confounders except for medications, which were moved to U_{it}.



In Figure 21, we observe that coefficient estimates and confidence interval widths are similar whether the longitudinal confounder adjustment is rich or sparse. The models that do include

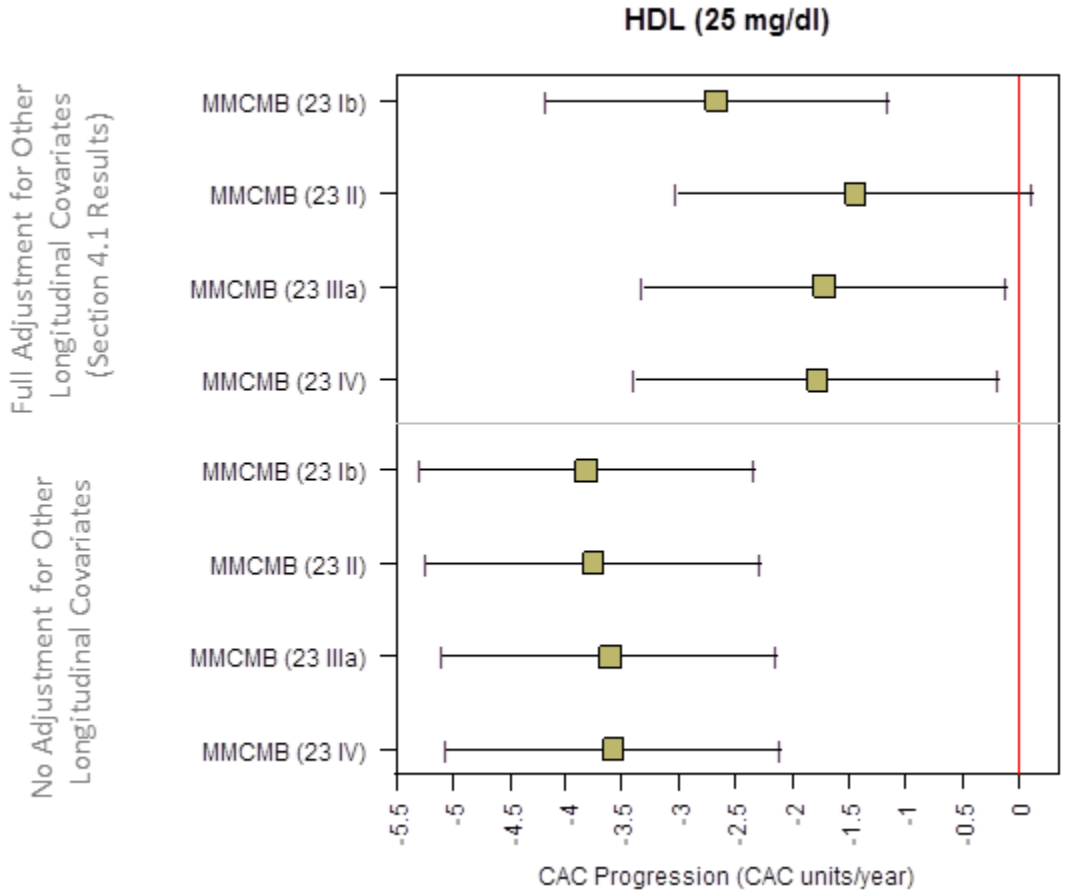
longitudinal adjustment for covariates show more sensitivity to model staging. Additional confounder sets (Ib, Ic, IIIa, and IIIb) were targeted at HDL and SBP and are not included here.

Figure 22. Longitudinal parameters (β_2) from staged models MMCMB (21) as outlined in Section 3.2. In the top section, W_{it} and X_{i0} contain all confounders as previously specified. In the bottom section, W_{it} contains only age. X_{i0} contains baseline values of confounders except for medications, which were moved to U_{it}.



In Figure 22, we observe that coefficient estimates and confidence interval widths are similar whether the longitudinal confounder adjustment is rich or sparse. As in Figure 21, we observe that models that do include longitudinal adjustment for other covariates are more sensitive to model staging. It is possible that this is partially due to the time-varying information that is introduced by those models in the more adjusted stages.

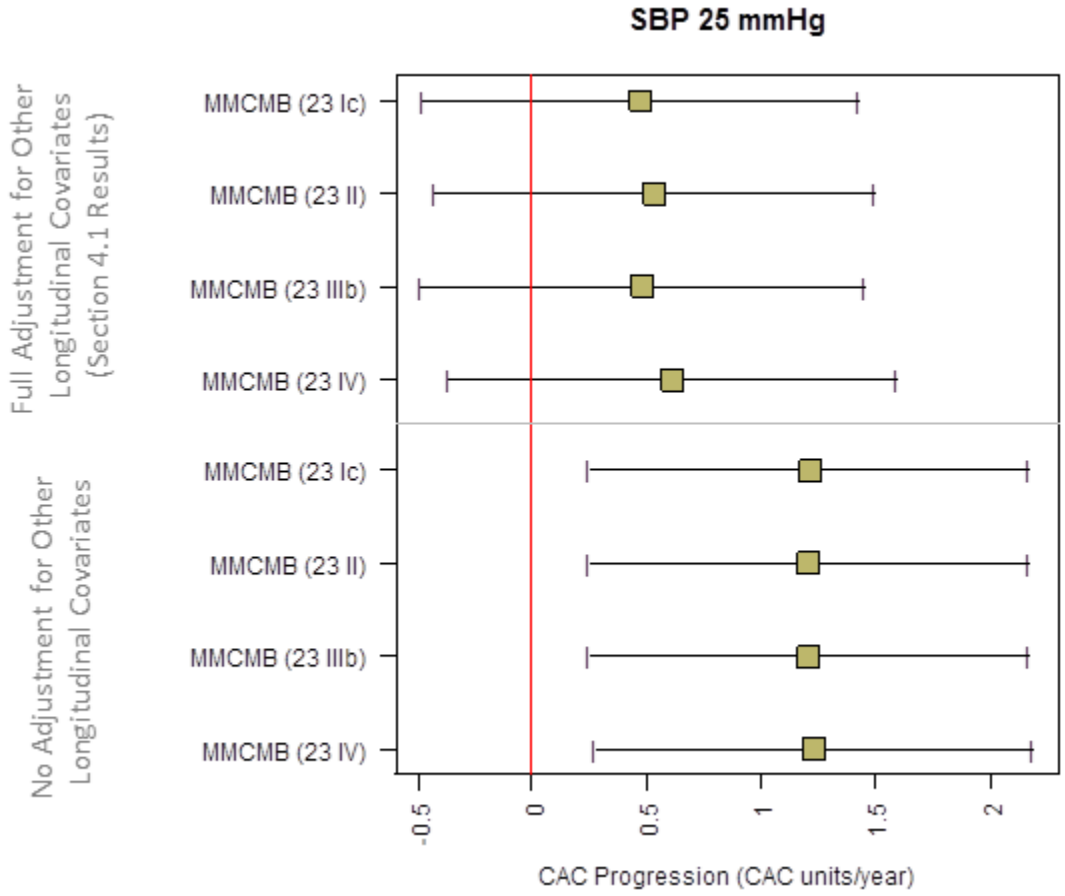
Figure 23. Longitudinal parameters (β_2) from staged models MMCMB (21) as outlined in Section 3.2. In the top section, W_{it} and X_{i0} contain all confounders as previously specified. In the bottom section, W_{it} contains only time-varying HDL. X_{i0} contains baseline values of confounders except for medications, which were moved to U_{it}.



In Figure 23, we observe that the results for HDL are sensitive to adjustment for other longitudinal covariates. The models without adjustment for other longitudinal covariates (bottom half of the figure) appear to be similar to the minimally adjusted model (MMCMB [23 Ib]). This model does not adjust for any time-varying covariates in the longitudinal relationship, just covariates defined at baseline. We note that results from Table 7 from models that include adjustment for the baseline covariate only (top section of Table 7), are similar to the models that do not include additional time-varying information. The difference between Figure 23 and the results from Table 7 is that the Table 7 results do not include time-varying information for HDL. The association between HDL and rate of change is stronger in the table. The other three models in Figure 23, MMCMB (23 II, 23 IIIa, and 23 IV) appear to be attenuated; these models include other time-varying covariates in the longitudinal adjustment. MMCMB (23 II, 23 IIIa, and 23 IV) are similar to the results from Table 7 where those models included time-varying covariates

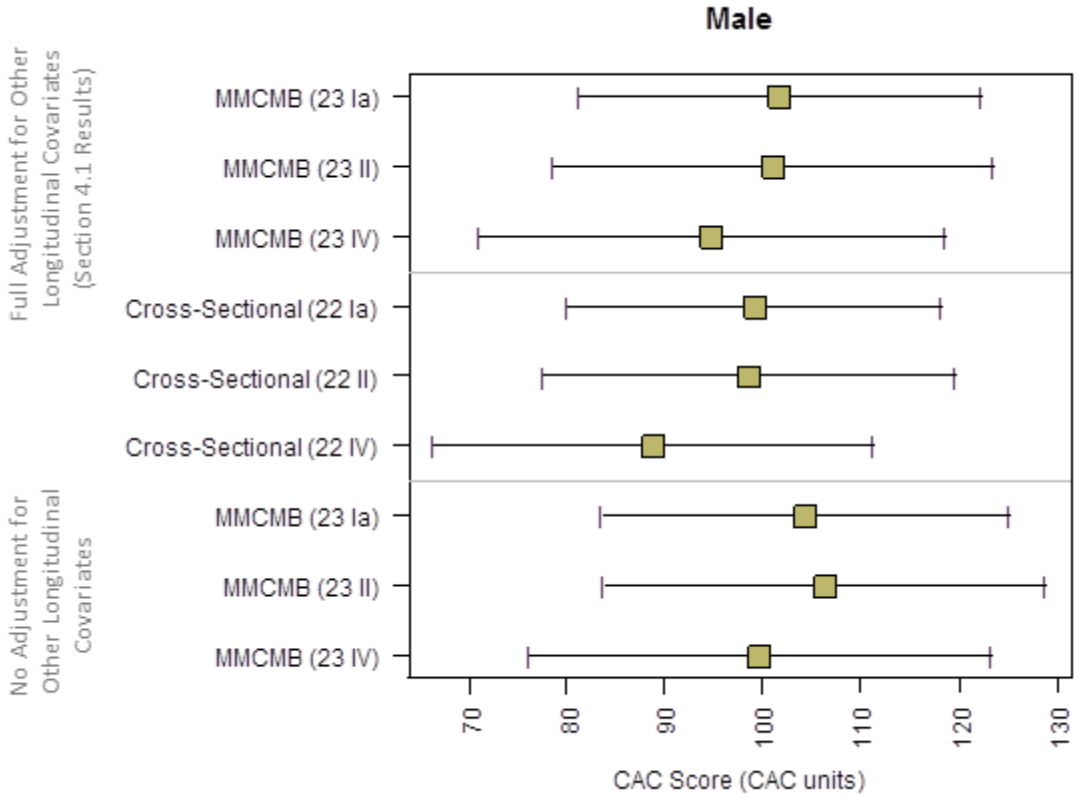
(bottom section of Table 7). In summary, the more time-varying information is included, the more the observed effects are attenuated.

Figure 24. Longitudinal parameters (β_2) from staged models MMCMB (21) as outlined in Section 3.2. In the top section, W_{it} and X_{i0} contain all confounders as previously specified. In the bottom section, W_{it} contains only time-varying systolic blood pressure. X_{i0} contains baseline values of confounders except for medications, which were moved to U_{it}.



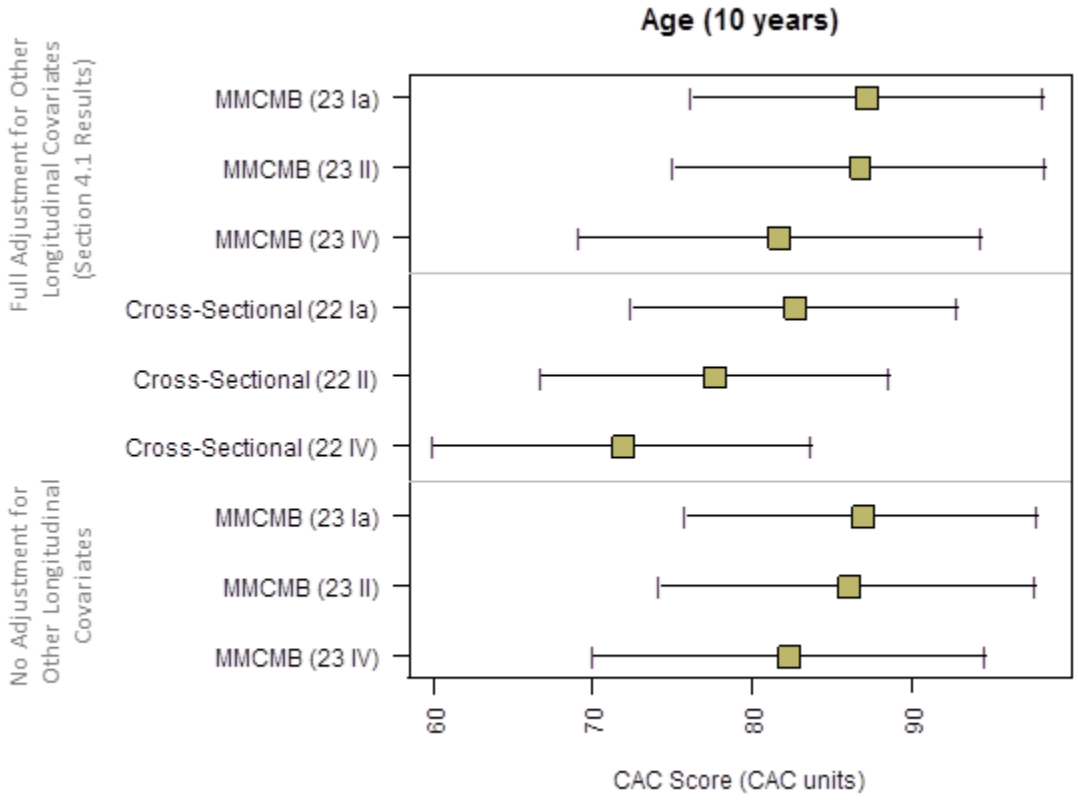
In Figure 24, we observe that the results for SBP are slightly sensitive to adjustment for other longitudinal covariates. We note that blood pressure results from the bottom section of Table 7 (time-varying covariates included) are similar to all of the results in Figure 24. All of the models in Figure 24 include time-varying blood pressure. The results in Figure 24 are very different from the results in the top section of Table 7 where only baseline blood pressure was used. Again, the more time-varying information is included, the more the observed effects are attenuated. The effect could be more pronounced for systolic blood pressure than for HDL because the correlation between serial measurements is lower for blood pressure.

Figure 25. Cross-sectional parameters (α_1) from staged models MMCMB (21) and cross-sectional models (22) as outlined in Section 3.2. In the top section, W_{it} and X_{i0} contain all confounders as previously specified. In the middle section, we show results from data restricted to Exam 1. In the bottom section, W_{it} contains only an indicator for "male". X_{i0} contains baseline values of confounders except for medications, which were moved to U_{it}.



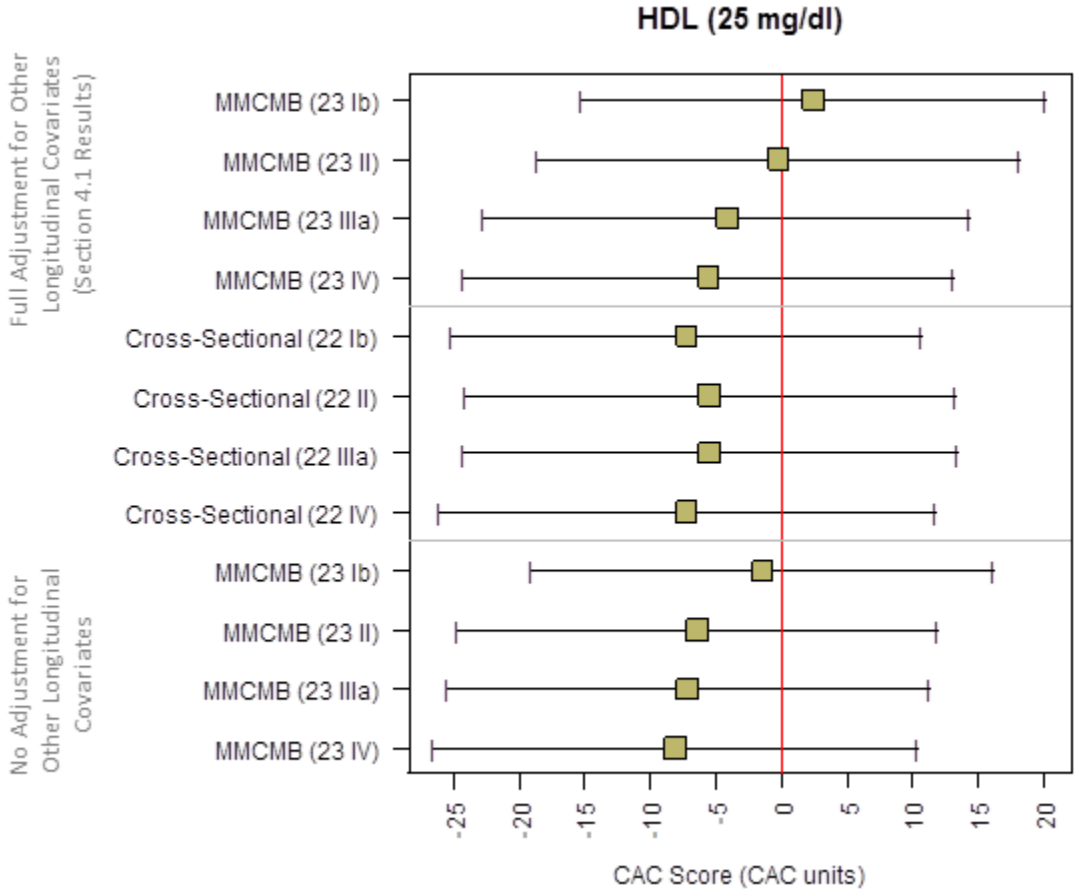
In Figure 25, we show cross-sectional associations between sex and CAC score at baseline as calculated by various models. Overall, results from MMCMB are comparable to results from cross-sectional (22) models.

Figure 26. Cross-sectional parameters (α_1) from staged models MMCMB (21) and cross-sectional models (22) as outlined in Section 3.2. In the top section, W_{it} and X_{i0} contain all confounders as previously specified. In the middle section, we show results from data restricted to Exam 1. In the bottom section, W_{it} contains only age at baseline. X_{i0} contains baseline values of confounders except for medications, which were moved to U_{it}.



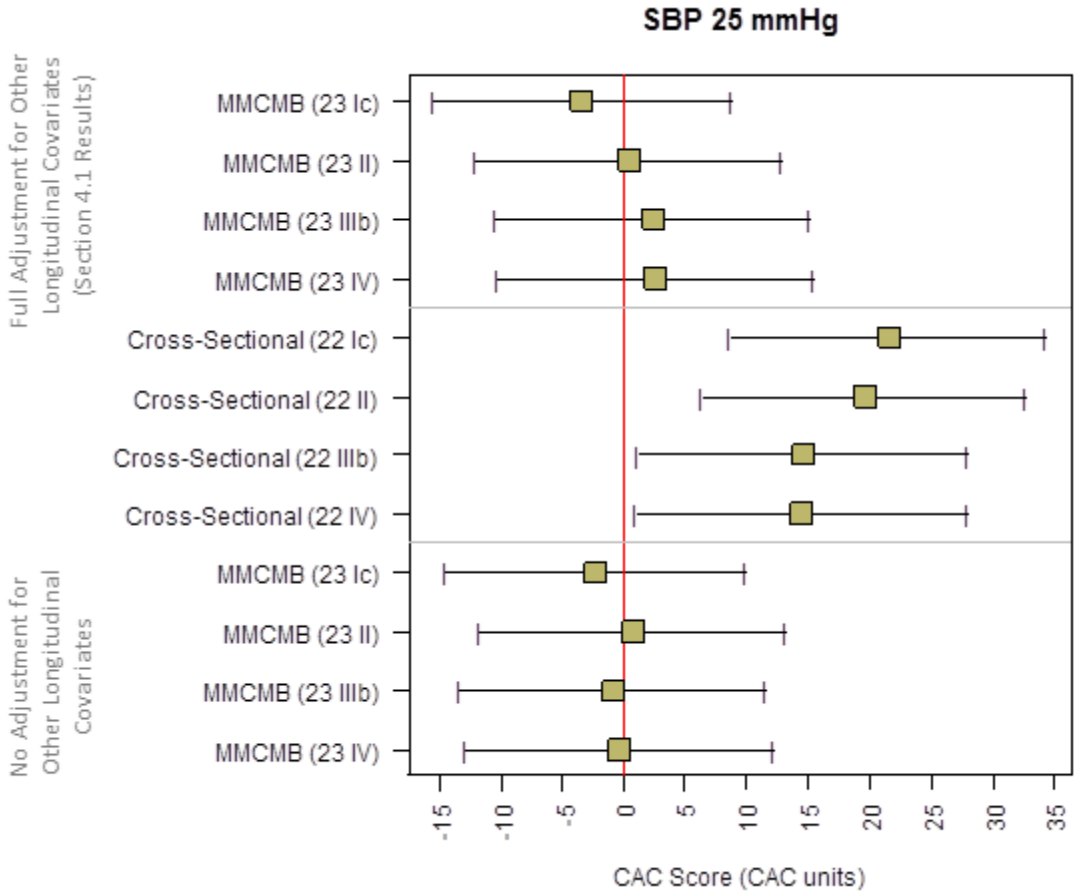
In Figure 26, we show cross-sectional associations between baseline age and CAC score at baseline as calculated by various models. Overall, results from MMCMB are comparable to results from cross-sectional (22) models. The effect for age is more sensitive to model staging that the results for sex are. The likely reason for this difference is that the association between the outcome and age is confounded by other indicators of health.

Figure 27. Cross-sectional parameters (α_1) from staged models MMCMB (21) and cross-sectional models (22) as outlined in Section 3.2. In the top section, W_{it} and X_{i0} contain all confounders as previously specified. In the middle section, we show results from data restricted to Exam 1. In the bottom section, W_{it} contains only time-varying HDL. X_{i0} contains baseline values of confounders except for medications, which were moved to U_{it}.



In Figure 27, we show cross-sectional associations between baseline HDL and CAC score at baseline as estimated by various models. Overall, results from MMCMB (23) are comparable to results from cross-sectional (22) models. Again we observe that effects are somewhat attenuated when more time-varying information is included, but for HDL the confidence intervals are huge and there is no evidence for an association with baseline CAC. Results in Figure 27 are consistent with results included in Table 8. As in the longitudinal parameters, we observe similarities between models in Figure 27 that include less time-varying information and results in Table 8 that adjust only for baseline covariates. We observe similarities between result in Figure 27 that include more time-varying information and results in Table 8 that include time-varying information.

Figure 28. Cross-sectional parameters (α_1) from staged models MMCMB (21) and cross-sectional models (22) as outlined in Section 3.2. In the top section, W_{it} and X_{i0} contain all confounders as previously specified. In the middle section, we show results from data restricted to Exam 1. In the bottom section, W_{it} contains only time-varying systolic blood pressure. X_{i0} contains baseline values of confounders except for medications, which were moved to U_{it}.



In Figure 28, we show cross-sectional associations between baseline systolic blood pressure and CAC score at baseline as calculated by various models. Overall, results from MMCMB (23) are different from results from cross-sectional (22) models. The MMCMB results in Figure 28 include time-varying information, and are similar to results presented in Table 8 from models that also included time-varying information. Results from Figure 28 from analyses restricted to Exam 1 data are similar to results presented in Table 8 from models that only include baseline values of covariates. The effect of time-varying covariates is more pronounced for SBP than for HDL. We hypothesize that the reason for this is that blood pressure is measured with more error, and the correlations between serial measurements of blood pressure are lower than correlations between serial measurements of HDL.

The main source of sensitivity in all results appears to be the inclusion of time-varying information. It is possible that the error inherent in the blood pressure measure is causing some instability in other coefficient estimates. If covariates are measured with error, it may be advantageous to control for confounding by these covariates in the cross-sectional relationship but not the longitudinal relationship, or to control for related covariates measured with less error. Additional avenues for analyzing the effects of time-varying covariates are discussed in Section 5.

Our other conclusion from these results is that when there is little temporal variation in a confounder, it does not confound the longitudinal relationship, and it is sufficient to include the confounder in the cross-sectional relationship. If the dataset is not large enough to support the estimation of a large number of fixed effect parameters and confounders are stable over time, it could be advantageous to control for confounding in the cross-sectional relationship but not the longitudinal relationship. Otherwise, an analyst interested in observing multiple longitudinal effects could include these effects in W_{it}.

4.6. Simulation Study Results

We begin this section by presenting specifics of data generation that depended on associations observed in the MESA dataset. The fixed effect model was chosen to be moderately complicated to provide for some realistic correlations between factors, but simple enough to ensure reasonable run times in fitting many iterations of the model. Exact coefficients are provided in Appendix A. We wanted to investigate the bias and uncertainty in the estimate of a parameter that was time-varying, continuous, and that had a small effect. The reason that we wanted to look at a predictor with these properties is that we hoped to generalize our conclusions to environmental exposures. As will become evident, we believe that this choice impacted the results that are presented in this section in unanticipated way. Nevertheless, in this section, we provide details of the data generation, results from the simulation as it was originally conceived, and a number of diagnostic scenarios.

Random effects at the subject level were generated as draws from a multivariate normal distribution with mean 0 and the following covariance matrix:

$$\begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{bmatrix} = \begin{bmatrix} 81647 & 5886 \\ 5886 & 1076 \end{bmatrix}$$

The residual variance was 8889 (SD = 94). In this dataset, the subject-specific random intercepts are about 10 times as variable as the residuals. Exam 1 and Exam 5 outcomes were correlated about 0.8. In the simulated data, iid normal errors were layered on top of the fixed effect + random effects, to bring the correlation within participant to the desired level. The following standard deviations ($\sqrt{var(\epsilon_{it})}$) for the errors were chosen to have this effect: 1 for 0.9, 200 for 0.7, 350 for 0.5, 550 for 0.3. In Table 9, we present estimated variances for the random effects from examples of the simulated data with residual variances.

Table 9. Estimated variance parameters from single realizations of error generation, using random effect scenario 1. Variance estimates for random effect components are similar between error realizations and between random effect scenarios 1-3. Variances in this table are estimated from simple models with fixed effects only for the overall intercept and slope to partition the total variability as between- and within- subject.

ρ	Random Intercept	Random Slope	Correlation of Random Effects	Residual Variance	Ratio of Intercept/Residual
0.9	92,736	1,250	0.66	27	3,434.7
0.7	92,315	1,159	0.68	39,727	2.3
0.5	96,405	1,334	0.58	119,972	0.8
0.3	103,881	1,417	0.50	299,573	0.3

Table 9 shows approximately what we would expect: in the $\rho = 0.9$ scenario, the residuals are less variable than they are in the observed data, and in other scenarios the residuals are more variable. The random intercepts and slopes have variances that are estimated to be in the ballpark of values in the variance-covariance matrix used to generate them. Qualitatively, differences observed in the estimated variance-covariance matrix of random effects between ransom effects scenarios seem minor. Ratios of random intercepts to residuals were consistent across different random effect realizations for the same level of iid error applied.

Statistics from the simulation study are presented in Tables 10 and 11. We show the mean, percent bias, standard deviation of the mean, mean standard error, and proportion of 95% confidence intervals that cover the true parameter for longitudinal HDL. We hypothesized that MMCMB would be less biased than ANCOVA and more precise than either of the competing models. We show results from three random effects scenarios to check the sensitivity of the bias result to the realized random effects.

Table 10. Summary of the longitudinal HDL parameter estimate and its uncertainty and coverage from the simulation study. Each group of columns conditions on one set of random effects.

		RE 1					RE 2			RE 3		
		Mean	**Bias**	**SD of Mean***	**Mean SE***	**Coverage**	**Mean**	**Bias**	**Coverage**	**Mean**	**Bias**	**Coverage**
GOAL		**-1.45**	**N/A**	**N/A**	**N/A**	**0.95**	**-1.45**	**N/A**	**0.95**	**-1.45**	**N/A**	**0.95**
	Correlation within person											
MMCMB	0.9	-1.45	0%	0.01	0.02	0.96	-1.45	0%	0.96	-1.45	0%	0.96
	0.7	-0.79	46%	0.92	1.19	0.96	-1.47	-1%	0.99	-2.46	-70%	0.93
	0.5	-0.66	55%	1.43	1.71	0.96	-1.46	-1%	0.98	-2.68	-85%	0.92
	0.3	-0.63	57%	2.17	2.39	0.96	-1.55	-7%	0.97	-2.81	-94%	0.94
ANCOVA	0.9	-1.45	0%	0.02	0.02	0.96	-1.45	0%	0.96	-1.45	0%	0.96
	0.7	-0.64	56%	0.86	1.24	0.97	-1.32	9%	0.99	-2.61	-80%	0.93
	0.5	-0.36	75%	1.31	1.72	0.96	-1.25	14%	0.99	-2.78	-92%	0.94
	0.3	-0.08	94%	1.88	2.30	0.95	-1.13	22%	0.99	-2.98	-106%	0.94
Scaled Change	0.9	-1.45	0%	0.02	0.02	0.96	-1.45	0%	0.96	-1.45	0%	0.96
	0.7	-0.46	68%	2.17	2.40	0.95	-1.03	29%	0.97	-3.09	-113%	0.92
	0.5	-0.55	62%	3.84	4.01	0.95	-1.16	20%	0.96	-3.30	-127%	0.93
	0.3	-0.53	64%	5.86	6.21	0.95	-1.15	20%	0.96	-3.31	-128%	0.95

*Standard deviations of the means and mean standard errors are not presented for random effects scenario 2 or 3 because these statistics are nearly identical across all the random effects.

Table 11. Relative efficiencies, defined as ratio of observed mean variances.

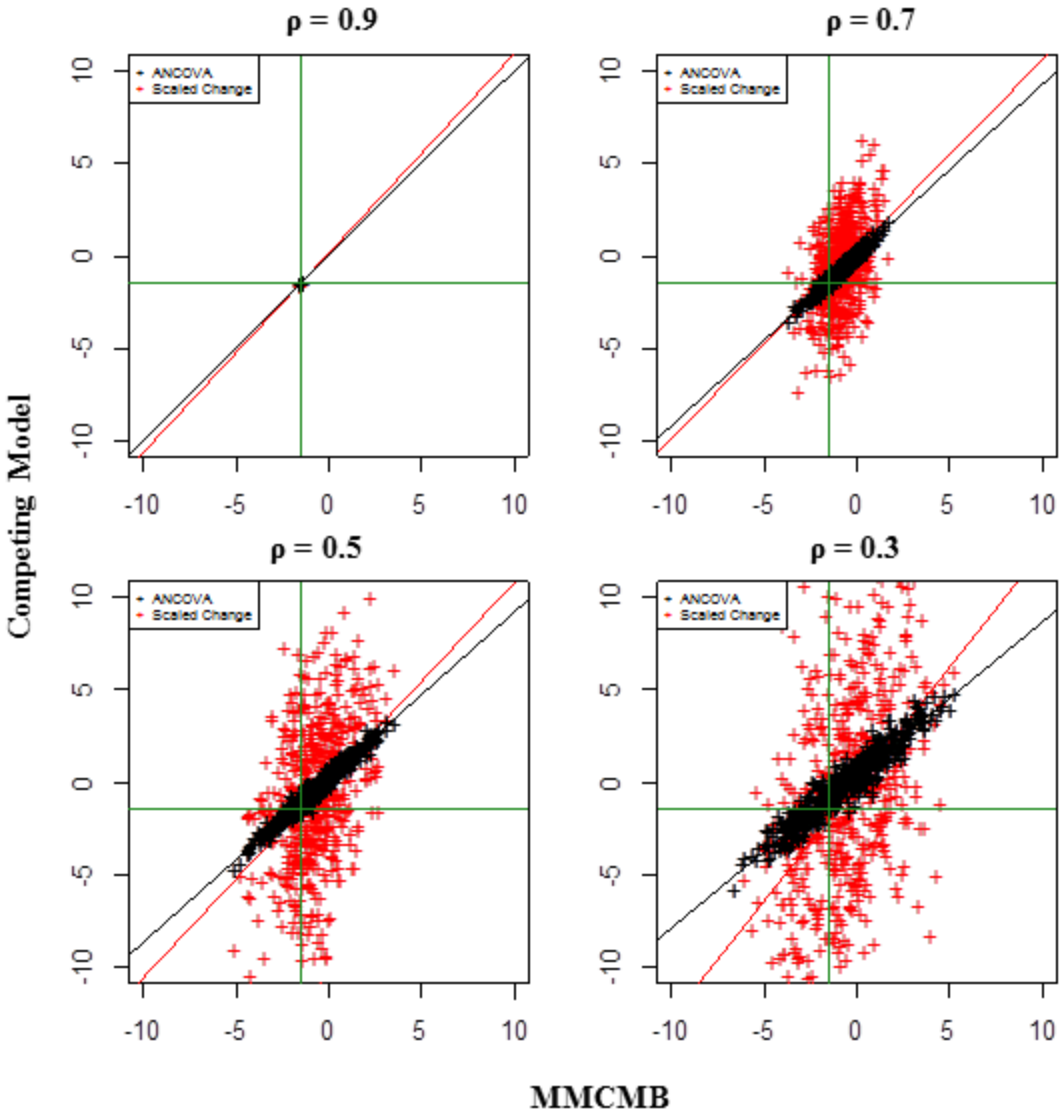
Correlation within person	MMCMB Scaled Change	ANCOVA Scaled Change	MMCMB ANCOVA
0.9	0.62	0.66	0.94
0.7	0.50	0.52	0.96
0.5	0.43	0.43	0.99
0.3	0.38	0.37	1.04

In Table 10, we observe that where $\rho = 0.9$, all models consistently estimate the true value of the parameter of interest. After adding any substantial degree of error, random effect scenarios 1 and 3 result in biased results across the board, but in opposite directions. MMCMB (23 II) is slightly less biased than the ANCOVA (24 II) model in each random effect scenario, which is in turn less biased than the scaled change model (25 II). Coverage is comparable between models, though slightly conservative in random effect scenarios 1 and 2 and slightly anti-conservative in random effect scenario 3. We suspect that results would be unbiased with correct coverage for all three models if averaged across a large number of random effect scenarios. Mean standard errors and the standard deviation of the estimated coefficient were virtually identical across random effect scenarios, which is unsurprising since the residuals should have a constant variance and the number of observations does not change between random effect scenarios.

In Table 11, we look at variance ratios to assess the relative efficiency of MMCMB compared to the other models. MMCMB (23 II) and ANCOVA (24 II) have similar efficiency between levels of ρ. As the level of error increases, the use of either of these models becomes increasingly advantageous compared to the scaled change model (25 II).

Correlation plots of the associations estimated by each model under identical data are presented in Figure 29.

Figure 29. Associations between longitudinal parameter estimates (β_2) from MMCMB and competing models using identical realizations of errors and random effect model 1. Best fit lines are shown in red and black. The "true" parameter value is shown in green. Plots are based on 500 replications.



In Figure 29, we present the associations estimated by each model under identical data after different levels of error were applied. Although measurements are not perfectly correlated where $\rho = 0.9$, the lack of correlation is attributed to the random effects rather than the iid errors. That is, 0.9 is the correlation between within-person measurements at baseline and follow-up in the "true" data for random effects covariance specification. This results in a consistently precise estimate of the true parameter for longitudinal HDL, although we will show that this is not always the case for every parameter. We observe a strong linear relationship between the

MMCMB (23 II) and ANCOVA (24 II) models. We suspect that this is due to the similarity between the form of the modeled baseline and the 'true' model used to generate the baseline value. The parameters estimated by the two models also have similar interpretations. The scaled change model (25 II) is subject to different constraints and estimates fewer parameters, which may be the reason for the substantial scatter. The best fit line (red) indicates that MMCMB coefficient estimates and scaled change estimates are associated under identical data, but this relationship is weak.

We observe that the results from the repeated ANCOVA (24 II) and MMCMB (23 II) are highly correlated even where $\rho = 0.3$, and this is also true for random effect scenarios 2 and 3 (shown in Appendix A, Figures 31 and 32). However, the association is not 1 to 1; there is some attenuation as the level of error increases. Between the scaled change (25 II) and MMCMB (23 II), the linear relationship is steeper than 1 to 1. Therefore, for cases where the MMCMB (23 II) result happens to be extreme relative to the true value, the ANCOVA (24 II) result tends to be closer on average to the true value and the scaled change result tends to be further on average from the true result. Our results imply that MMCMB (23 II) and the repeated ANCOVA (24 II) generally have similar performance in terms of both variance and average bias, and the performance becomes more similar as the level of error increases. On average, the MMCMB (23 II) results presented in Table 10 are the least biased, but the advantage to using MMCMB (23 II) is surprisingly small.

These results are different from what we expected based on the general pattern in the results in Tables 5 – 7, where the scaled difference (25) usually produced a result that was more similar to MMCMB (23). Upon closer inspection, the result for HDL looks less sensitive to adjustment for baseline than the other parameters. How can we reconcile these observations?

Carefully parsing the equation given for bias (Equation [9]) that was presented in Section 1.3.3., we realized that there were two reasons that a longitudinal parameter (β_2) would be estimated with minimal bias in ANCOVA. The obvious reason is that when the ratio of the measurement error to the total residual error is small, the bias is also small. The less obvious reason is that the bias is small when a covariate is not associated with the true outcome at baseline. The

coefficient for HDL at baseline in (23 II) is nearly 0 (see Table 13 in Appendix A); this is the parameter used to generate the simulated data. We hypothesize that the lack of association between baseline HDL and baseline CAC could lead to minimally biased estimates for ANCOVA (24) in both the data analysis and simulation study, and this would be consistent with the formula for bias in β_2 cited in Equation (9).

We did not recognize this property of the HDL parameter prior to executing the simulation and did not anticipate how this would translate to our simulated results. Preliminary evidence from a simulation that focuses on a different parameter supports our hypothesis that the lack of bias is related to the lack of association at baseline. To check the hypothesis, we examined the parameter estimates for the association between male sex and rate of change of CAC for 500 replications of the simulation. The result table and figures are presented in Appendix A, Table 16 and Figures 35 - 37. Interestingly, in the figures, we observe the same linear relationship between ANCOVA (24 II) and MMCMB (23 II) that we see in Figure 29. However, the center of the association is shifted away from the intersection of the green lines that indicate the true value of the parameter. In Table 16, we observe that coefficient estimates are more biased for all models for male sex than the coefficient estimates were for HDL. Between estimation methods, we observe the pattern we originally expected to see, where the ANCOVA (24 II) coefficient estimates show more bias as the variance of ε increases. MMCMB (23 II) and the scaled change analysis (25 II) show less bias than ANCOVA (24 II), but comparable bias to each other. Coefficient estimates from both methods (23 and 25) show more bias than we expected and poor coverage. We hypothesize that the bias is related to the categorical nature of the parameter.

Careful inspection of the full set of parameter estimates for a few examples of MMCMB (23 II) and ANCOVA (24 II) indicated that while the coefficients for categorical variables were estimated with bias under MMCMB (23 II), coefficients for continuous variables were estimated more accurately (results not shown). We were particularly interested in the coefficient for age, since we observe age to be associated with baseline and rate of change. Consistent with the hypotheses outlined in this section, age appeared to be more biased when estimated using ANCOVA (24 II) rather than MMCMB (23 II) when the variance of ε was large. These observations are based on a limited number of examples, but further work could confirm these

findings by averaging over a large number of replications. We will return to this point in Section 5.

Due to concerns about the dissimilarity between the random effect structure for CAC and other outcomes, we fit additional random effects scenarios to check the simulation results' sensitivity to the random effects covariance structure. The primary goals of the additional random effects scenarios were: 1) to examine a scenario in which random intercepts and slopes were independent of each other, rather than positively correlated, and 2) to examine a scenario in which the random intercepts were still variable, but where there was limited variability in the slopes.

Random effects scenarios 4 - 6 were generated as draws from a multivariate normal distribution with mean 0 and the following covariance matrix:

$$\begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{bmatrix} = \begin{bmatrix} 81647 & 0 \\ 0 & 100 \end{bmatrix}$$

Although we intended to use the same variance for intercepts and slopes as in random effects scenarios 1 – 3, we found that the resulting "true" data had properties that were problematic for our simulation. Choosing these effects to be independent of one another caused the "true" measurements at baseline and follow-up to have lower correlation ($\rho = 0.8$) than the "true" measurements in the original simulation (where $\rho = 0.9$). Therefore, we lowered the variability in the slopes until we observed $\rho = 0.9$ between measurements. This, in turn, caused us to adjust the distributions used to generate "measurement" errors. For random effects scenarios 4 – 6, we used SDs 1, 180, 300, and 475. We notice that low variability in the measurement error results in consistent estimates of the parameter of interest, regardless of random effect scenario.

Similarly, we checked random effect scenarios in which nearly all of the between-person variability was observed in the intercept, with practically none in the slope. The covariance used for random effect scenarios 7-9 was:

$$\begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{bmatrix} = \begin{bmatrix} 81647 & 0 \\ 0 & 1 \end{bmatrix}$$

For these scenarios, baseline and follow-up were very highly correlated, so the following SDs were chosen: 75, 200, 350, 550.

Result tables and figures for these analyses are provided in Appendix A. (Figures 33 - 34 and Tables 14 – 15). We note primarily that "unlucky" biases that we observe in random effects scenarios 1 and 3 do not appear in these results. This seems reasonable since without random slopes, the random intercepts could be seen as a simple component of the cross-sectional errors. We also observe that when we add a non-trivial amount of error to bring the within-person correlation of measurements to $\rho = 0.9$, we see a small amount of scatter in the coefficient estimate. This is in contrast to the random effect scenarios where, due to the variance in the random slopes, the correlation between the true baseline and true follow-up was 0.9 before any iid "measurement error" was added.

5. Discussion and Conclusions

This discussion focuses on the questions that our results brought up and proposes directions for future research in order to answer these questions. We open with general remarks about using statistical methods to answer scientific questions and about the limitations of observed data in Section 5.1. Next, several sections are dedicated to specific issues that were raised throughout the document: in Section 5.2, we cover selection bias and varying number of observations; in Section 5.3, we discuss time-varying covariates; in Section 5.4 we consider cross-sectional parameters. In Section 5.5 we present the limitations of the simulation study that are not otherwise covered in Section 5.3. In Section 5.6, we conclude with remarks that briefly reconnect our results to the literature presented in Section 1 and with final thoughts regarding the mixed model approach.

5.1. General Comments

Statistical methods should be chosen, first and foremost, to answer a scientific question. The nature of the exposure and type of effect an exposure is expected to cause will affect the specification of the model. Directed, acyclic graphs (DAGs) of the causal pathways involved may help clarify whether the effect should be seen as a rate or as a sudden change, and which covariates should be included as time-varying. If the baseline outcome is related to rate of change in the outcome over time, it is fruitful to consider whether the baseline measure can account for some unmeasured confounding, or whether it may be a mediator.

Features of the dataset can also affect which parameters can be estimated well. If the heterogeneity of measurements across time is limited, the difference between a sudden change and an effect on the rate of change will not be well distinguished. For datasets with only two time points for most participants, we have the option to use any of the methods outlined in Section 1.3, as well as MMCMB. If MMCMB is used, then we should be careful not to over-interpret the β parameter as a true slope, but it may still be useful to parameterize the model as an MMCMB rather than as a change model. This could be particularly useful in datasets where transient effects affect the outcome, or where participants change groups in important ways. We may want to adjust the baseline measurement and the follow-up measurement for different values of the same covariate, for example, treatment group. Transient effects were not a focus of

this thesis, so further research could explore the gains and uncertainties associated with these effects.

Another reason to use MMCMB for datasets with two time points per person, rather than using a simplified model, is that we may want to model the baseline and follow-up outcomes jointly because we believe the exposure affects both. Further simulations could be developed to explore exposures that affect both baseline and change.

5.2. Selection Bias and Varying Number of Observations

In analyses of CAC in the MESA cohort, we observed that rate of change parameters were generally comparable regardless of the method used, as long as we used the same dataset for all analyses. For the purpose of this thesis, which was primarily to illustrate the MMCMB and its similarities and differences compared to simplified methods, the results presented throughout Section 4 were based on a relatively healthy subset of 3191 participants. We observed in Table 1 that these participants' CAC scores at baseline were lower and less variable than CAC scores in the rest of the sample, pointing to strong selection bias in that subset.

An important feature of the MMCMB is its ability to leverage information from participants with fewer observations, even participants with only a single observation. None of the simplified methods provides a mechanism by which to include the participants with a single observation. To enrich our understanding of the impact of the inclusion of single-observation individuals, additional analyses could be conducted to compare results from MMCMB models that include:

1) all observations from all participants,
2) all observations from participants with 2 or more observations,
3) observations from participants with 2 or more observations, dropping observations at random so that some participants in the subset drop to having one observation.

This analysis could help us understand what additional information the single-observation participants provide under selection bias [i.e., comparing 1) to 2)], and what additional information these participants provide under missingness at random [i.e., comparing 2) to 3)].

5.3. Time-Varying Covariates

Our results were sensitive to the inclusion of time-varying covariates to an unsettling degree. Participants were followed over ten years and experienced many changes. As noted in Section 2.3, approximately one-third of participants were prescribed lipid-lowering medications and/or anti-hypertensive medications during the course of follow-up. Given the dramatic impacts these drugs can have on health outcomes, it seems appropriate to adjust for these covariates in a time-varying way, especially as the temporal spacing between observations increases.

However, there are complications associated with time-varying covariates that are not completely understood. The issues include: 1) the scientific interpretation of the coefficients for some time-varying covariates; 2) the best way to adjust for time-varying covariates; 3) whether 2) depends on the amount of error with which a covariate is measured; and 4) whether 2) depends on the expected lag between the time a covariate changes and the time we observe an effect on the outcome.

For exogenous covariates, we interpret the coefficient for an exposure by time interaction as the effect of a cumulative exposure on the rate of change of an outcome. For endogenous covariates, the interpretation may be less clear, because those covariates are changing both in response to external forces and may also exert changes on the outcome. Again, it may be a useful exercise to construct a DAG to determine which variables may be confounders, to clarify whether it is necessary to adjust for time-varying information and what the meaning is of those that are included.

Covariate values at baseline and at follow-up may be correlated to varying degrees, so in some cases most of the relevant information may be incorporated through adjustment for the baseline covariate value. Work by Adar et. al.[24] suggests that the change from baseline in a covariate could be an exposure of interest. Whether change from baseline could be a useful way to adjust for confounding is an open question. A potential simulation study that could answer this question would be one in which the outcome is associated with a time-varying covariate in a fixed and known way, and the result from 1) an MMCMB (21) that controls for the baseline value of that covariate and adjusts for the time-varying values of the covariate at each follow-up

time, to 2) an MMCMB (21) that controls for the baseline value of the covariates and adjusts for the change from baseline at each follow-up time. This comparison could be staged over several levels of correlation between baseline and follow-up covariate values. The study could examine the effect on a different parameter of interest, i.e. with the time-varying adjustment variable treated as a confounder, and/or could investigate the time-varying variable itself as the exposure of interest. We hypothesize that where the correlation between measurements of the covariate is high, and where the time-varying covariate is not the exposure of interest, that the effect on the estimate of the parameter of interest will be minimal.

We observed sensitivity in our results in Section 4 where we adjusted for time-varying blood pressure or treated time-varying blood pressure as an exposure or risk factor. This covariate is measured with error. The sensitivity we observed in our results could be related to the measurement error in the covariate. A fruitful direction for exploration could be to control for a smoothed covariate value or summary statistic, rather than the concurrently measured covariate. For example, the linear trajectory over time for the covariate could be calculated for each participant, and the predicted concurrent covariate could be the adjustment variable. A potential simulation in this vein would be to generate data according to "true" underlying values of time-varying covariates, and then analyze the data adjusting for time-varying covariates with variable levels of error layered onto them.

The last issue is that there may be a lag between the time a covariate changes and the time we observe an effect on the outcome. We would expect this to vary from outcome to outcome and exposure to exposure, but we propose as few possible analyses to explore the impact of a lagged effect. One possible analysis would be to adjust for covariates in a time-varying way, but incorporate a lag. For example, our analysis could adjust for the medication use or covariate value at the *prior* exam, rather than the current exam. Using observed data has the limitation that that the lag would not be consistent from person to person and exam to exam, e.g., a longer lag would occur between Exam 4 and 5 compared to Exam 1 and 2. A simulation could be implemented to generate data based on the assumption that the effect on the rate of change is associated with a known lag. Then the model would adjust for the current value of a covariate to see whether the result is consistent with the underlying truth.

5.4. Cross-Sectional Parameters

Additional work could explore the importance and interpretability of the cross-sectional parameters. We showed some evidence in Section 4.5 that indicates that cross-sectional control for confounding is more critical than longitudinal control for confounding, at least for covariates that are not time-varying. Further analyses could compare performance of over-specified models, which may contain terms with more co-linearity than we would usually care to have, to "best" scientific models. Very richly specified cross-sectional models may have good predictive properties, although individual coefficients become more difficult to interpret (e.g., control for total cholesterol, HDL, and triglyercides means that we cannot tell whether an effect found for HDL is truly related to an increase in HDL or a decrease in LDL which would be collinear). However, we also observed in Section 4.5 that our results were not sensitive to model staging. We might want to analyze a dataset with more apparent confounding, or develop a simulation study that has confounding embedded in it, to make sure that the lack of sensitivity to model staging is a generalizable result. We hypothesize that careful control for confounding in the cross-sectional relationship is often important, and other datasets would show sensitivity to model staging.

The parameterization of a model affects the interpretation of each coefficient in the model. Inclusion or exclusion of longitudinal covariates impacts the estimation of the cross-sectional parameters, so careful selection of the confounder model is necessary if both cross-sectional and longitudinal parameters are of interest. Cross-sectional results from MMCMB (23) were generally consistent with results from cross-sectional models (22), although we observed inconsistencies where time-varying covariates, particularly those measured with error, were included. Where exposure is related to the baseline outcome and the baseline outcome may be in the causal pathway, it may be appropriate to report cross-sectional effect estimates from MMCMB (23) as results of secondary interest. However, a more thorough understanding of the influence of time-varying covariates should be developed before reporting the cross-sectional parameters broadly.

5.5. Limitations of the Simulation Study

The simulation study included in this thesis showed results that were sensitive to properties of the parameter chosen for scrutiny. In MMCMB (23 II), HDL was not estimated to be strongly associated with CAC at baseline. Thus, in the fixed effect model used to generate the simulated data, HDL was not strongly related to the baseline outcome. The consequence of that choice resulted in an unexpected result, which was that control for the baseline measurement did not significantly impact the estimation of the parameter of interest. Upon reflection, we could have anticipated this, since it is consistent with both the background from Yanez et. al.[4] that was presented in Section 1.3.3. and our own results, presented in Table 7. We also observed that estimates for coefficients of categorical variables (as presented for male sex in Table 16, and observed informally for race and site during diagnostic analyses) are possibly more biased than estimates for coefficients of continuous variables. Future work could carefully characterize covariates and the impact of those characteristics on bias and uncertainty in estimates of association between the covariates and the outcome in the presence of measurement error.

Additional simulations similar to those included in this thesis could explore the impact of cross-sectional covariate-outcome relationships on the estimates of longitudinal relationships. Specifically, we could repeat our simulation with a different fixed effect model and the following additions:

1) Focus on a longitudinal coefficient for a covariate that is strongly associated with both baseline outcome and change over time.
2) Examine a longitudinal coefficient for a covariate that is strongly associated with baseline outcome but not change.
3) Examine a longitudinal coefficient for a covariate that is not associated with baseline at all, in any model, but is strongly associated with change.
4) Examine both categorical and continuous covariates for each of 1) through 3).

We hypothesize that in analysis 1) we will observe bias in the ANCOVA model for all simulations. We hypothesize that in analysis 2) we might see that ANCOVA induces an association in the longitudinal parameter where no true association exists. We hypothesize that in analysis 3) we would not see bias in ANCOVA. We hypothesize that MMCMB would be

unbiased in 1) – 3) for continuous covariates. We expect that this analysis would uncover some of the reasons for the bias observed in the categorical parameter estimate.

Our simulation study was subject to additional limitations. We did not explore the effect of model mis-specification on variance or bias in the mixed model. Model mis-specification could impact the results of the model in unknown ways. Because the model specified was taken to be true and known, the repeated ANCOVA controlled for a baseline that was very similar to the baseline modeled by the MMCMB. That is, by virtue of the data generating mechanism, the measured baseline was related to the same covariates as were included in the model for the baseline in the MMCMB. Additional simulations could be developed in which the baseline and rate of change are associated due to the influence of an unobserved covariate. Then we could observe the effect of control for measured and modeled baselines in circumstances where the baseline could be viewed as a confounder. Because our results were not sensitive to model staging, perhaps we would see minimal sensitivity to model mis-specification.

Our simulation conditioned on a small number of random effects realizations, and some sensitivity to the particular realization of the random effects was observed. Further research could more thoroughly characterize the impact of random effects realizations on parameter estimates. One possibility would be to hold the final errors ("measurement errors") constant, but generate new random effects for each replication of the analysis, and then average over realizations of the random effects. Characterizing the variability in effect estimates that is due to random effects could be a useful descriptive analysis. We hypothesize that the result would show that coefficient estimates are unbiased on average. Additional covariance structures for the random effects could also be explored to simulate the behavior of other outcomes.

5.6. Concluding Remarks

The MMCMB (21) exhibits desirable properties under a variety of assumptions, and it is an excellent analytic approach for cases where exposures and confounders are time-varying, and where the exposure may influence the level of the outcome at baseline. Because it is a more sophisticated model, it may be inaccessible to non-mathematicians. A simpler alternative model could be chosen that would be targeted at the scientific question. The scaled change (12 or 26)

that seemed like the best alternative choice *a priori* showed reasonable performance in terms of bias in the simulation results, and produced comparable results in the data analysis. Therefore, it should be a reasonable simplified model choice in many cases, since the other "simplified" analyses we presented actually have parameterizations that are as complicated as the MMCMB. We note that according to (12 or 26), we estimate fewer parameters for the scaled change model than we do for other models. This model is less flexible than other models we considered, which may be a reason that its var(ε) is higher and its efficiency is lower. Since the precision of the scaled change (12 or 26) is lower than other methods, we may prefer a more precise method when effect sizes are expected to be small.

In keeping with the literature, we observed that control for the measured baseline biased our results where baseline and rate of change were associated with the same covariate. Bias formulas were not presented in the literature for SLAIN or for the analysis of slopes, but we observed the analysis of slopes to be consistent with other methods and SLAIN not to be so. These methods were not assessed in the simulation study since they were found to be too computationally intensive for what was supposed to be a simple exercise, so further simulation studies could assess their performance. However, we note that these methods have limitations at the outset which might discourage their use where we have other methods available. Disadvantages include properties that we mentioned in Section 1, such as not having a natural way to handle varying numbers of observations, and concerns that they cannot incorporate time-varying information. In cases where health outcomes could be strongly related to time-varying covariates, then we might want to incorporate this information, but not have enough within-person information to correctly adjust our slopes.

Overall, we are satisfied with the performance of MMCMB (21) compared to the other methods assessed and are optimistic that we can communicate its form, interpretation, and advantages in a way that will lead to broader use.

6. Appendices

A. Additional Figures and Tables

Figure 30. Cartoon of hypothetical outcome trajectories in three individuals. Individual A (red) is never treated and progresses linearly over time. Individual B's progression (shown in green) is parallel to Individual A until time t = 1, when treatment is prescribed. Individual B's outcome still increases parallel to Individual A over time, but at any time t > 1, his outcome is lower than it would have been if he were untreated. An intercept estimated for him at time t = 0 (shown as an open green circle) is much lower if it is based only on data for t > 1. Individual C (blue) has been taking a different kind of treatment during the entire course of follow-up. This type of treatment prevents progression, so her outcome is relatively stable over time.

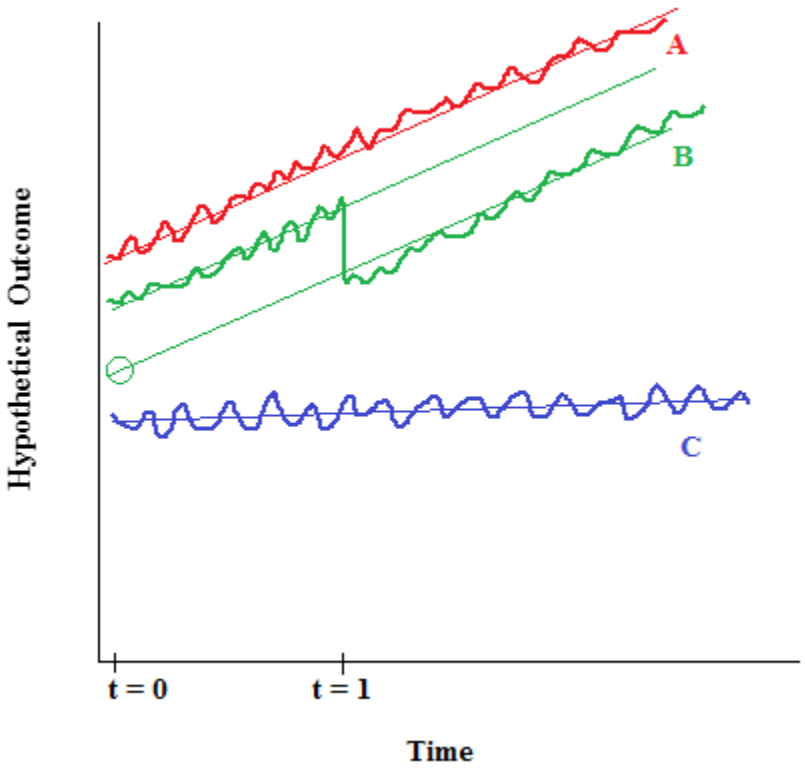


Figure 31. Associations between estimates for the parameter of interest from MMCMB and from competing models under random effects scenario 2, with identical errors applied to the "true" simulated data. Best fit lines are shown in red and black. The "true" value of the coefficient of interest is shown in green. Plots are based on 500 replications.

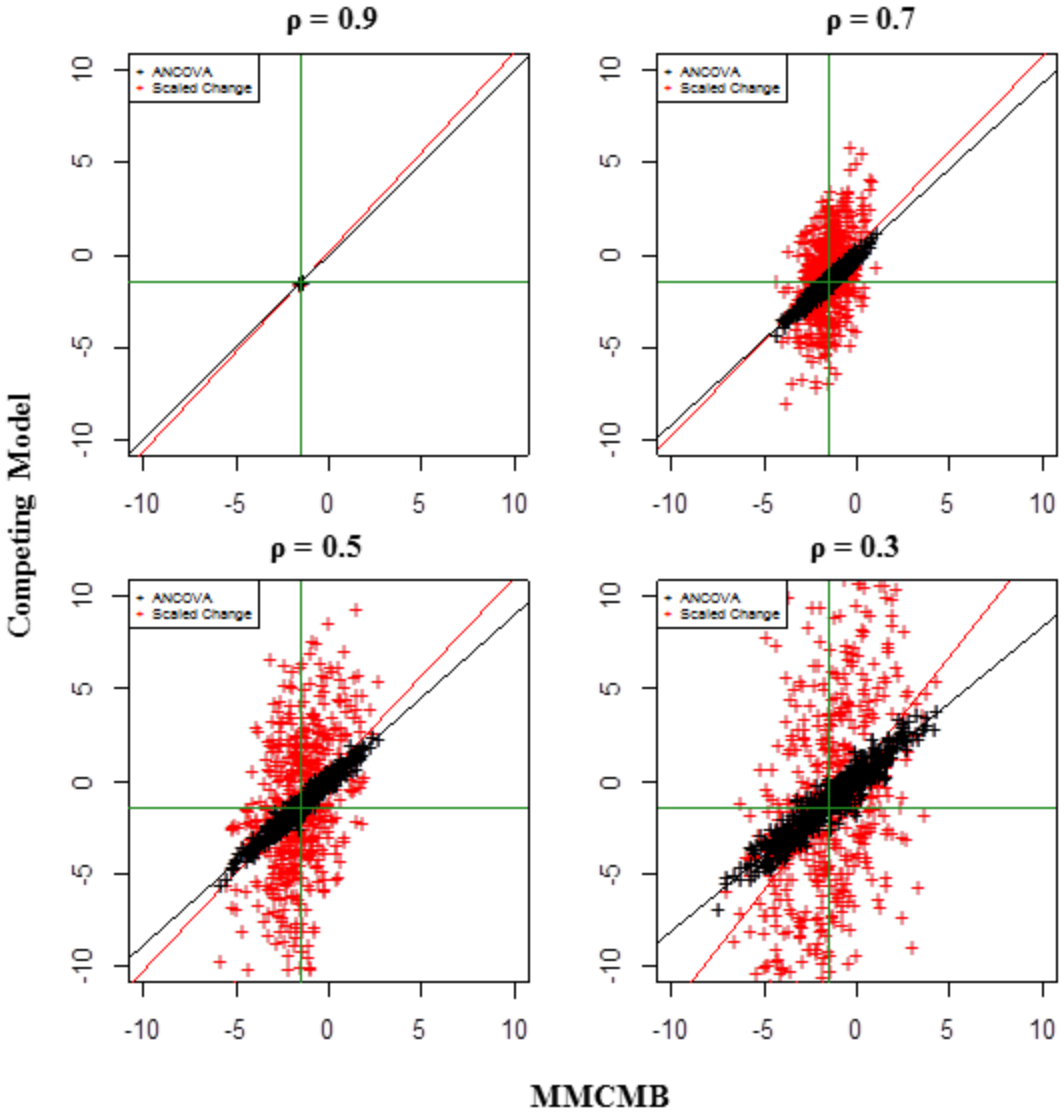


Figure 32. Associations between estimates for the parameter of interest from MMCMB and from competing models under random effects scenario 3, with identical errors applied to the "true" simulated data. Best fit lines are shown in red and black. The "true" value of the coefficient of interest is shown in green. Plots are based on 500 replications.

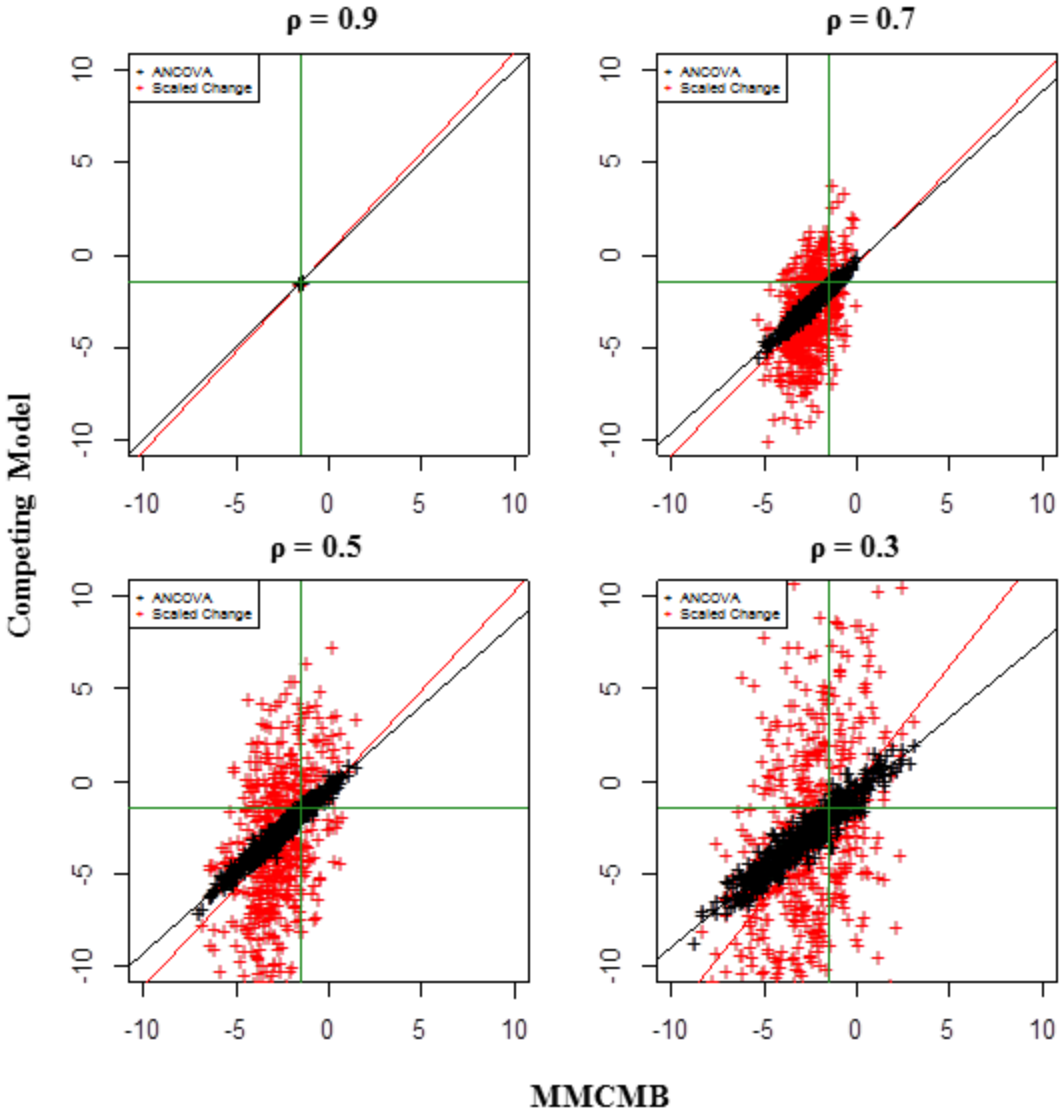


Figure 33. Associations between estimates for the parameter of interest from MMCMB and from competing models under random effects scenario 4, with identical errors applied to the "true" simulated data. Random effects scenarios 5 and 6 were nearly identical. Best fit lines are shown in red and black. The "true" value of the coefficient of interest is shown in green. Plots are based on 500 replications.

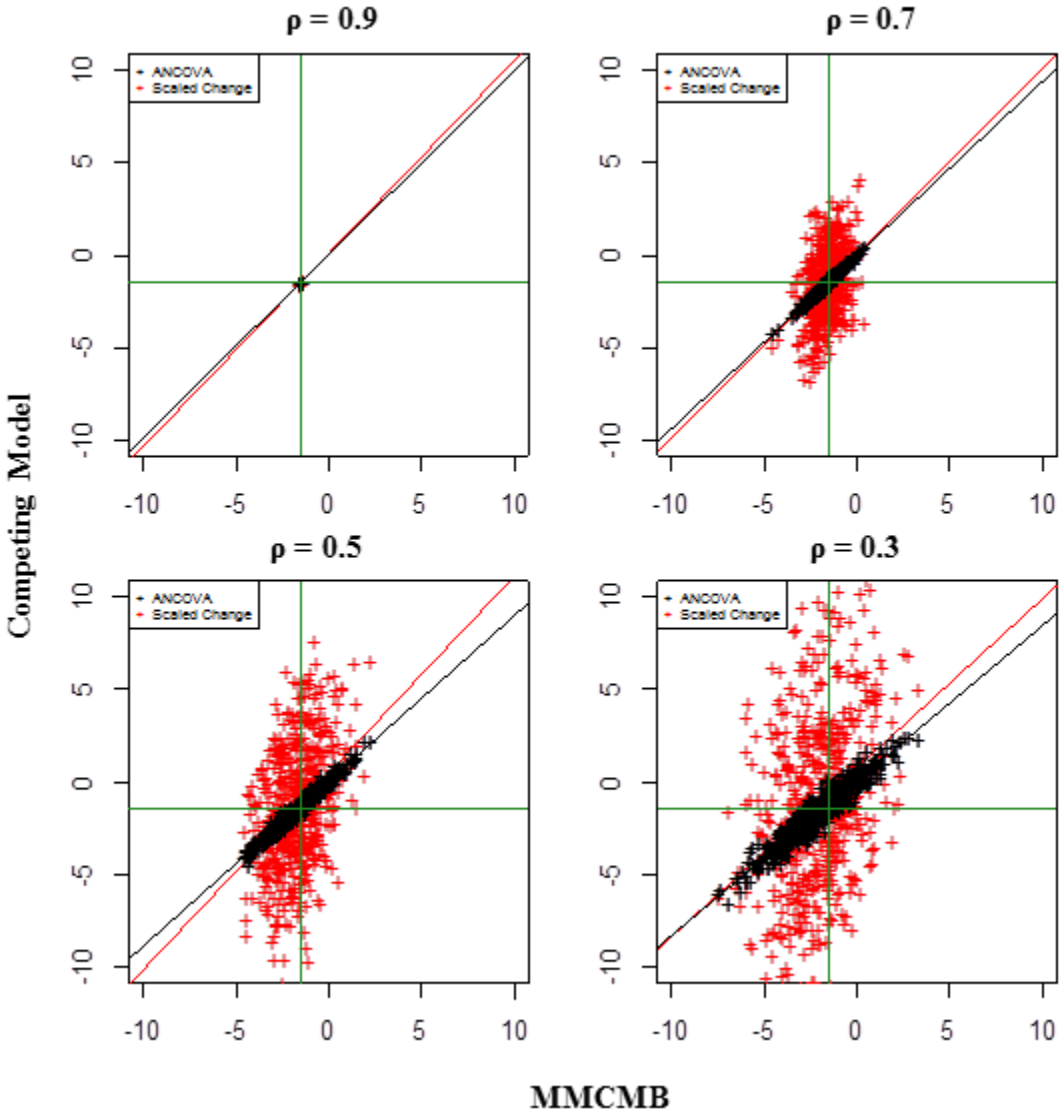


Figure 34. Associations between estimates for the parameter of interest from MMCMB and from competing models under random effects scenario 7, with identical errors applied to the "true" simulated data. Random effects scenarios 8 and 9 were nearly identical. Best fit lines are shown in red and black. The "true" value of the coefficient of interest is shown in green. Plots are based on 500 replications.

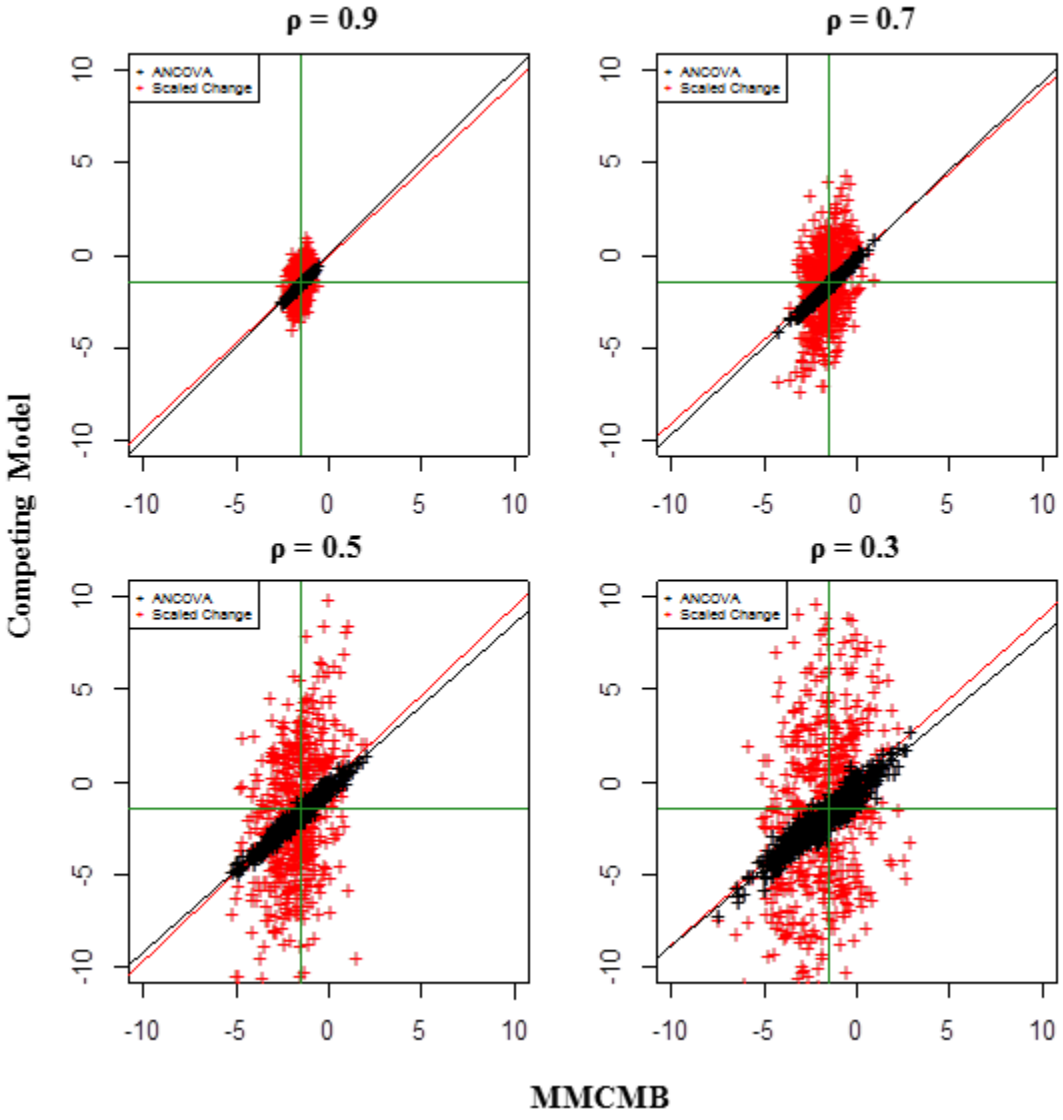


Table 12. Observed coefficients for longitudinal parameters (β) used as the basis for the simulation study fixed effect model. Each coefficient indicates the additional amount of increase per year for an increment increase in the covariate.

Covariate	Coefficient Estimate	Standard Error
Time (per year)	14.67	1.85
Age (10 years)	10.39	0.70
Male	12.86	1.32
Hispanic	-1.76	1.94
Black	-4.67	1.68
Chinese	1.12	2.43
White		Reference
Systolic Blood Pressure (25 mmHg)	0.54	0.49
BMI (5 kg/m^2)	2.01	0.56
Total Cholesterol (50 mg/dl)	-2.11	0.55
HDL (25 mg/dl) (selected as parameter of interest)	**-1.45**	**0.80**
Winston-Salem	3.70	2.22
NYC	-0.69	2.27
St. Paul	2.92	2.30
LA	-1.93	2.29
Baltimore	-6.92	2.42
Chicago		Reference

Table 13. Observed coefficients for cross-sectional (α) parameters used as the basis for the simulation study fixed effect model.

Covariate	Coefficient Estimate	Standard Error
Intercept	53.93	15.34
Age at Baseline (10 years)	86.79	5.95
Male	101.03	11.48
Hispanic	-39.55	16.22
Black	-32.46	14.20
Chinese	-45.96	20.36
White		Reference
Systolic Blood Pressure at Baseline (25 mmHg)	0.42	6.37
BMI at Baseline (10 kg/m^2)	-3.43	5.35
Total Cholesterol at Baseline (50 mg/dl)	-2.03	6.77
HDL at Baseline (25 mg/dl)	-0.24	9.43
Winston-Salem	16.18	18.56
NYC	18.14	18.92
St. Paul	51.99	19.10
LA	2.76	19.03
Baltimore	48.36	20.14
Chicago		Reference

Table 14. Summary of the longitudinal HDL parameter estimate and its uncertainty and coverage from the simulation study for RE scenarios 4-6. Each group of columns conditions on one set of random effects. The table is based on 1000 replications.

		RE 4					RE 5			RE 6		
		Mean	**Bias**	**SD of Mean***	**Mean SE***	**Coverage**	**Mean**	**Bias**	**Coverage**	**Mean**	**Bias**	**Coverage**
GOAL		**-1.45**	**N/A**	**N/A**	**N/A**	**0.95**	**-1.45**	**N/A**	**0.95**	**-1.45**	**N/A**	**0.95**
	Correlation within person											
MMCMB	0.9	-1.45	0%	0.02	0.02	0.95	-1.45	0%	0.95	-1.45	0%	0.95
	0.7	-1.51	-4%	0.70	0.79	0.97	-1.85	-28%	0.95	-1.36	6%	0.97
	0.5	-1.68	-16%	1.17	1.23	0.94	-1.97	-36%	0.93	-1.06	27%	0.95
	0.3	-2.08	-43%	1.75	1.86	0.95	-2.29	-58%	0.94	-0.84	42%	0.95
ANCOVA	0.9	-1.45	0%	0.02	0.02	0.96	-1.45	0%	0.96	-1.45	0%	0.96
	0.7	-1.39	4%	0.67	0.77	0.98	-1.96	-35%	0.94	-1.22	16%	0.96
	0.5	-1.42	2%	1.07	1.17	0.97	-2.19	-51%	0.92	-0.79	46%	0.93
	0.3	-1.59	-10%	1.54	1.70	0.97	-2.58	-78%	0.92	-0.43	70%	0.94
Scaled Change	0.9	-1.45	0%	0.02	0.02	0.94	-1.45	0%	0.95	-1.45	0%	0.95
	0.7	-1.35	7%	1.91	2.03	0.97	-1.67	-15%	0.96	-1.45	0%	0.96
	0.5	-1.27	12%	3.32	3.36	0.95	-1.58	-9%	0.95	-1.36	6%	0.95
	0.3	-1.22	16%	4.79	5.32	0.97	-1.53	-6%	0.97	-1.31	9%	0.97

*Standard deviations of the means and mean standard errors are not presented for random effects scenario 2 or 3 because these statistics are nearly identical across all the random effects.

Table 15. Summary of the longitudinal HDL parameter estimate and its uncertainty and coverage from the simulation study for RE scenarios 7-9. Each group of columns conditions on one set of random effects.

		RE 7					RE 8			RE 9		
		Mean	**Bias**	**SD of Mean***	**Mean SE***	**Coverage**	**Mean**	**Bias**	**Coverage**	**Mean**	**Bias**	**Coverage**
GOAL		**-1.45**	**N/A**	**N/A**	**N/A**	**0.95**	**-1.45**	**N/A**	**0.95**	**-1.45**	**N/A**	**0.95**
	Correlation within person											
MMCMB	0.9	-1.50	-3%	0.31	0.31	0.94	-1.46	-1%	0.95	-1.56	-7%	0.94
	0.7	-1.53	-5%	0.77	0.77	0.96	-1.45	0%	0.96	-1.71	-18%	0.94
	0.5	-1.64	-13%	1.30	1.29	0.94	-1.50	-3%	0.94	-2.02	-39%	0.92
	0.3	-1.76	-22%	1.81	1.83	0.95	-1.57	-8%	0.95	-2.35	-62%	0.93
ANCOVA	0.9	-1.51	-4%	0.31	0.31	0.94	-1.50	-3%	0.95	-1.57	-8%	0.94
	0.7	-1.60	-11%	0.73	0.75	0.96	-1.66	-15%	0.95	-1.78	-23%	0.94
	0.5	-1.79	-23%	1.17	1.21	0.94	-1.94	-34%	0.93	-2.14	-47%	0.91
	0.3	-1.94	-34%	1.59	1.68	0.95	-2.18	-50%	0.94	-2.44	-69%	0.92
Scaled Change	0.9	-1.44	0%	0.80	0.84	0.97	-1.43	1%	0.97	-1.50	-3%	0.97
	0.7	-1.40	3%	2.05	2.13	0.96	-1.40	4%	0.96	-1.46	-1%	0.96
	0.5	-1.52	-5%	3.60	3.64	0.95	-1.51	-4%	0.95	-1.58	-9%	0.95
	0.3	-1.49	-2%	5.01	5.31	0.96	-1.48	-2%	0.96	-1.54	-6%	0.96

*Standard deviations of the means and mean standard errors are not presented for random effects scenario 2 or 3 because these statistics are nearly identical across all the random effects.

Table 16. Summary of the longitudinal male sex parameter estimate and its uncertainty and coverage from the simulation study for RE scenarios 1 - 3. Each group of columns conditions on one set of random effects. Results are based on 500 replications.

		RE 1					RE 2			RE 3		
		Mean	**Bias**	**SD of Mean***	**Mean SE***	**Coverage**	**Mean**	**Bias**	**Coverage**	**Mean**	**Bias**	**Coverage**
GOAL		**12.86**	**N/A**	**N/A**	**N/A**	**0.95**	**12.86**	**N/A**	**0.95**	**12.86**	**N/A**	**0.95**
	Correlation within person											
MMCMB	0.9	15.44	20%	0.01	1.17	0.00	11.09	-14%	1.00	15.07	17%	1.00
	0.7	15.13	18%	0.99	1.65	0.80	10.81	-16%	0.90	14.34	11%	0.96
	0.5	15.03	17%	1.95	2.28	0.87	10.93	-15%	0.90	14.42	12%	0.92
	0.3	14.92	16%	2.68	3.23	0.96	11.04	-14%	0.94	14.64	14%	0.97
ANCOVA	0.9	15.45	20%	0.03	1.17	0.00	11.09	-14%	1.00	15.06	17%	1.00
	0.7	16.56	29%	0.95	1.70	0.35	12.08	-6%	1.00	15.81	23%	0.67
	0.5	20.00	56%	1.72	2.27	0.03	14.92	16%	0.91	19.29	50%	0.14
	0.3	23.66	84%	1.96	2.95	0.00	17.92	39%	0.59	22.90	78%	0.00
Scaled Change	0.9	15.44	20%	0.02	1.17	0.00	11.09	-14%	1.00	15.07	17%	1.00
	0.7	15.02	17%	2.52	3.05	0.94	10.94	-15%	0.92	13.57	6%	0.98
	0.5	15.59	21%	4.38	5.04	0.94	11.55	-10%	0.97	14.11	10%	0.97
	0.3	15.82	23%	8.34	7.76	0.92	11.79	-8%	0.95	14.32	11%	0.93

*Standard deviations of the means and mean standard errors are not presented for random effects scenario 2 or 3 because these statistics are nearly identical across all the random effects.

Figure 35. Associations between estimates for the longitudinal association between male sex and CAC score from MMCMB and from competing models under random effects scenario 1, with identical errors applied to the "true" simulated data. Best fit lines are shown in red and black. The "true" value of the coefficient of interest is shown in green. Plots are based on 100 replications.

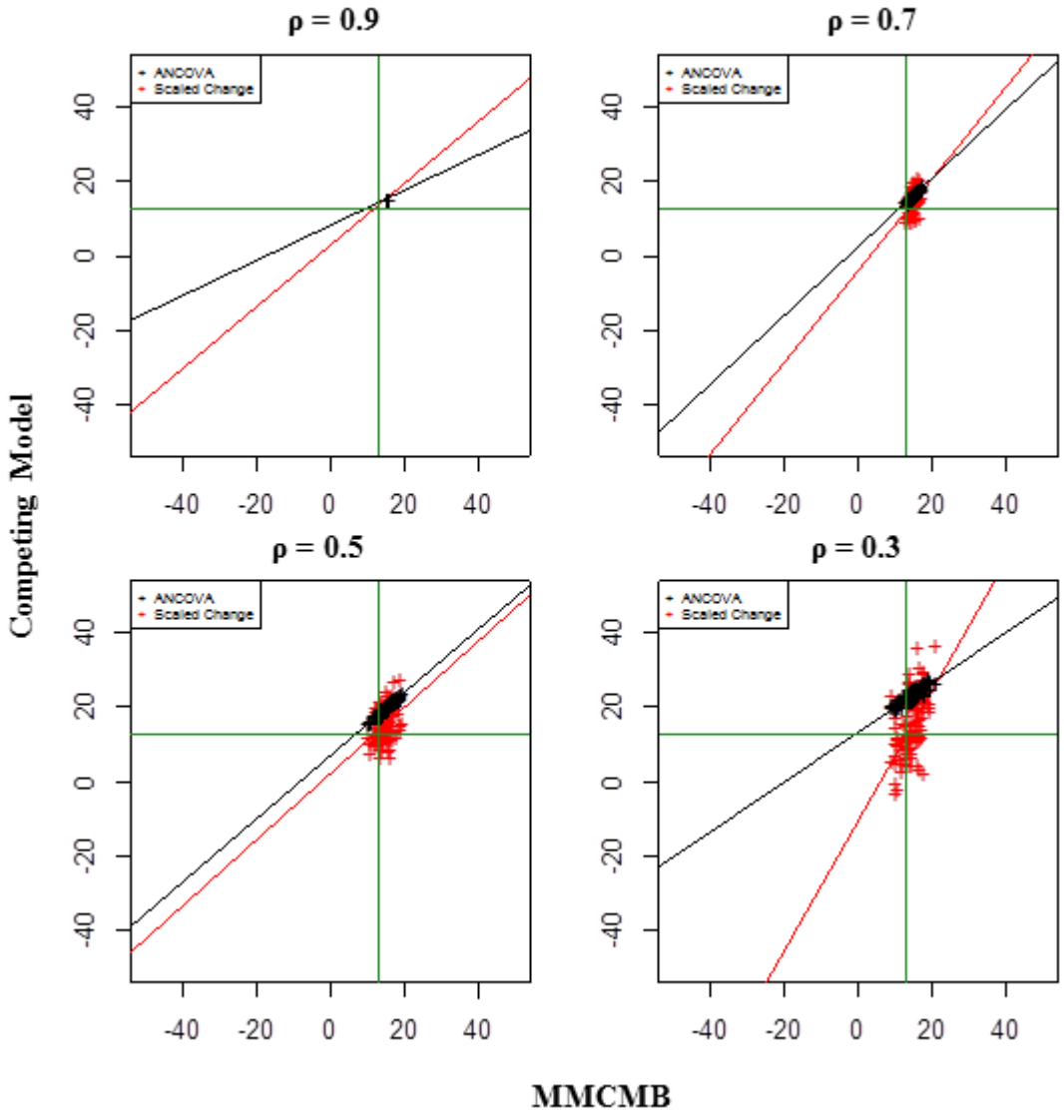


Figure 36. Associations between estimates for the longitudinal association between male sex and CAC score from MMCMB and from competing models under random effects scenario 2, with identical errors applied to the "true" simulated data. Best fit lines are shown in red and black. The "true" value of the coefficient of interest is shown in green. Plots are based on 100 replications.

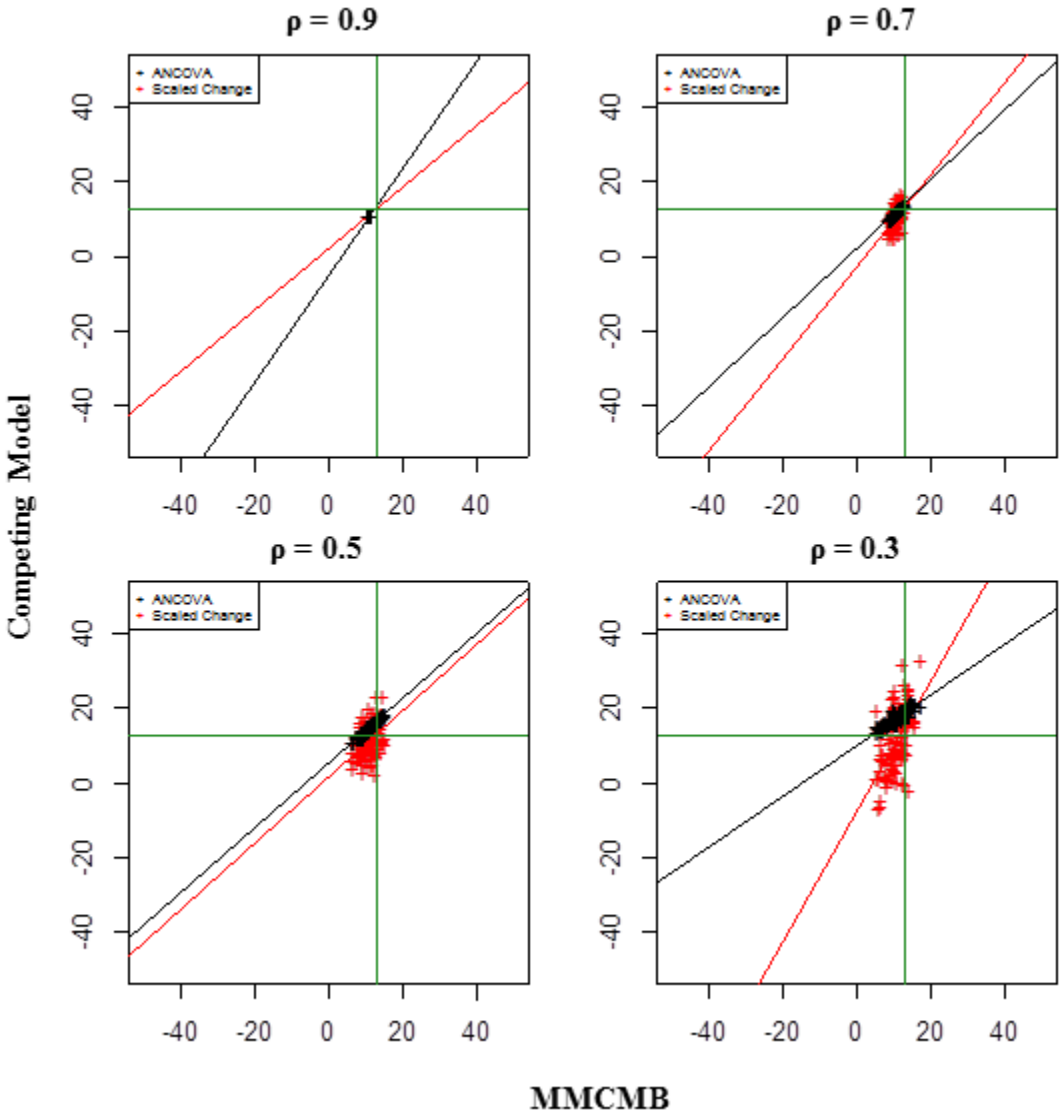
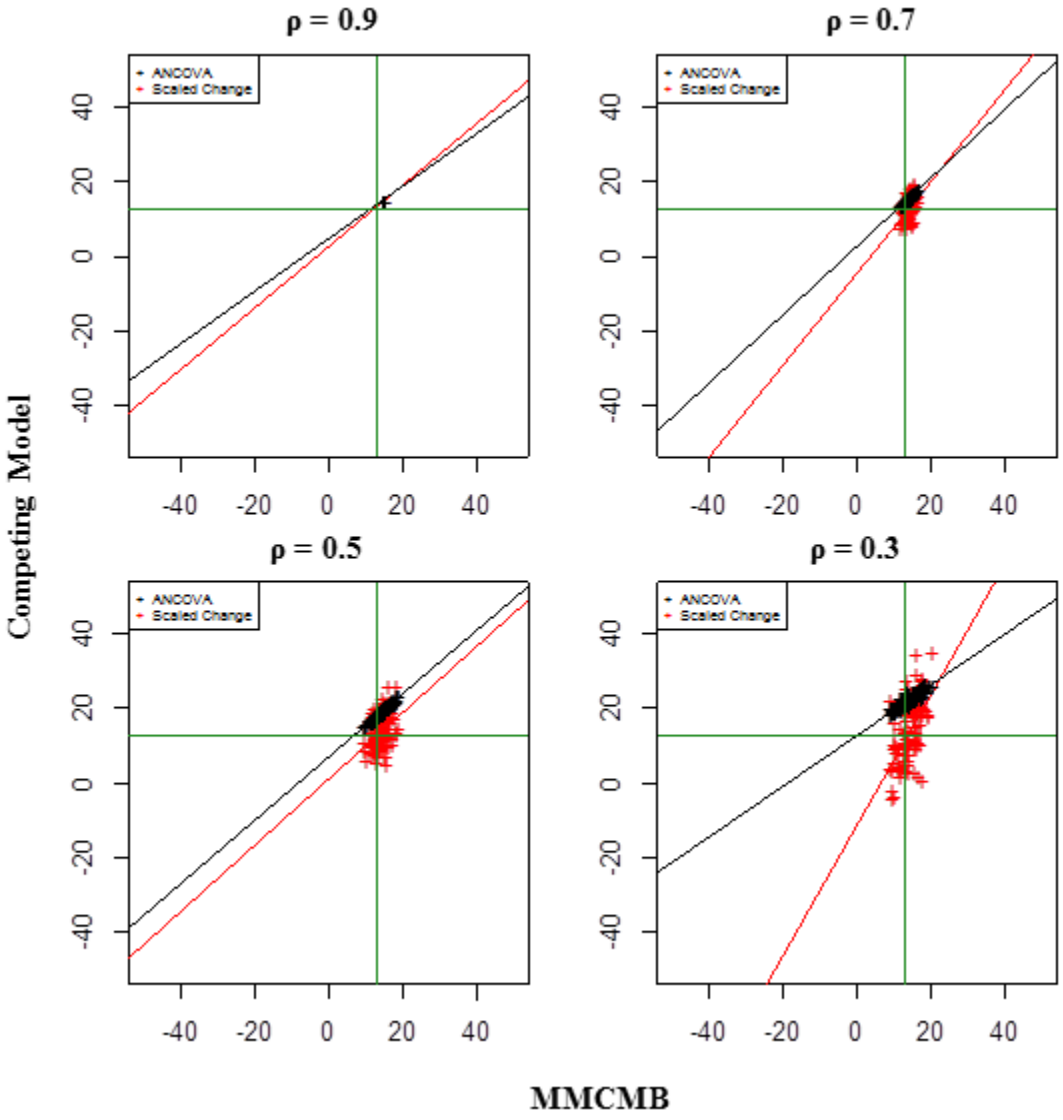


Figure 37. Associations between estimates for the longitudinal association between male sex and CAC score from MMCMB and from competing models under random effects scenario 3, with identical errors applied to the "true" simulated data. Best fit lines are shown in red and black. The "true" value of the coefficient of interest is shown in green. Plots are based on 100 replications.



B. Examples of Mixed Model Analysis Code

Once the model is clearly understood, coding the mixed model is straightforward. Here is an example of R code, which corresponds to MMCMB (23 II):

```
/* We use lmer from the lme4 package */
require(lme4)
result <- summary(lmer(
  score ~ gender + as.factor(site) + as.factor(race) +
            bmi1 + sbp1 + hdl1 + chol1 + age   +
            years + years:gender + years:as.factor(site) +
            years:as.factor(race) + years:bmi2 + years:age +
            years:chol2 + years:hdl2 + years:sbp2 +
            (1 + years | ppt_id), data = sim.data))
```

Baseline covariate values are in columns ending with the numeral "1". Follow-up values are in columns ending with the numeral "2". The first two lines of the call to `lmer` model the baseline. The next three lines estimate slope parameters. Random intercepts and slopes are specified in parenthesis. The `lmer` default behavior is to estimate an unstructured covariance for the random effects. Baseline and follow-up measurements are modeled jointly.

Similarly, this is an example of SAS code:

```
/* Sorting the data sets the reference category */
proc sort data = sim_data; by descending race ppt_id; run;

/* Here we use proc mixed, but hpmixed is an option for complex */
/* random effect models                                         */
proc mixed data = sim_data order = data noclprint = 10;
  class site ppt_id race;
  model score = age gender race sbp1 bmi1 chol1 hdl1 site
                years age*years gender*years race*years sbp2*years
                bmi2*years chol2*years hdl2*years site*years
                / solution;
  random int years /  subject = ppt_id type = un;
  ods output SolutionF = c;
run;
```

Again, the first line of the `model` statement models the baseline, and the next two lines estimate the slope parameters. Random effects are specified in the `random` statement. The `proc mixed` default behavior is to estimate an independent variance-covariance matrix for the random effects, so we must specify that we want an unstructured covariance in order to allow a nonzero covariance parameter to be estimated.

C. Simulation Code

```
############################
# Generates Random Effects #
############################

sim.data<-sqlFetch(channel, "tr0006.marg_all")
sim.wide<-sim.data[sim.data$days > 0,]

ct.exam1 <- sim.data[sim.data$days == 0,]
n1<-length(ct.exam1[,1])

test <- mvrnorm(n=n1, mu = c(0,0), Sigma = matrix(c(81647, 5886, 5886, 1076),2,2),
empirical = TRUE)
# Residual variance 8889

# 1
test <- mvrnorm(n=n1, mu = c(0,0), Sigma = matrix(c(81647, 5886, 5886, 1076),2,2),
empirical = TRUE)
test <- as.data.frame(test)
names(test) <- c("r_int", "r_slope")
ct.exam1$r_int1   <- test$r_int
ct.exam1$r_slope1 <- test$r_slope

# 2
test <- mvrnorm(n=n1, mu = c(0,0), Sigma = matrix(c(81647, 5886, 5886, 1076),2,2),
empirical = TRUE)
test <- as.data.frame(test)
names(test) <- c("r_int", "r_slope")
ct.exam1$r_int2   <- test$r_int
ct.exam1$r_slope2 <- test$r_slope

# 3
test <- mvrnorm(n=n1, mu = c(0,0), Sigma = matrix(c(81647, 5886, 5886, 1076),2,2),
empirical = TRUE)
test <- as.data.frame(test)
names(test) <- c("r_int", "r_slope")
ct.exam1$r_int3   <- test$r_int
ct.exam1$r_slope3 <- test$r_slope

mini.reff<-
ct.exam1[,c("ppt_id","r_int1","r_int2","r_int3","r_slope1","r_slope2","r_slope3")]

#################################################################
# Take marginal predictions, add random effects, then add noise #
#################################################################


sim.data<-sim.data[,1:21]
sim.wide<-sim.wide[,1:21]

sim.data<-merge(sim.data, mini.reff, by="ppt_id")
sim.wide<-merge(sim.wide, mini.reff, by="ppt_id")
sim.test<-sim.wide[sim.wide$fu_days > 7,]
sim.test$sim1 <- sim.test$marg_pred + sim.test$r_int2
sim.test$sim2 <- sim.test$combined2 + sim.test$r_int2 +
sim.test$r_slope2*sim.test$fu_days


# R ~ .9
test.a<-sim.test$sim1 + rnorm(3191, sd = 25)
test.b<-sim.test$sim2 + rnorm(3191, sd = 25)
sqrt(summary(lm(test.b ~ test.a))$r.squared)
```

```
# R ~ .7
test.a<-sim.test$sim1 + rnorm(3191, sd = 200)
test.b<-sim.test$sim2 + rnorm(3191, sd = 200)
sqrt(summary(lm(test.b ~ test.a))$r.squared)

# R ~ .5
test.a<-sim.test$sim1 + rnorm(3191, sd = 350)
test.b<-sim.test$sim2 + rnorm(3191, sd = 350)
sqrt(summary(lm(test.b ~ test.a))$r.squared)

# R ~ .3
test.a<-sim.test$sim1 + rnorm(3191, sd = 550)
test.b<-sim.test$sim2 + rnorm(3191, sd = 550)
sqrt(summary(lm(test.b ~ test.a))$r.squared)
plot(test.a~test.b)

#sqlSave(channel, sim.data, "tr0006.sim_data_all")

sim.data$err9<-sim.data$sim_score_true + rnorm(length(sim.data[,1]), sd = 25)
sim.data$err7<-sim.data$sim_score_true + rnorm(length(sim.data[,1]), sd = 200)
sim.data$err5<-sim.data$sim_score_true + rnorm(length(sim.data[,1]), sd = 350)
sim.data$err3<-sim.data$sim_score_true + rnorm(length(sim.data[,1]), sd = 550)

summary(lmer(sim_score_true ~ days + gender + days:gender + (1 + days | ppt_id), data
= sim.data))@coefs[,1:2]

############################
# Mixed Model Simulation #
############################
#sim.data<-sqlFetch(channel, "tr0006.sim_data_all")
#sim.data$fu.ind<-(sim.data$days>0)*9.59

# ------------- #
#  "true" data  #
# ------------- #
sim.data$sim_score_true1 <- sim.data$marg_pred + sim.data$r_int1 +
sim.data$days*sim.data$r_slope1
sim.data$sim_score_true2 <- sim.data$marg_pred + sim.data$r_int2 +
sim.data$days*sim.data$r_slope2
sim.data$sim_score_true3 <- sim.data$marg_pred + sim.data$r_int3 +
sim.data$days*sim.data$r_slope3

sim.wide$sim1.1<-sim.wide$marg1 + sim.wide$r_int1
sim.wide$sim1.2<-sim.wide$marg_pred + sim.wide$r_int1 +
sim.wide$days*sim.wide$r_slope1

sim.wide$sim2.1<-sim.wide$marg1 + sim.wide$r_int2
sim.wide$sim2.2<-sim.wide$marg_pred + sim.wide$r_int2 +
sim.wide$days*sim.wide$r_slope2

sim.wide$sim3.1<-sim.wide$marg1 + sim.wide$r_int3
sim.wide$sim3.2<-sim.wide$marg_pred + sim.wide$r_int3 +
sim.wide$days*sim.wide$r_slope3

# -------------------- #
#  "true" association  #
# -------------------- #

summary(lmer(sim_score_true1 ~ days*gender + days*as.factor(site) +
days*as.factor(race) + days*age +
                  days:bmi2 + days:chol2 + days:hdl2 + days*sbp2 + hdl1 + bmi1 + chol1
+ (1 + days | ppt_id), data = sim.data))@coefs
```

```
summary(lmer(sim_score_true2 ~ days*gender + days*as.factor(site) +
days*as.factor(race) + days*age +
                  days:bmi2 + days:chol2 + days:hdl2 + days*sbp2 + hdl1 + bmi1 + chol1
+ (1 + days | ppt_id), data = sim.data))@coefs
summary(lmer(sim_score_true3 ~ days*gender + days*as.factor(site) +
days*as.factor(race) + days*age +
                  days:bmi2 + days:chol2 + days:hdl2 + days*sbp2 + hdl1 + bmi1 + chol1
+ (1 + days | ppt_id), data = sim.data))@coefs


# --------------------------- #
# instantiate result matrices #
# --------------------------- #

result1      <- matrix(NA, ncol=8,  nrow=5000)
result2      <- matrix(NA, ncol=8,  nrow=5000)
result3      <- matrix(NA, ncol=8,  nrow=5000)
result1r2.f  <- matrix(NA, ncol=8,  nrow=5000)
result2r2.f  <- matrix(NA, ncol=8,  nrow=5000)
result3r2.f  <- matrix(NA, ncol=8,  nrow=5000)
result1r2.r  <- matrix(NA, ncol=8,  nrow=5000)
result2r2.r  <- matrix(NA, ncol=8,  nrow=5000)
result3r2.r  <- matrix(NA, ncol=8,  nrow=5000)
result1r2.i  <- matrix(NA, ncol=8,  nrow=5000)
result2r2.i  <- matrix(NA, ncol=8,  nrow=5000)
result3r2.i  <- matrix(NA, ncol=8,  nrow=5000)
result.sl1   <- matrix(NA, ncol=16, nrow=5000)
result.sl2   <- matrix(NA, ncol=16, nrow=5000)
result.sl3   <- matrix(NA, ncol=16, nrow=5000)
result.sc1   <- matrix(NA, ncol=8,  nrow=5000)
result.sc2   <- matrix(NA, ncol=8,  nrow=5000)
result.sc3   <- matrix(NA, ncol=8,  nrow=5000)
result.an1   <- matrix(NA, ncol=8,  nrow=5000)
result.an2   <- matrix(NA, ncol=8,  nrow=5000)
result.an3   <- matrix(NA, ncol=8,  nrow=5000)
result.sc1r2 <- matrix(NA, ncol=8,  nrow=5000)
result.sc2r2 <- matrix(NA, ncol=8,  nrow=5000)
result.sc3r2 <- matrix(NA, ncol=8,  nrow=5000)
result.an1r2 <- matrix(NA, ncol=8,  nrow=5000)
result.an2r2 <- matrix(NA, ncol=8,  nrow=5000)
result.an3r2 <- matrix(NA, ncol=8,  nrow=5000)

temp.new     <- as.data.frame(matrix(NA, nrow=length(unique(sim.data$ppt_id)),
ncol=5))
temp.new[,1] <- unique(sim.data$ppt_id)
names(temp.new)<-c("ppt_id","err9","err7","err5","err3")
temp.fu      <- sim.data[sim.data$days>0,c("ppt_id","days")]
sl.int       <- sim.data[sim.data$days == 0,]
nppts <- length(temp.new[,1])
nfu   <- length(temp.fu[,1])

# --------------------------- #
#        error objects        #
# --------------------------- #

errs0.9 <- matrix(NA, nrow=nppts,  ncol=5000)
errs0.7 <- matrix(NA, nrow=nppts,  ncol=5000)
errs0.5 <- matrix(NA, nrow=nppts,  ncol=5000)
errs0.3 <- matrix(NA, nrow=nppts,  ncol=5000)

errs1.9 <- matrix(NA, nrow=nfu,  ncol=5000)
errs1.7 <- matrix(NA, nrow=nfu,  ncol=5000)
errs1.5 <- matrix(NA, nrow=nfu,  ncol=5000)
```

```
errs1.3 <- matrix(NA, nrow=nfu,  ncol=5000)

for (i in 1:5000){
  # generate errors
  errs0.9[,i] <- rnorm(nppts, sd = 25)
  errs0.7[,i] <- rnorm(nppts, sd = 200)
  errs0.5[,i] <- rnorm(nppts, sd = 350)
  errs0.3[,i] <- rnorm(nppts, sd = 550)

  errs1.9[,i] <- rnorm(nfu, sd = 25)
  errs1.7[,i] <- rnorm(nfu, sd = 200)
  errs1.5[,i] <- rnorm(nfu, sd = 350)
  errs1.3[,i] <- rnorm(nfu, sd = 550)
}

# -------------------------- #
#        simulation          #
# -------------------------- #

# track time
a1<-proc.time()

for (i in 1:100){
  # generate errors
  temp.new[,2] <- errs0.9[,i]
  temp.new[,3] <- errs0.7[,i]
  temp.new[,4] <- errs0.5[,i]
  temp.new[,5] <- errs0.3[,i]

  temp1.9 <- errs1.9[,i]
  temp1.7 <- errs1.7[,i]
  temp1.5 <- errs1.5[,i]
  temp1.3 <- errs1.3[,i]

  sim.data$err1.9[sim.data$days > 0] <- sim.data$sim_score_true1[sim.data$days > 0] +
temp1.9
  sim.data$err1.7[sim.data$days > 0] <- sim.data$sim_score_true1[sim.data$days > 0] +
temp1.7
  sim.data$err1.5[sim.data$days > 0] <- sim.data$sim_score_true1[sim.data$days > 0] +
temp1.5
  sim.data$err1.3[sim.data$days > 0] <- sim.data$sim_score_true1[sim.data$days > 0] +
temp1.3

[Similar code removed]

  sim.data$err1.9[sim.data$days == 0] <- sim.data$sim_score_true1[sim.data$days == 0]
+ temp.new[,2]
  sim.data$err1.7[sim.data$days == 0] <- sim.data$sim_score_true1[sim.data$days == 0]
+ temp.new[,3]
  sim.data$err1.5[sim.data$days == 0] <- sim.data$sim_score_true1[sim.data$days == 0]
+ temp.new[,4]
  sim.data$err1.3[sim.data$days == 0] <- sim.data$sim_score_true1[sim.data$days == 0]
+ temp.new[,5]

[Similar code removed]

  sim.wide$err1.9.2 <- sim.wide$sim1.2 + temp1.9
  sim.wide$err1.7.2 <- sim.wide$sim1.2 + temp1.7
  sim.wide$err1.5.2 <- sim.wide$sim1.2 + temp1.5
  sim.wide$err1.3.2 <- sim.wide$sim1.2 + temp1.3

[Similar code removed]
```

```
  temp.new2 <- merge(sim.wide, temp.new, by="ppt_id")
  sim.wide$err1.9.1 <- sim.wide$sim1.1 + temp.new2$err9
  sim.wide$err1.7.1 <- sim.wide$sim1.1 + temp.new2$err7
  sim.wide$err1.5.1 <- sim.wide$sim1.1 + temp.new2$err5
  sim.wide$err1.3.1 <- sim.wide$sim1.1 + temp.new2$err3

[Similar code removed]

  print(proc.time()-a1)

  # MMCMB, RE 1

  result1[i,1:2]<-summary(lmer(err1.9 ~ days*gender + days*as.factor(site) +
days*as.factor(race) + days*age +
                  days:bmi2 + days:chol2 + days:hdl2 + days:sbp2 + hdl1 + bmi1 + chol1
+ sbp1 + (1 + days | ppt_id), data = sim.data))@coefs[29,1:2]
  result1[i,3:4]<-summary(lmer(err1.7 ~ days*gender + days*as.factor(site) +
days*as.factor(race) + days*age +
                  days:bmi2 + days:chol2 + days:hdl2 + days:sbp2 + hdl1 + bmi1 + chol1
+ sbp1 + (1 + days | ppt_id), data = sim.data))@coefs[29,1:2]
  result1[i,5:6]<-summary(lmer(err1.5 ~ days*gender + days*as.factor(site) +
days*as.factor(race) + days*age +
                  days:bmi2 + days:chol2 + days:hdl2 + days:sbp2 + hdl1 + bmi1 + chol1
+ sbp1 + (1 + days | ppt_id), data = sim.data))@coefs[29,1:2]
  result1[i,7:8]<-summary(lmer(err1.3 ~ days*gender + days*as.factor(site) +
days*as.factor(race) + days*age +
                  days:bmi2 + days:chol2 + days:hdl2 + days:sbp2 + hdl1 + bmi1 + chol1
+ sbp1 + (1 + days | ppt_id), data = sim.data))@coefs[29,1:2]

  # MMCMB, RE 2
  # MMCMB, RE 3
[Similar code removed]

  # MMCMB, RE 1, REDUCED TO 2 TIME POINTS

  result1r2.f[i,1:2]<-summary(lmer(err1.9 ~ days*gender + days*as.factor(site) +
days*as.factor(race) + days*age +
                  days:bmi2 + days:chol2 + days:hdl2 + days:sbp2 + hdl1 + bmi1 + chol1
+ sbp1 + (1 + days | ppt_id), data = sim.data, subset = (days == 0 | days >
7)))@coefs[29,1:2]
  result1r2.f[i,3:4]<-summary(lmer(err1.7 ~ days*gender + days*as.factor(site) +
days*as.factor(race) + days*age +
                  days:bmi2 + days:chol2 + days:hdl2 + days:sbp2 + hdl1 + bmi1 + chol1
+ sbp1 + (1 + days | ppt_id), data = sim.data, subset = (days == 0 | days >
7)))@coefs[29,1:2]
  result1r2.f[i,5:6]<-summary(lmer(err1.5 ~ days*gender + days*as.factor(site) +
days*as.factor(race) + days*age +
                  days:bmi2 + days:chol2 + days:hdl2 + days:sbp2 + hdl1 + bmi1 + chol1
+ sbp1 + (1 + days | ppt_id), data = sim.data, subset = (days == 0 | days >
7)))@coefs[29,1:2]
  result1r2.f[i,7:8]<-summary(lmer(err1.3 ~ days*gender + days*as.factor(site) +
days*as.factor(race) + days*age +
                  days:bmi2 + days:chol2 + days:hdl2 + days:sbp2 + hdl1 + bmi1 + chol1
+ sbp1 + (1 + days | ppt_id), data = sim.data, subset = (days == 0 | days >
7)))@coefs[29,1:2]

  # MMCMB, RE 2, REDUCED TO 2 TIME POINTS

[Similar code removed]

  # MMCMB, RE 3, REDUCED TO 2 TIME POINTS

[Similar code removed]
```

```
  # Difference
    sim.wide$diff1.9<-sim.wide$err1.9.2 - sim.wide$err1.9.1
    sim.wide$diff1.7<-sim.wide$err1.7.2 - sim.wide$err1.7.1
    sim.wide$diff1.5<-sim.wide$err1.5.2 - sim.wide$err1.5.1
    sim.wide$diff1.3<-sim.wide$err1.3.2 - sim.wide$err1.3.1

[Similar code removed]

  # Scaled Difference
    sim.wide$scdiff1.9<-(sim.wide$err1.9.2 - sim.wide$err1.9.1)/sim.wide$days
    sim.wide$scdiff1.7<-(sim.wide$err1.7.2 - sim.wide$err1.7.1)/sim.wide$days
    sim.wide$scdiff1.5<-(sim.wide$err1.5.2 - sim.wide$err1.5.1)/sim.wide$days
    sim.wide$scdiff1.3<-(sim.wide$err1.3.2 - sim.wide$err1.3.1)/sim.wide$days

[Similar code removed]

result.sc1[i,1:2]<-summary(lmer(scdiff1.9 ~ gender  + as.factor(site) +
as.factor(race) + sbp2 + hdl2 + age + chol2 + bmi2 + (1 | ppt_id), data =
sim.wide))@coefs[12,1:2]
    result.sc1[i,3:4]<-summary(lmer(scdiff1.7 ~ gender  + as.factor(site) +
as.factor(race) + sbp2 + hdl2 +
         age + chol2 + bmi2 + (1 | ppt_id), data = sim.wide))@coefs[12,1:2]
    result.sc1[i,5:6]<-summary(lmer(scdiff1.5 ~ gender  + as.factor(site) +
as.factor(race) + sbp2 + hdl2 +
         age + chol2 + bmi2 + (1 | ppt_id), data = sim.wide))@coefs[12,1:2]
    result.sc1[i,7:8]<-summary(lmer(scdiff1.3 ~ gender  + as.factor(site) +
as.factor(race) + sbp2 + hdl2 +
         age + chol2 + bmi2 + (1 | ppt_id), data = sim.wide))@coefs[12,1:2]

[Similar code removed]

  # Reduced to 2 time points: Scaled Difference


  # ANCOVA
  result.an1[i,1:2]<-summary(lmer(err1.9.2 ~ err1.9.1 + days + days:gender +
days:as.factor(site) + days:as.factor(race) + days:age +
                  days:bmi2 + days:chol2 + days:hdl2  + days:sbp2   + (1 + days |
ppt_id), data = sim.wide))@coefs[16,1:2]
  result.an1[i,3:4]<-summary(lmer(err1.7.2 ~ err1.7.1 + days + days:gender +
days:as.factor(site) + days:as.factor(race) + days:age +
                  days:bmi2 + days:chol2 + days:hdl2  + days:sbp2   + (1 + days |
ppt_id), data = sim.wide))@coefs[16,1:2]
  result.an1[i,5:6]<-summary(lmer(err1.5.2 ~ err1.5.1 + days + days:gender +
days:as.factor(site) + days:as.factor(race) + days:age +
                  days:bmi2 + days:chol2 + days:hdl2  + days:sbp2   + (1 + days |
ppt_id), data = sim.wide))@coefs[16,1:2]
  result.an1[i,7:8]<-summary(lmer(err1.3.2 ~ err1.3.1 + days + days:gender +
days:as.factor(site) + days:as.factor(race) + days:age +
                  days:bmi2 + days:chol2 + days:hdl2  + days:sbp2   + (1 + days |
ppt_id), data = sim.wide))@coefs[16,1:2]

[Similar code removed]

}
a2<-proc.time()
a2-a1
```

7. References

1. Fitzmaurice GM, Laird NM, Ware JH. Applied Longitudinal Analysis. Second Edition ed. Hoboken, NJ: John Wiley & Sons, Inc; 2011.
2. Council NR. The Prevention and Treatment of Missing Data in Clinical Trials. Washington, DC: The National Academies Press; 2010.
3. Frison L, Pocock SJ. Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. Stat Med 1992;11:1685-704.
4. Yanez ND, Kronmal RA, Shemanski LR. The effects of measurement error in response variables and tests of association of explanatory variables in change models. Statistics in Medicine 1998;17.
5. Fitmaurice GM, Laird NM, Ware JH. Applied Longitudinal Analysis. Second Edition ed. Hoboken, NJ: John Wiley & Sons, Inc; 2011.
6. Frison LJ, Pocock SJ. Linearly divergent treatment effects in clinical trials with repeated measures: efficient analysis using summary statistics. Stat Med 1997;16:2855-72.
7. Diggle PJ, Heagerty P, Liang K-Y, Zeger SL. Analysis of Longitudinal Data. Second Edition ed. United States: Oxford Univeristy Press; 2002.
8. Hodges JS, Reich BJ. Adding Spatially-Correlated Errors Can Mess Up the Fixed Effect You Love. American Statistical Association 2010;64:325-34.
9. Bild DE, Bluemke DA, Burke GL, et al. Multi-ethnic study of atherosclerosis: objectives and design. Am J Epidemiol 2002;156:871-81.
10. Agatston AS, Janowitz WR, Hildner FJ, Zusmer NR, Viamonte M, Detrano R. Quantification of coronary artery calcium using ultrafast computed tomography. J Am Coll Cardiol 1990;15:827-32.
11. Greenland P. Coronary artery calcium score combined with Framingham score for risk prediction in asymptomatic individuals (vol 291, pg 210, 2004). Jama-Journal of the American Medical Association 2004;291:563-.
12. Detrano R, Guerci AD, Carr JJ, et al. Coronary calcium as a predictor of coronary events in four racial or ethnic groups. New England Journal of Medicine 2008;358.
13. O'Malley PG, Taylor AJ, Jackson JL, Doherty TM, Detrano RC. Prognostic value of coronary electron-beam computed tomography for coronary heart disease events in asymptomatic populations. American Journal of Cardiology 2000;85:945-8.
14. Kondos GT, Hoff JA, Sevrukov A, et al. Electron-beam tomography coronary artery calcium and cardiac events - A 37-month follow-up of 5635 initially asymptomatic low- to intermediate-risk adults. Circulation 2003;107:2571-6.
15. Budoff MJ, Nasir K, McClelland RL, et al. Coronary Calcium Predicts Events Better With Absolute Calcium Scores Than Age-Sex-Race/Ethnicity Percentiles MESA (Multi-Ethnic Study of Atherosclerosis). Journal of the American College of Cardiology 2009;53:345-52.

16. Leening MJG, Elias-Smale SE, Kavousi M, et al. Coronary Calcification and the Risk of Heart Failure in the Elderly The Rotterdam Study. Jacc-Cardiovascular Imaging 2012;5:874-80.

17. Raggi P, Callister TQ, Shaw LJ. Progression of coronary artery calcium and risk of first myocardial infarction in patients receiving cholesterol-lowering therapy. Arterioscler Thromb Vasc Biol 2004;24:1272-7.

18. Budoff MJ, Hokanson JE, Nasir K, et al. Progression of coronary artery calcium predicts all-cause mortality. JACC Cardiovasc Imaging 2010;3:1229-36.

19. Budoff MJ, Raggi P. Coronary artery disease progression assessed by electron-beam computed tomography. Am J Cardiol 2001;88:46E-50E.

20. Budoff MJ, Young R, Lopez VA, et al. Progression of Coronary Calcium and Incident Coronary Heart Disease Events: The Multi-Ethnic Study of Atherosclerosis. J Am Coll Cardiol 2013;61:1231-9.

21. McEvoy JW, Blaha MJ, DeFilippis AP, et al. Coronary Artery Calcium Progression: An Important Clinical Measurement? A Review of Published Reports. Journal of the American College of Cardiology 2010;56.

22. Kronmal RA, McClelland RL, Detrano R, et al. Risk factors for the progression of coronary artery calcification in asymptomatic subjects - Results from the Multi-Ethnic Study of Atherosclerosis (MESA). Circulation 2007;115.

23. Liang CJ, Budoff MJ, Kaufman JD, Kronmal RA, Brown ER. An alternative method for quantifying coronary artery calcification: the multi-ethnic study of atherosclerosis (MESA). BMC Med Imaging 2012;12:14.

24. Adar SD, Sheppard L, Vedal S, et al. Fine particulate air pollution and the progression of carotid intima-medial thickness: a prospective cohort study from the multi-ethnic study of atherosclerosis and air pollution. PLoS Med 2013;10:e1001430.