Estimation of an Exposure Effect on Outcome Rate of Change in Observational Study Settings

Cooper S. Schumacher

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Science

University of Washington
2017

Committee:
Lianne Sheppard
Patrick Heagerty

Program Authorized to Offer Degree:
Public Health - Biostatistics

©Copyright 2017
Cooper S. Schumacher

University of Washington

Abstract

Estimation of an Exposure Effect on Outcome Rate of Change in Observational Study Settings

Cooper S. Schumacher

Chair of the Supervisory Committee:
Lianne Sheppard, Professor
Department of Biostatistics

Mixed models for longitudinal analysis have several well-established benefits, including gaining information from unbalanced data and depending on less restrictive assumptions about missing data than other methods such as semi-parametric methods (e.g. generalized estimating equations). While there are multiple longitudinal analysis methods to estimate the effect of exposure on an outcome's rate of change, in many applications linear mixed models with a model for the baseline outcome and longitudinal change may provide the best rate of change estimates. These models also give two distinct exposure effect estimates: the cross-sectional effect at baseline and the longitudinal effect of interest.

We conduct simulations to evaluate characteristics of longitudinal exposure effect estimates from the modeled baseline mixed model (i.e. mixed model with separate cross-sectional and longitudinal effects) and competing methods of rate of change analysis. We also compare the longitudinal and cross-sectional effect estimates from the modeled baseline mixed model. We apply our insights to an analysis of how ambient air pollution affects the progression of a measure of coronary heart disease in a subset of the MESA Air study.

The modeled baseline model avoids bias caused by controlling for the outcome's measured baseline as a covariate when the exposure affects the baseline outcome and the outcome is measured with error. The modeled baseline estimates are also more precise in comparison to models that use scaled change since baseline as the outcome variable. In the modeled baseline model, we find the cross-sectional exposure effect estimate is primarily influenced by characteristics of the baseline measurements, while the longitudinal estimate is influenced by characteristics of both the baseline and temporal changes in the follow-up data. Exposure qualities and model parameterization may induce a correlation in these two exposure estimates. We recommend deemphasizing the cross-sectional estimate and focusing on the longitudinal estimate as the cross-sectional effect estimate is generally more dependent on characteristics of the study design and are expected to be more prone to confounding.

Our work provides strong support for the use of the modeled baseline mixed model in environmental epidemiology and more broadly in many non-randomized longitudinal study contexts.

Contents

1 Introduction

In many scientific contexts we are interested in the progression of a health outcome and the role of environmental exposures on that progression. Longitudinal study designs allow for exposures to be linked with outcome growth or rate of change. One such example is the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air), which investigated an association between long-term ambient air pollution and progression of coronary artery calcification (CAC), a subclinical measure of atherosclerosis [Kaufman et al., 2016]. These analyses were motivated by previous research which has shown both one-time measurements of CAC and progression of CAC (i.e. change in CAC over time) to predict cardiovascular events [McEvoy et al., 2010] [Budoff et al., 2013].

Many longitudinal analysis methods are available for modeling outcome change over time. For simplistic study contexts where only a baseline and one follow-up measure are taken, simple methods may be applied such as regression of change, where the two measures are used to derive a single outcome representing the difference between follow-up and baseline measurements. When the structure of repeated measures is more complex, we require models that can account for these types of study designs. Some of the most common longitudinal methods include mixed models, marginal models, and transition models. We will focus on mixed models. Marginal models estimated using semi-parametric methods [Liang and Zeger, 1986] require that missing data are missing completely at random (MCAR)[Fitzmaurice et al., 2012], which is unreasonable for the study contexts we are interested in. Note that generalized estimating equations can be used with missing at random (MAR) data, a less restrictive assumption than MCAR [Fitzmaurice et al., 2012], but would require inverse probability weights that account for missing data. We also do not consider transition models as they are not appropriate for unbalanced data (i.e. data where subjects do not have the same number of measures and follow-up timing) [Colosimo et al.,

2012]. Furthermore, transition models condition on past health outcomes and these are often intermediate outcomes for chronic exposure effects. Mixed models allow for unbalanced data and only requires that data are MAR, given that the full multivariate model including longitudinal mean model and covariance structure are correctly specified [Fitzmaurice et al., 2012].

The selection of a good statistical model depends on many factors of the study context such as study design, target population and characteristics of the variables of interest. When there are more than one follow-up measure per subject or the timing of follow-ups varies considerably, methods which estimate a rate of change are usually most appropriate for quantifying the progression of an outcome. In mixed models we include a time covariate in order to consider a linear trend, or constant outcome rate of change, along with interaction terms of covariates with time to consider differences in the outcome rate of change across subpopulations of interest. Another approach to modeling rate of change is to conduct regression on a time-scaled measure of the outcome change since baseline.

In this paper we focus on linear mixed models for rate of change that incorporate a flexible model for the baseline outcome (See Equation 1.1). By using a modeled baseline, estimates avoid bias that can occur from controlling for a measured outcome baseline in an observational study design. In randomized studies, regression of change models may gain power by controlling for the outcome's measured baseline [Van Breukelen, 2006]. However, in a non-randomized setting where the exposure is associated with the baseline outcome, models adjusted for the measured baseline have been shown to produce biased estimates of the exposure effect if the outcome is measured with error [Lepage et al., 2015] [Van Breukelen, 2006] . This bias can be present even if the exposure is measured without error [Yanez et al., 1998]. In the Supplement Section A, we give an explanation of this bias using directed acyclic graphs (DAG's) similar to as is done in Lepage et al. [2015]. In almost all environmen-

tal epidemiological studies we expect the exposure to be associated with the baseline outcome, making this bias a legitimate concern. Much of the literature on this topic has focused primarily on regression of change for study designs with a baseline and follow-up. We will demonstrate the challenges of adjusting for the measured baseline in a more general setting with multiple follow-up times that are not necessarily evenly spaced and we are interested in rate of change.

The form of the linear mixed model with a modeled baseline, the same model used in MESA Air analyses, may be expressed as [Gassett et al., 2015]:

$$Y_{iv} = (\alpha_0 + \alpha_1 E_{i0} + \boldsymbol{\alpha}_2 \boldsymbol{X}_{i0} + a_i) + (\beta_0 + \beta_1 E_{iv} + \boldsymbol{\beta}_2 \boldsymbol{Z}_{iv} + b_i) t_{iv} + (\boldsymbol{\gamma} \boldsymbol{U}_{iv} + \epsilon_{iv}). \quad (1.1)$$

In this model form, Y_{iv} is the outcome for subject i at the vth follow-up measurement; t_{iv} is the follow-up time since baseline; E_{iv} is the exposure of interest, $\boldsymbol{X}_{i0}$ is a set of baseline covariates, $\boldsymbol{Z}_{iv}$ is a set of baseline and/or time-varying covariates; $\boldsymbol{U}_{iv}$ is a set of transient covariate(s); the α's, β's and $\boldsymbol{\gamma}$ are fixed effects; a_i and b_i are a random intercept and slope respectively; and ϵ_{iv} is an error term. The first term in parentheses is a cross-sectional component that models the outcome at baseline. The second term in parentheses is a longitudinal component that includes all of the time interactions and models the outcome rate of change. The third term in parentheses is a transient term, which controls for variables specific to the measurement time in order to account for measurements taken under different conditions. An example of a transient variable is the technology used to measure the outcome. Advances in technology may lead to improved measurement accuracy, but if devices on average produce different measurements, it may be difficult in a longitudinal study to distinguish actual outcome change from technology changes if the variable is not properly accounted for. Note that E_{i0} and E_{iv} represent the exposure variables at baseline and current follow-up time, but that the specification of this variable is flexible. For

example, the exposure at a specific time could be a average of an exposure history under several different time-scales (e.g. short-term, long-term, cumulative etc.). Also note that this model has the interaction $E_{iv} * t_{iv}$ but not a main effect for E_{iv}. This assumes that the slope of the average outcome rate of change is influenced by the most recent exposure levels, and the modeled outcome at baseline is only influenced by the exposure at baseline, E_{i0}, and not by future exposure levels.

The primary goal of our analyses is to make inference on the exposure effect on outcome rate of change, β_1. The cross-sectional/baseline exposure effect α_1 is interpretable as well, but we will argue against it being given too much emphasis in an analysis.

In this paper we conduct simulation studies to investigate how the modeled baseline linear mixed model compares to other competing models in estimation of an exposure effect on rate of change. We further consider how the modeled baseline model performs under various contexts, and clarify the model's feature of producing two exposure estimates. We then demonstrate the use of the modeled baseline mixed model in an analysis of the MESA Air cohort from Baltimore, in order to investigate an association between exposure to long-term ambient air pollution and progression of coronary artery calcium (CAC).

2 Simulation Framework

We designed our simulation studies to address two primary questions. First, we are interested in how the modeled baseline mixed model compares to other methods of estimating an exposure effect on rate of change. Second, we aim to gain deeper insight on the application of the modeled baseline mixed model. Most notably, we consider how the two exposure effect estimates (i.e. cross-sectional/baseline effect, $\hat{\alpha}_1$, and longitudinal/rate of change effect, $\hat{\beta}_1$) are influenced differently by features of the

study design, outcome variable and characteristics of the exposure. We also consider how the two exposure estimates are related.

In our simulations, we consider a simplified outcome model roughly representing CAC and its association with a generated $PM_{2.5}$ exposure and one additional covariate. Our simulated outcome model has the form $Y_{iv} = (\alpha_0 + \alpha_1 E_{i0} + \alpha_2 X_{i0} + a_i) + (\beta_0 + \beta_1 E_{iv} + \beta_2 X_{iv} + b_i)t_{iv} + \epsilon_{iv}$, where the terms have the same definition as for the model (1.1) given in Section 1. The measured outcome has an additional measurement error term (i.e. the measured outcome is $Y^m_{iv} = Y_{iv} + e^m_{iv}$, where $e^m_{iv} \sim \mathrm{N}(0, \sigma^2_{e^m})$), although this error in indistinguishable from the random outcome error ϵ_{iv}. The simulations assume complete data of 4 observations ($v = 0, 1, 2, 3$) which are equally spaced over the same length of follow-up ($t_{i3} = 5$) for each subject. The parameters and uncertainty structures of the model are chosen based on previous analysis of the MESA Air data. The exposure variable, E_{iv}, is composed of three parts: a (linear) temporal decline the same for each subject in the population, a subset-specific average exposure level, and random temporal variation at each measurement time. These exposure and outcome models are not intended to be an accurate representation of the true mechanisms influencing coronary artery calcium, but rather to demonstrate how model performance is influenced by various study contexts. However, we did base parameter values in the data generating model on results from analyses of the subset of MESA Air subjects from Baltimore. More details of the simulation framework are provided in the Supplement Section B.

We compare the performance of our modeled baseline mixed model to rate of change estimates from a measured baseline mixed model and repeated scaled change mixed model (See Table 1). The measured baseline mixed model has the same longitudinal component as the modeled baseline model but replaces the covariates in the modeled baseline component with the measured baseline outcome, Y_{i0}. There are several options in defining this model, such as whether to include the main effects

Model
Modeled Baseline Mixed Model $Y_{iv} = (\alpha_0 + \alpha_1 E_{i0} + \boldsymbol{\alpha}_2 \boldsymbol{X}_{i0} + a_i) + (\beta_0 + \beta_1 E_{iv} + \boldsymbol{\beta}_2 \boldsymbol{Z}_{iv} + b_i) t_{iv} + (\boldsymbol{\gamma} \boldsymbol{U}_{iv} + \epsilon_{iv})$
Measured Baseline Mixed Model $Y_{iv} = (\alpha_0^* + \alpha_1^* Y_{i0} + a_i^*) + (\beta_0^* + \beta_1^* E_{iv} + \boldsymbol{\beta}_2^* \boldsymbol{Z}_{iv} + b_i^*) t_{iv} + (\boldsymbol{\gamma}^* \boldsymbol{U}_{iv} + \epsilon_{iv}^*)$
Repeated Scaled Change Mixed Model $\frac{Y_{iv} - Y_{i0}}{t_{iv}} = (\beta_0^{**} + \beta_1^{**} E_{iv} + \boldsymbol{\beta}_2^{**} \boldsymbol{Z}_{iv} + b_i^{**}) + \epsilon_{iv}^{**}$

Table 1: Model forms of fitted rate of change models used in the simulations and to estimate the effect of particulate matter ($PM_{2.5}$) on CAC rate of change in the Baltimore MESA Air cohort. A description of terms is provided in Section 1.

of the covariates adjusted for in the longitudinal component and how to specify the random effects. However, regardless of theses choices, the model will have the potential for bias in non-randomized settings, and so we do not present these options in detail. A more appropriate alternative rate of change model is the repeated scaled change mixed model. This model considers regression on the time-scaled outcome change since baseline, $\frac{Y_{iv} - Y_{i0}}{t_{iv}}$. Since we are doing regression on the scaled outcome change rather than the outcome itself, we do not require the terms involving t_{iv} on the right side of the equation and only estimate the effect estimates for the longitudinal terms in the model. This model has the advantage of being straightforward and easily interpreted, but also has some drawbacks. The model is not properly specified to adjust for transient variables, because of how the estimated effects in the model are all effects on the scaled change. Therefore, the estimated effects will depend on the time that has passed since baseline. A transient variable effects the outcome at a specific measurement time and should not have an effect that depends on the time since baseline.

We will explore the performance of estimates of the exposure effect on outcome rate of change from each of these three models (i.e. estimates of β_1, β_1^* and β_1^{**} in Table 1). Each of these estimates may be interpreted similarly: as the difference in the outcome rate of change for a one unit increase in the exposure.

3 Simulation Results

We present several simulation results in this section. Further results are presented in the Supplement. A summary of the simulation studies and corresponding sections is provided in Table 5 in Supplement Section C.

3.1 Comparing Exposure Effect Estimates from Rate of Change Mixed Models

3.1.1 Findings

We compare the performance of the mixed model exposure effect estimates in both randomized and observational settings, determined by whether the exposure has an effect on the baseline outcome. Table 2 shows that in the randomized setting, all models provide unbiased estimates of the exposure effect on rate of change. The modeled baseline model has the most precise estimates, while the repeated scaled change estimates are the least precise. In this simulation, the relative efficiency (ratio of $E[SE]^2$ for the repeated scaled change model relative to the modeled baseline model) is 1.59. The estimated standard errors for the repeated scaled change models are slightly lower than the true standard deviation on average, leading to slight under-coverage in the confidence intervals. The modeled baseline and measured baseline models both have accurate estimates of the estimator precision and have proper confidence interval coverage.

In an observational setting, the measured baseline estimate experiences some bias. The precision of the estimate is similar to that in the randomized setting, and thus the confidence interval coverage suffers. The modeled baseline and repeated scaled change models each perform similarly with regard to accuracy and precision in the randomized and observational settings.

In Figure 1, we see that the magnitude of the measured baseline estimate's bias

Randomized Trial

Model	Estimate	SD	E[SE]	95% CI Coverage
Modeled Baseline	4.98	1.43	1.44	0.95
Measured Baseline	5.00	1.68	1.67	0.95
Repeated Scaled Change	4.98	1.92	1.81	0.94

Non-Randomized Trial

Model	Estimate	SD	E[SE]	95% CI Coverage
Modeled Baseline	4.98	1.43	1.44	0.95
Measured Baseline	4.51	1.69	1.67	0.93
Repeated Scaled Change	4.98	1.92	1.81	0.94

Table 2: Estimated exposure effects on rate of change (estimating a true parameter $\beta_1 = 5$), standard deviation (SD), average estimated standard error (E[SE]) and 95% confidence interval coverage from the modeled baseline, measured baseline and repeated scaled change mixed models. Both randomized (i.e. $\alpha_1 = 0$) and non-randomized (i.e. $\alpha_1 = 50$) settings were simulated.

depends on the outcome error. When there is a small random intercept (interpreted as natural variation across subjects unexplained by the covariates), the bias becomes larger in the positive direction when the error increases (Figure 1C). With a large random intercept (Figure 1A), the bias of the modeled baseline estimate is negative for lower levels of the outcome and moves in the positive direction as the error increases, but the influence of the error is not as strong as in Figure 1C. The modeled baseline and repeated scaled change estimates are unbiased for all levels of outcome error.

The standard error of the estimates from each model increases as the level of outcome error increases. The modeled baseline model always has the smallest standard error. The repeated scaled change model has the largest standard error, and the difference in precision compared to the modeled baseline is largest when there is more outcome error. With an outcome error of 20,000 and the random effects structure of Figures 1A/1B, the relative efficiency is 1.59 since it is the same outcome error as in the simulations used for Table 2. However, if the outcome error is increased to 100,000, the relative efficiency becomes 1.86.

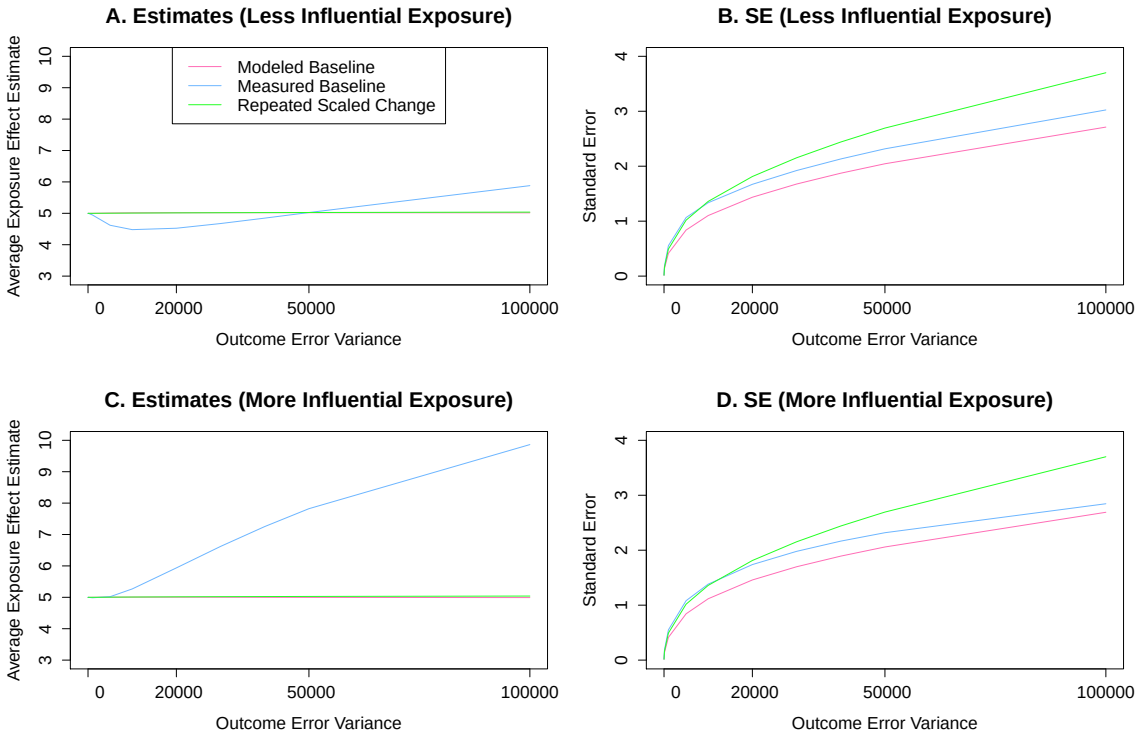

Figure 1: Influence of outcome error on the bias and precision of rate of change exposure effect estimates from the modeled baseline, measured baseline and repeated scaled change mixed models. Average exposure effect estimates (estimating a true value of $\beta_1 = 5$) are provided in the left two plots and standard error estimates are provided in the right two plots. Note that outcome error can be either outcome residual variance (natural randomness in the outcome variable) or outcome measurement error (due to inaccurate measurement of the true outcome value). We consider two different scenarios. In **A** and **B** there is more random variation in the baseline outcome and the exposure contributes very little to the total outcome variation (approximately 0.7% to 1.1% of the total variation of the outcome baseline). In **C** and **D** the exposure is the same, but there is less random variation in the baseline outcome (achieved in the simulations by lowering the variance of the random intercept) and so the exposure is more influential (contributes approximately 1.7% to 9.6% of the total variance of the baseline outcome). Note that in **A** and **C**, the modeled baseline and repeated scaled change estimates are unbiased and so the dashed line showing the true value of $\beta_1 = 5$ and the line for the modeled baseline model are both covered by the line for the repeated scaled change estimates.

3.1.2 Discussion

Consistent with the literature [Lepage et al., 2015] [Van Breukelen, 2006] [Yanez et al., 1998], adjusting for the measured baseline in the presence of outcome error and an exposure which affects the baseline outcome leads to biased exposure estimates. The magnitude of this bias is influenced by many factors, such as the amount of outcome measurement error or how much of the variation in the outcome is explained by the exposure. When there is little "randomness" in the outcome unexplained by the covariates (Figures 1C and 1D) and the exposure explains a large portion of the total outcome variation (approximately 1.7% to 9.6%, depending on the amount of outcome error), outcome error biases the measured baseline mixed model estimate more heavily. In Figures 1A and 1B, when there is more "random" outcome variation across-subjects making the exposure less influential (contributes approximately 0.7% to 1.1% of the total variance of the baseline outcome), there is less bias caused by outcome error. With more "randomness" across-subjects we also see negative bias for small amounts of outcome error, followed by positive bias as outcome error increases. This shape may be explained by how the random intercept, a_i^*, is included twice in the measured baseline model (i.e. the a_i^* term and within the Y_{i0} term). Therefore, the fixed and random effects of the cross-sectional term are dependent leading to bias in $\hat{\beta}_1$ towards zero. However, if the outcome error is large enough the positive bias will overpower this other form of bias. This initial negative bias is not apparent when there is a small random effect because the randomness makes up a smaller portion of Y_{i0} and so the dependence between Y_{i0} and a_i^* is not as strong.

In the Supplement Sections A and C.9, we also show through DAG's and simulations that adjusting for the measured baseline may increase the bias caused by unadjusted confounders. Confounding is often a concern in observational studies, further showing how in an observational context, we want to avoid adjusting for the measured baseline.

The repeated scaled change model does not suffer the same bias as the measured baseline model. In certain settings such as studies with very long follow-up the repeated scaled change estimate may have comparable precision to the modeled baseline. However, in general the modeled baseline model will provide more precise estimates and so may be preferable to the repeated scaled change model. In our simulations, the modeled baseline and repeated scaled change exposure effect estimates were highly correlated. However, this may be in large part due to the simplistic model without many covariates. We would expect there to be less correlation between the two models' estimates when there are more variables included in the models. Even though in certain circumstances both models may be valid choices, it should be clear to the researcher that a single method should be chosen a priori to avoid random high bias from choosing the estimate that gives a more desired result [Fleming, 2010].

3.2 Exposure Effect Estimates in the Modeled Baseline Mixed Model

We have argued that the modeled baseline linear mixed model is usually most appropriate for an observational setting, and so we now focus on characteristics of the modeled baseline model and how the two exposure estimates it produces perform under different contexts.

3.2.1 Performance of the Modeled Baseline Exposure Effect Estimates in Different Contexts

In many ways, the exposure effect estimates from the modeled baseline mixed model are influenced by factors in a similar way as compared to estimates from a common regression model without repeated measures. However, it is important to understand the caveats of a model with two exposure estimates, and how these estimates are influenced differently by certain factors. The cross-sectional exposure effect, $\hat{\alpha}_1$, will rely

primarily on across-subject variation observed at baseline. On the other hand, the longitudinal effect, $\hat{\beta}_1$, will depend on both exposure variation at baseline and temporal exposure variation, in addition to general temporal information such as follow-up time. This is exemplified by how the characteristics of the exposure variation influence the precision of each exposure effect estimate (See Supplement Sections C.6, C.7 and C.8 for further details). In general when we use regression models to estimate an exposure effect, we will have more precision if the exposure variable takes on a wide range of values. In a longitudinal study context, we have different forms of exposure variation to consider and the cross-sectional and longitudinal exposure estimates will be influenced differently by different characteristics of the exposure variation. Across-subject variation in the subject average exposure levels (Supplement Section C.8) contributes to the variation in exposure at baseline (and any other visit time), but not to the temporal variation in the exposure. Therefore, this across-subject variation has a large influence on the precision of the cross-sectional exposure estimate (See Supplement Figure 24). The across-subject variation also has a fairly large influence on the longitudinal estimate. We see a different influence of the exposure variation caused by a temporal trend in the average exposure levels in a population (Supplement Section C.6). If in the population the average exposure levels are changing over time, this will add temporal exposure variation to each subject, but not add variation across subjects at baseline. This population-level trend therefore influences the precision of the longitudinal term and has almost no influence on the cross-sectional term (See Supplement Figure 20).

In the Supplement Section C, we consider how other characteristics of the population, variables of interest and study design affect each exposure estimate. We find that we must consider both cross-sectional and longitudinal confounding of our estimates, which may bias the cross-sectional and longitudinal exposure effect estimates respectively (See Supplement Figure 26). There is positive bias if the effect of the

confounder on the outcome is the same sign (i.e. positive or negative) as the correlation of the confounder and the exposure. The bias is negative if these quantities have opposite signs. In Figure 26 of the Supplement, we also see that confounding can bias one of the exposure estimates without biasing the other if the confounding variable only affects either the outcome baseline or outcome rate of change. In practice, we might expect that such a scenario is unusual, and so it is generally safest to assume a potential confounder affects both the baseline outcome and outcome rate of change and adjust for the covariate in both the cross-sectional and longitudinal components of the model. The effects of confounding on the precision of the estimates are small and unimportant compared to the bias produced.

3.2.2 Findings: Relationship of the Modeled Baseline Exposure Effect Estimates

In addition to understanding the differences of the two exposure estimates, we also consider their relationship with each other. If the baseline exposure levels are close in value to the follow-up levels (e.g. when exposure does not vary over time), we observe a positive correlation in the biases of the two exposure estimates across realizations of our simulations. This correlation may be induced by either the characteristics of the exposure variable or how the exposure variable is specified in the fitted model. For example, Figures 2A and 2B show the exposure effect estimates in settings with and without a population-level exposure decline, respectively. When there is no population level exposure trend (2B), there is less temporal exposure variation and the baseline and follow-up exposure values used in each component of the model are very similar. This leads to a stronger correlation of $\hat{\alpha}_1$ and $\hat{\beta}_1$. When an exposure trend is present (2A), the baseline and follow-up exposure measurements are farther apart on average, and the exposure effect estimates are less correlated. Note that in these simulations for Figure 2 the length of follow-up was increased from 5 to 9 years

so that the influence of the exposure trend would be more apparent.

In Figure 3, we show that how the researcher chooses to specify the exposure variable in the fitted model may also induce a correlation in the exposure estimates. In Figure 3A, the exposure variable used in the fitted model is the true exposure, which varies over time. In Figure 3B, the exposure used for all measurement times in the fitted model is each subject's baseline exposure level. Using the baseline levels may be done in practice either for simplicity or because the available exposure data is limited. When the exposure variable is misspecified as time-constant in this way, the baseline and follow-up exposure levels are forced to be identical, thus inducing a stronger correlation between $\hat{\alpha}_1$ and $\hat{\beta}_1$. In Figure 3 we also see how misspecification of a time-varying exposure as time-constant may result in a loss of precision and/or bias. Further exploration of exposure specification is given in the Supplement Section C.10.

3.2.3 Discussion: Differences and Interpretations of the Modeled Baseline Exposure Effect Estimates

We clarify the behavior and relationship of the two exposure effect estimates, $\hat{\alpha}_1$ and $\hat{\beta}_1$. Our primary interest is in the estimate of the exposure effect on outcome rate of change, $\hat{\beta}_1$. The second exposure effect estimate, $\hat{\alpha}_1$, has been listed as an added benefit of using the modeled baseline model [Gassett et al., 2015]. However, depending on the study context and population this may not be meaningful to the researcher, as we discuss below.

The cross-sectional exposure effect estimate $\hat{\alpha}_1$ may be interpreted as the estimated effect of exposure on baseline levels of the outcome. Unlike the longitudinal estimate $\hat{\beta}_1$, the cross-sectional effect is not a difference in the outcome change per unit of time, and may depend on a longer exposure time duration. In most observational contexts, is it expected that an exposure effect on baseline will be larger than

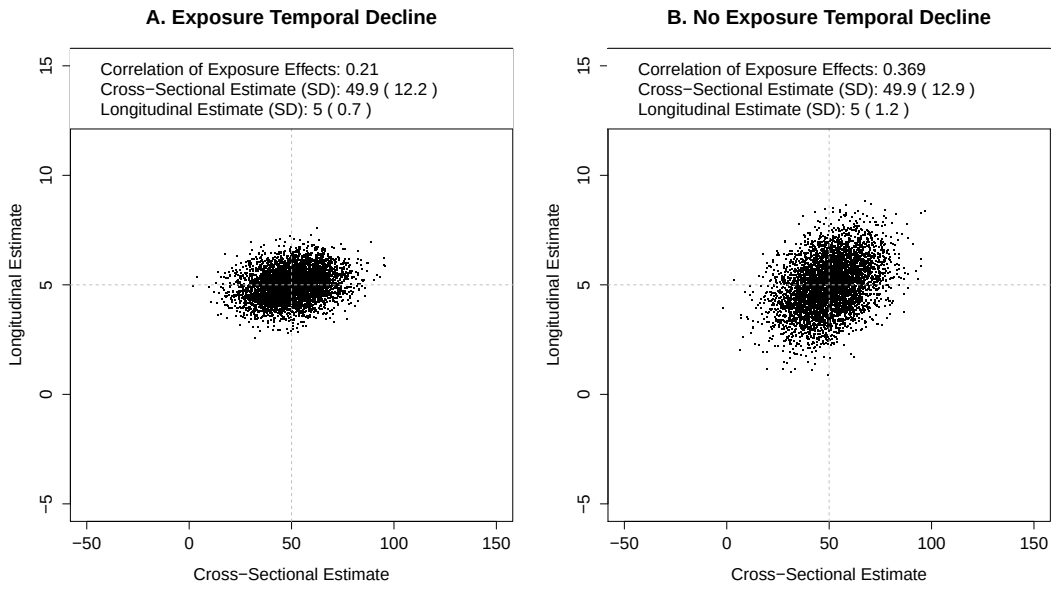

Figure 2: Cross-sectional and longitudinal exposure effect estimates from modeled baseline mixed models under populations with different exposure characteristics. The horizontal and vertical dashed lines show the true values of the exposure effects being estimated ($\alpha_1 = 50$ and $\beta_1 = 5$). **A** provides estimates when the average population exposure levels decline over time. **B** show the estimates when the average population exposure levels are constant across time. The averages and standard deviations of the cross-sectional and longitudinal exposure effect estimates are provided in each plot, along with the correlation of the two estimates

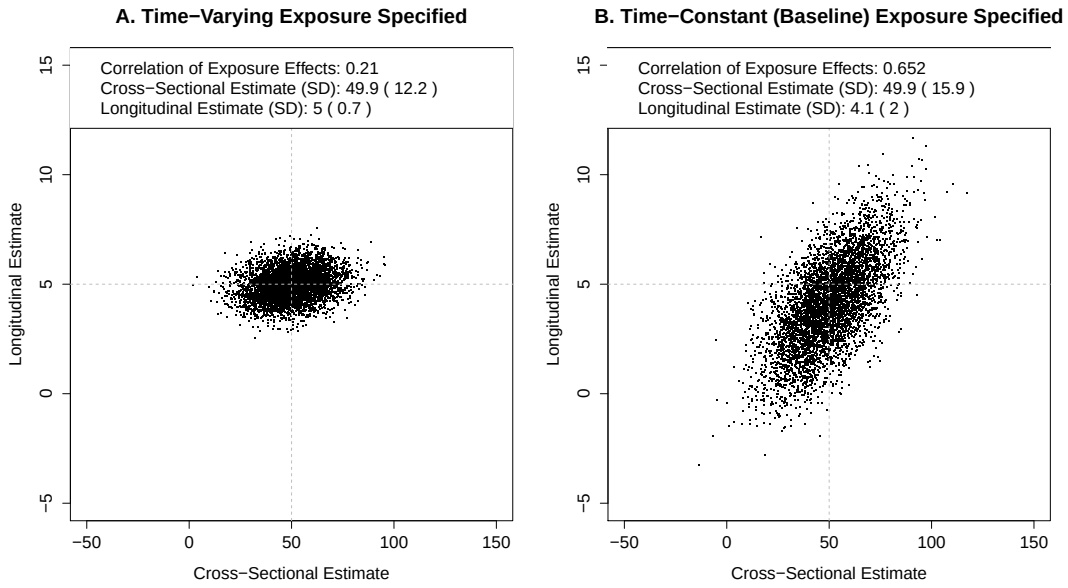

Figure 3: Cross-sectional and longitudinal exposure effect estimates from modeled baseline mixed models under fitted models which specify the exposure correctly as time-varying (**A**) and incorrectly as time-constant (baseline levels) (**B**). The horizontal and vertical dashed lines show the true values of the exposure effects being estimated ($\alpha_1 = 50$ and $\beta_1 = 5$). The averages and standard deviations of the cross-sectional and longitudinal exposure effect estimates are provided in each plot, along with the correlation of the two estimates

and in the same direction as the effect on rate of change, since the effect on baseline may be thought of as the cumulative effect of potentially many years during which the exposure affects the outcome progression. Furthermore, we generally expect for there to be more potential for bias due to confounding on the baseline exposure estimate, as many of the relevant factors related to the baseline outcome may have occurred long before the study and be poorly understood or accounted for. The types of variables that confound the cross-sectional and longitudinal estimates may be very different. The variation at baseline is important for cross-sectional confounders, whereas both the variation at baseline and temporal variation may be relevant for longitudinal confounders. Depending on the context we may believe that the exposure is less likely to be correlated with these types of time-varying covariates.

We have also shown that the exposure effects may be correlated under conditions where the baseline and follow-up exposure measurements are similar in value. In the Supplement Section C.11, we find that this correlation is partially dependent on if the time specified for the cross-sectional term is the baseline or a later visit. In practice, we would not parameterize a model with a cross-sectional term corresponding to a time other than baseline for several reasons. The estimate would be less interpretable, impractical if subjects have different follow-up timing, and potentially model outcome values from early measurements with future covariate data in the cross-sectional term. In our simulations, we get around these issues by giving all subjects the same follow-up timing and specifying the covariates as time-constant. We find that if the exposure is time-constant and we change the model parameterization so that the cross-sectional term corresponds to a different time than baseline, the cross-sectional exposure effect, $\hat{\alpha}_1$, will shift by the product of the longitudinal effect estimate and the time since baseline specified for the cross-sectional component. The longitudinal estimate, $\hat{\beta}_1$, remains constant. Even though this parameterization would never be used in practice, this finding still illustrates how the cross-sectional component is more susceptible to

being influenced by study design and analysis choices. In general we found that there are many situations in which the cross-sectional effect may be estimated poorly, but that the longitudinal estimate is usually not strongly affected by this. For these reasons provided in the last two paragraphs, we recommend focusing on the rate of change estimate of interest $\hat{\beta}_1$, and treating the cross-sectional effect, $\hat{\alpha}_1$, primarily as a tool for better estimating this longitudinal effect.

4 Analysis of Particulate Matter Effect on Coronary Artery Calcium

To demonstrate the use of the rate of change linear mixed models, we will consider the effect of ambient air pollution on progression of coronary artery calcium (CAC) in a subset of the MESA Air Study. We provide a summary of the study population, analysis methods and results of the analyses. Further details are presented in Supplement Section D, and R code is given in Supplement Section E.2.

4.1 Analysis Methods

4.1.1 Study Characteristics

We consider the subgroup of 984 MESA Air subjects from the Baltimore location with complete data for CAC, $PM_{2.5}$ exposure and the adjustment variables of interest. This is one of the six metropolitan areas considered in Kaufman et al. [2016], and the same group of subjects analyzed by Keller et al. [2017]. In this subgroup, subjects have an average of 2.2 visits (range of 0 to 4) and an average of 4.8 years of follow-up (range of 0 to 11.2). Our outcome of interest is CAC, a subclinical measure of cardiovascular disease. The exposure we focus on is fine particulate matter less than 2.5 μm in diameter ($PM_{2.5}$). Air pollutant concentrations were predicted at

each subject's residence using spatiotemporal models described in Keller et al. [2015]. A subject's long term exposure levels were calculated by the average of residence-specific two-week measurements [Kaufman et al., 2016]. Participant characteristics of the Baltimore location are provided in Table 9 of the Supplement Section D, and differ only slightly from that of the Table in Kaufman et al. [2016] due to small differences in selection criteria.

4.1.2 Description of Fitted Models

To estimate the exposure effect on rate of change, we fit a modeled baseline mixed model. Since our primary purpose is to understand model performance and we are analyzing a subset of data which was already investigated in Kaufman et al. [2016] and Keller et al. [2017], we fit a repeated scaled change mixed model as well. A measured baseline model of the form in Table 1 could not be fit because the model requires at least three measurements per subject on average in order to fit both a random intercept and slope. The measured baseline model runs into this fitting issue while the modeled baseline does not because the measured baseline model uses only the follow-up data and accounts for the baseline outcome through a covariate in the model. Because the modeled baseline model is fit with vectors of the baseline and follow-up observations, it is easier to fit both a random intercept and slope since there only needs to be on average two or more measures per subject. The repeated scaled change model is fit with just the follow-up data similarly to the measured baseline. This is because the outcome is the scaled change, which is undefined at baseline. However, the repeated scaled change model is only fit with one random effect, and so like the modeled baseline model only needs an average of more than two observations per subject in order to be fit.

4.1.3 Adjustment Covariates

In both the modeled baseline and repeated scaled change models we adjust for risk factors and potential confounders including age, sex, race/ethnicity, income, employment outside of the home, smoking status, second-hand smoke exposure, packyears, physical activity, adiposity measurements, total cholesterol, high density lipoprotein (HDL), triglycerides, statin use, neighborhood socio-economic status (SES) index and education. In the modeled baseline mixed model, we adjust for these risk factors in both the cross-sectional and longitudinal components as failure to adjust for potential confounders in both components may produce biased estimates if the variable in fact influences both the outcome baseline and rate of change (See Supplement Section C.9 for confounder simulations).

In the modeled baseline model we also adjust for the CAC scanner type as a time-varying covariate in the transient component of the model. We do not adjust for scanner type in the repeated scaled change model, which highlights one of the major drawbacks of the model. Scanner type is a variable that influences the outcome at the specific measurement time, and has the same effect regardless of how much time has elapsed since baseline. It is therefore illogical to model scanner type as having an effect on the scaled change. One alternative adjustment method would be to pre-adjust CAC by the scanner type prior to fitting the repeated scaled change model, but this adds new complexity and limitations. We discuss the importance of transient variables further in Section 4.3.

4.2 Analysis Results

Particulate matter exposure effect estimates from each model are provided in Table 3. With the modeled baseline model we estimate that for a $1\frac{\mu g}{m^3}$ higher $PM_{2.5}$ exposure, CAC will on average progress at a rate that is 4.0 Agatston units per year higher (95% CI: 2.1,6.0). With the repeated scaled change model the estimated ex-

	Estimate	SE	95% CI Lower	95% CI Upper
Rate of Change Effect *(Agaston units per year per 1 $\frac{\mu g}{m^3}$ difference in $PM_{2.5}$)*				
Modeled Baseline	4.00	1.00	2.10	6.00
Scaled Change (unadj.)	5.20	1.20	2.90	7.40
Scaled Change (pre-adj.)	5.20	1.20	2.90	7.50
Baseline Effect *(Agaston units per 1 $\frac{\mu g}{m^3}$ difference in $PM_{2.5}$)*				
Modeled Baseline	26.2	16.8	-6.8	59.2
Scaled Change (unadj.)	-	-	-	-
Scaled Change (pre-adj.)	-	-	-	-

Table 3: Estimates, standard error and 95% confidence intervals from the modeled baseline and repeated scaled change models for the effect of exposure to particulate matter ($PM_{2.5}$) on progression of coronary artery calcium (CAC). Exposure effects on CAC rate of change (i.e. $\hat{\beta}_1$ and $\hat{\beta}_1^*$) are provided above and the modeled baseline exposure effect on baseline (i.e. $\hat{\alpha}_1$) is provided below. An estimate is provided for both the repeated scaled change model unadjusted for scanner type, and the model that pre-adjusts the outcome by the modeled baseline scanner type effect estimate. The rate of change estimates provide the estimated effect of a $1\frac{\mu g}{m^3}$ higher $PM_{2.5}$ exposure on rate of change in CAC (Agatston units per year). The modeled baseline cross-sectional effect is the estimated effect of a $1\frac{\mu g}{m^3}$ higher $PM_{2.5}$ exposure on the baseline CAC (Agatston units).

posure effect on rate of change is 5.2 (95% CI: 2.9, 7.4). Although the estimate is slightly higher than the modeled baseline estimate, their confidence intervals overlap considerably. Consistent with our simulations, the repeated scaled change model has a larger standard error (1.2) than that of the modeled baseline (1.0). We also report the modeled baseline cross-sectional exposure effect estimate. For a $1\frac{\mu g}{m^3}$ higher $PM_{2.5}$ exposure, we estimate that the baseline CAC will on average be 26.2 Agatston units higher (95% CI: -6.8, 59.2).

4.3 Discussion of Analysis

In our analyses of the Baltimore subset of the MESA Air study, we produce estimates of the effect of particulate matter ($PM_{2.5}$) exposure on rate of CAC progression. In Section 4.2, we report the effect of a $1\frac{\mu g}{m^3}$ higher $PM_{2.5}$ exposure on CAC progression, since those are the units which the simulations are roughly based around. To be

consistent with the previous analyses done by Kaufman et al. [2016] and Keller et al. [2017], we could also choose to report the estimates of a $5\frac{\mu g}{m^3}$ difference. With the modeled baseline model we have an estimate of 20.2 Agatston units per year for a $5\frac{\mu g}{m^3}$ higher $PM_{2.5}$ exposure (95% CI: 10.6, 29.8). This is consistent with the analyses reported in Keller et al. [2017], which also focused on the Baltimore subset. The estimate is higher than the estimate by Kaufman et al. [2016], which was based on the entire MESA Air study. Including the data from all 5 metropolitan areas and adjusting for the location factor, CAC progression was estimated to be on average 4.1 Agatston units per year higher with a $5\frac{\mu g}{m^3}$ higher $PM_{2.5}$ exposure (95% CI: 1.4, 6.8) by Kaufman et al. [2016]. This difference between the Baltimore-specific and whole-study estimates may be due to several reasons, such as differences in the cohorts or composition of the air pollution across different locations [Keller et al., 2017].

We report the estimated effect of $PM_{2.5}$ exposure on the baseline CAC values, which is roughly seven times as large as the estimated effect on rate of change. This estimate is reasonable from a scientific perspective. The cross-sectional estimate is in the same direction as the longitudinal estimate (i.e. exposure to higher levels of particulate matter tend to have higher baseline CAC and increased progression of CAC). The estimated effect on baseline is also much larger than the effect on rate of change, although there is much more uncertainty in the estimate. In fact, the 95% confidence interval is consistent with no effect. This difference in scale between the cross-sectional and longitudinal effects makes sense as the effect on baseline is a cumulative effect that likely depends on a longer history of exposure.

Although the cross-sectional exposure effect estimate is interpretable, we still recommend that it is not emphasized. A cross-sectional estimate is not even reported in either Kaufman et al. [2016] or Keller et al. [2017] since the longitudinal effect was the primary interest. We report the cross-sectional estimate since it is relevant to our investigation of model performance. We also provide estimates from a repeated

scaled change model for this reason. If the application were the primary focus of the analysis, we would want to choose one method a priori to avoid random high bias [Fleming, 2010].

Consistent with our simulations, the estimate from the repeated scaled change model is less precise than the modeled baseline. A major limitation of the repeated scaled change model that is highlighted is the inability to easily adjust for transient variables such as CAC scanner type. We could pre-adjust the outcome for scanner type prior to fitting a repeated scaled change model, but this would require an ad hoc approach such as fitting the modeled baseline model first in order to estimate the effect of scanner type. If we fit this repeated scaled change model pre-adjusted by the modeled baseline scanner type effect estimate, we get a similar estimate as the unadjusted model. However, in other situations where a transient variable is very influential, failure to adjust for it may result in estimator bias in the exposure effect. We did not address transient variable confounding in the confounder simulations in the Supplement Section C.9. However, it is reasonable for a transient variable to be associated with exposure and affect the outcome, which produce a potential for confounding. For example, in the Baltimore subset of the MESA Air study, one scanner type was used for early CAC measurements, and a second type was used for later visits (See Table 8 in the Supplement Section D). The exposure levels decline on average during the study, and so this study design feature of replacing technology halfway through induces an association between exposure and scanner type. A study design such as this with very little temporal overlap of technology may make it especially difficult to properly correct for a transient variable and distinguish actual changes in the outcome from differences in technology used to measure the outcome. Outcomes which are dependent on changing technology are a generic problem for longitudinal studies. These problems may be especially problematic for long-term studies where technology much be replaced, or in situations where only certain subpopulations may

receive advances in technology and this subpopulation happens to be systematically different from the population as a whole.

5 Discussion

Linear mixed models with a modeled baseline are a useful method of analysis in many longitudinal study contexts. There are several already established advantages such as the ability to handle data that are missing at random and gain information from unbalanced data (even in subjects with a single measurement time) [Fitzmaurice et al., 2012]. We have shown that the model avoids bias due to adjusting for the measured baseline, and is more precise than regression on the scaled change. We also show how the two exposure estimates (i.e. cross-sectional and longitudinal) are influenced differently by different features of the scientific context and study design and that the estimates may be correlated under certain conditions. Our findings along with what is already known about the model suggests that the cross-sectional effect be treated primarily as a tool to help estimate the longitudinal effect of interest. In an application to the MESA Air study, we demonstrate several other advantages of the modeled baseline mixed model when data is not as well behaved as the simulations. Most notably, we highlight the ability to easily adjust for transient variables and fit a model with few repeated measures per subject.

Consistent with the literature on regression of change models [Lepage et al., 2015] [Van Breukelen, 2006] [Yanez et al., 1998], we demonstrated in a more general context with multiple follow-up measurements that controlling for a measured baseline outcome will cause bias in a rate of change estimate when the outcome is measured with error and the exposure is correlated with the baseline outcome. It is common in observational studies for the exposure to effect the baseline outcome, and so it is important to understand that inference in an observational setting may be fundamen-

tally different than in a randomized one. In the Supplement Section A, we show with DAG's that adjusting for the measured baseline may even increase the bias caused by unadjusted confounders. This claim is supported by simulations done in the Supplement Section C.9. Confounding is often a concern in observational studies, which further supports the use of a model that avoids adjusting for the measured baseline. In our discussion of the MESA Air analyses in Section 4.3, we also highlight how a measured baseline model requires a higher number of measures per subject than is necessary to fit a modeled baseline model with a similar random effects structure. Other measured baseline model forms could be used, but in all cases the estimates may suffer from bias from controlling for the measured baseline.

In our simulations, we also show that a repeated scaled change model will generally provide less precise estimates than the modeled baseline model. In our analysis of the MESA Air study (Sections 4.1.2 and 4.3), we highlight how the model cannot adequately adjust for transient variables that are relevant to the specific observation time and not the rate of change. In our simulations the repeated scaled change model did not produce biased estimates, and was strongly correlated with the modeled baseline estimates. However, this correlation may not be generalizable to non-simulation settings or where there are more complex outcome models.

In addition to demonstrating the advantages of the modeled baseline mixed model, we gain insight from our simulations on how to properly use the model and interpret the two exposure estimates (i.e. the cross-sectional/baseline effect, $\hat{\alpha}_1$, and the longitudinal/rate of change effect, $\hat{\beta}_1$). We recommend the cross-sectional effect be treated as a tool for estimating the longitudinal effect of interest for several reasons. The cross-sectional effect is dependent on a potentially long-term history of exposure prior to the study period, making it more susceptible to bias from confounding. The cross-sectional effect and correlation between the two exposure effect estimates is also partially driven by the baseline time specified for the cross-sectional effect (See Sup-

plement Section C.11); this model specification does not affect the longitudinal effect estimate.

There are many considerations to using a modeled baseline mixed model that we have not considered in depth in this paper. In the Supplement Section C, we show how many design features and characteristics of the exposure influence the exposure effect estimates. In The Supplement Sections C.1 and C.3 we show that the length of follow-up and sample size greatly influence the precision of the rate of change exposure effect estimates. The frequency of follow-up was less influential on the rate of change exposure effect estimates (Supplement Section C.4), although in a more complex time-varying exposure framework the frequency may be more important. These results may be used to help inform design of an effective longitudinal study, although other factors must be considered. For example, longer studies allow for more precise inference on an exposure effect on rate of change, but there may be added complexity such as advancements in technology or practices which can potentially bias estimates. This, of course, is in addition to considerations such as financial constraints.

In Section 3.2.2, we briefly consider specification of the exposure variable and the consequences of using a time-constant exposure when the true exposure is time-varying. We explore the topic of exposure specification further in the Supplement Section C.10. We find that incorrect specification of long-term vs. short term exposure, or a cumulative vs. moving average effect may lead to large amounts of estimator bias. In practice, the scientific context should drive many of the exposure specification decisions, such as the relevant time scale in which exposure matters and whether the exposure has a cumulative effect. However, understanding the statistical properties of these specifications may also inform decisions as certain methods may be more robust to misspecification. Exposure specification is a topic that deserves further research.

Another way in which our treatment of exposure is simplistic is the assumption of a fixed exposure effect for all subjects. This is assumed in both the fitted models and the data generating models used in the simulations. In most situations, it is expected that there may be subject-to-subject differences in the effect of the exposure. One way to relax our assumptions could be to include a random effect on the exposure (i.e. random effect on E_{i0} in the cross-sectional component and on E_{iv} in the longitudinal component). There are many other aspects of the exposure specification we could consider. For example, in equation 1.1 we specify the interaction $E_{iv} * t$ but have no main effect for E_{iv}. A main effect for E_{iv} could be considered in the transient component of the model. Exploring these model specifications is a logical next step for future work.

Exposure error is another issue relevant to many environmental epidemiological study contexts. Often exposures are not directly measured at subject-specific locations and must be predicted. These predicted values are then plugged into health models as a covariate. For example, in the MESA Air Study, air pollution exposure was measured at locations misaligned with subject residences, and spatiotemporal models were used to predict the subject-specific exposure levels [Keller et al., 2015]. When predicted exposures are plugged into a health model, there may be differences in the true and predicted exposure values. Generally these errors are thought to be a combination of classical-like and Berkson-like forms of measurement error, and the presence of exposure error in a model can lead to attenuated exposure effect estimates [Szpiro and Paciorek, 2013][Szpiro et al., 2011a][Szpiro et al., 2011b]. Methods may be applied in attempt to correct for these types of measurement error [Szpiro and Paciorek, 2013].

One of the main features of linear mixed models which we did not investigate is the assumption of a linear exposure effect on the outcome and outcome rate of change. In many contexts, it is not believable that the exposure and other predictors have

a linear relationship with the outcome, and alternative approaches may be preferred such as adding higher order terms or more complex models [Bryan and Heagerty, 2014]. For example, linear splines could be used to allow the exposure effect to be stronger or weaker for different exposure levels. However, it is worth noting that like most models, the modeled baseline mixed model is not meant to perfectly reflect the underlying scientific mechanisms of the outcome. Rather, the model can be thought of as a useful way of estimating the impact of exposure contrasts and understanding the general trend of an exposure effect.

The modeled baseline linear mixed model is a method that is appropriate in many longitudinal study contexts where the goal is estimating an exposure effect on rate of change. Understanding how the context of the study influences estimates and further development of methods to address these issues is important for making proper inference in longitudinal analyses.

Bibliography

M Bryan and P.J. Heagerty. Direct regression models for longitudinal rates of change. *Statistics in medicine*, 33(12):2115–2136, 2014.

M.J. Budoff, R Young, V.A. Lopez, R.A. Kronmal, K Nasir, R.S. Blumenthal, R.C. Detrano, D.E. Bild, A.D. Guerci, K Liu, et al. Progression of coronary calcium and incident coronary heart disease events: Mesa (multi-ethnic study of atherosclerosis). *Journal of the American College of Cardiology*, 61(12):1231–1239, 2013.

E.A. Colosimo, M.A. Fausto, M.A. Freitas, and J.A. Pinto. Practical modeling strategies for unbalanced longitudinal data analysis. *Journal of Applied Statistics*, 39(9): 2005–2013, 2012.

G.M. Fitzmaurice, N.M. Laird, and J.H. Ware. *Applied longitudinal analysis*, volume 998. John Wiley & Sons, 2012.

T.R. Fleming. Clinical trials: discerning hype from substance. *Annals of internal medicine*, 153(6):400–406, 2010.

A.J. Gassett, L Sheppard, R.L. McClelland, C Olives, R Kronmal, M.J. Blaha, M Budoff, and J.D. Kaufman. Risk factors for long-term coronary artery calcium progression in the multi-ethnic study of atherosclerosis. *Journal of the American Heart Association*, 4(8):e001726, 2015.

J.D. Kaufman, S.D. Adar, R.G. Barr, M Budoff, G.L. Burke, C.L. Curl, M.L. Daviglus, A.V.D. Roux, A.J. Gassett, D.R. Jacobs, et al. Association between air pollution and coronary artery calcification within six metropolitan areas in the usa (the multi-ethnic study of atherosclerosis and air pollution): a longitudinal cohort study. *The Lancet*, 2016.

J.P. Keller, C Olives, K Sun-Young, L Sheppard, P.D. Sampson, A.A. Szpiro, A.P. Oron, J Lindström, S Vedal, and J.D. Kaufman. A unified spatiotemporal modeling approach for predicting concentrations of multiple air pollutants in the multi-ethnic study of atherosclerosis and air pollution. *Environmental Health Perspectives*, 123 (4):301–309, 2015.

J.P. Keller, E Austin, T Larson, L Sheppard, A Szpiro, S Vedal, and J.D. Kaufman. Multi-pollutant clustering of traffic-related air pollution and the progression of coronary artery calcification: the multi-ethnic study of atherosclerosis and air pollution (in progress). 2017.

B Lepage, S Lamy, D Dedieu, N Savy, and T Lang. Estimating the causal effect of an exposure on change from baseline using directed acyclic graphs and path analysis. *Epidemiology*, 26(1):122–129, 2015.

K Liang and S.L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.

J.W. McEvoy, M.J. Blaha, A.P. DeFilippis, M.J. Budoff, K Nasir, R.S. Blumenthal, and S.R. Jones. Coronary artery calcium progression: an important clinical measurement?: A review of published reports. *Journal of the American College of Cardiology*, 56(20):1613–1622, 2010.

A.A. Szpiro and C.J. Paciorek. Measurement error in two-stage analyses, with application to air pollution epidemiology. *Environmetrics*, 24(8):501–517, 2013.

A.A. Szpiro, C.J. Paciorek, and L Sheppard. Does more accurate exposure prediction necessarily improve health effect estimates? *Epidemiology (Cambridge, Mass.)*, 22 (5):680, 2011a.

A.A. Szpiro, L Sheppard, and T Lumley. Efficient measurement error correction with spatially misaligned data. *Biostatistics*, 12(4):610–623, 2011b.

Gerard JP Van Breukelen. Ancova versus change from baseline had more power in randomized studies and more bias in nonrandomized studies. *Journal of clinical epidemiology*, 59(9):920–925, 2006.

N.D. Yanez, R.A. Kronmal, and L.R. Shemanski. The effects of measurement error in response variables and tests of association of explanatory variables in change models. *Statistics in medicine*, 17(22):2597–2606, 1998.

Supplement

In the Supplement, we provide further information not covered in the main paper. We give an explanation of bias from controlling for the measured baseline (Section A),

details on the simulation framework (Section B), additional simulation study results (Section C), MESA Air analysis details and descriptive tables (Section D), and R code for the simulations and MESA Air analysis (Section E). A summary of the population and study design features investigated in the main paper and Supplement are provided in Table 5 of Section C.

A Bias from Controlling for Baseline

When estimating an outcome change from baseline or rate of change, how to control for the baseline is regularly debated. Whether to adjust for a measured baseline outcome in a model depends on the study design, whether confounding is thought be be properly adjusted for and regression to the mean [Lepage et al., 2015].

Change models that control for the measured baseline have been shown to have more power in a randomized setting. However, in a non-randomized setting where the exposure is associated with the baseline outcome, a model adjusted for the measured baseline can produce biased estimates of the exposure effect in the presence of outcome measurement error [Van Breukelen, 2006]. This bias can be present even if the exposure is measured without error [Yanez et al., 1998]. In our simulations we expand this result into a more general setting with multiple follow-up measurements and estimating rate of change.

Bias from controlling for the measured baseline was explained with directed acyclic graphs (DAG's) by Lepage et al. [2015]. In this section we summarize much of their findings as it applies to our observational study setting, which is a combination of their Figures 2 and 4. We attempt to stay fairly consistent with their format and notation in order to draw a connection between their research and the simulations in our paper.

In Figure 4, we present a DAG similar to those in Lepage et al. [2015] looking at

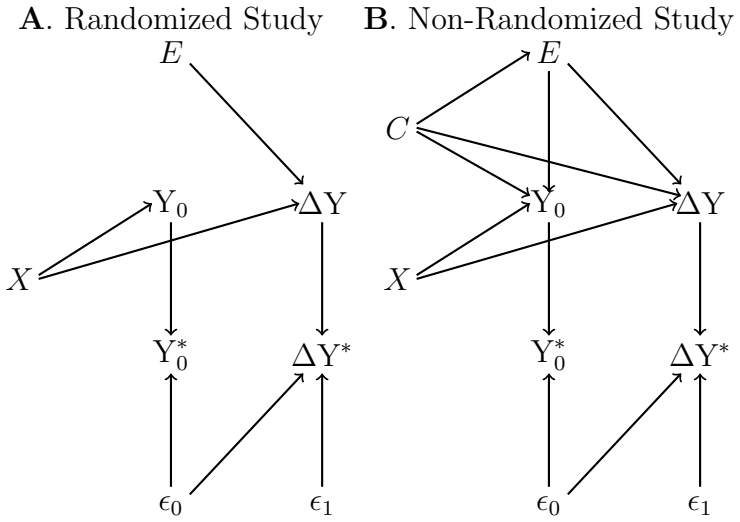

Figure 4: Directed Acyclic Graphs (DAG's) of a simplified longitudinal study setting with one follow-up time, similar to those in Lepage et al. [2015]. **A** gives a randomized trial where exposure has no affect on the outcome baseline levels. **B** gives a non-randomized trial with confounding and the exposure affecting the outcome baseline levels. In the DAG's, E= exposure, X = other predictors independent of the exposure, Y_0 = true baseline outcome, Y_0^* = observed baseline outcome, ϵ_0 =baseline measurement error, ϵ_1 = follow-up measurement error, $\Delta Y = Y_1 - Y_0$=true outcome change from baseline, $\Delta Y^* = Y_1^* - Y_0^*$=observed outcome change from baseline, and C= confounder.

the simplified setting of two measurements (i.e. baseline and follow-up) and a single exposure that does not change over time. We are interested in the exposure effect on change in outcome ($E \rightarrow \Delta Y$). Because we do not observe the true outcome change, ΔY, we can only make inference on the exposure effect on the observed outcome change, ΔY^*. However, using path analysis, the coefficient estimate for ΔY is equivalent to that of ΔY^* [Lepage et al., 2015]. We therefore are interested in whether any backdoor paths exist between E and (ΔY^*), which will create bias in the estimate of the exposure effect on outcome change.

In a randomized study (Figure 4A), the exposure has no effect on the baseline outcome and a change model will produce an unbiased estimate regardless of if the model is adjusted or unadjusted for the measured outcome baseline. However, in a non-randomized study (Figure 4B) where the exposure affects the baseline outcome, con-

trolling for the measured baseline of the outcome can induce an association between the baseline measurement error (ϵ_0) and the true outcome at baseline, (Y_0) [Lepage et al., 2015]. This association creates an unblocked backdoor path between the exposure, E, and the measured change in outcome, ΔY^* (i.e. $E \rightarrow Y_0 - - - \epsilon_0 \rightarrow \Delta Y^*$).

As we can see, the addition of a confounder (C) adds an unblocked backdoor path $E \leftarrow C \rightarrow \Delta Y \rightarrow \Delta Y^*$ regardless of whether the unadjusted or adjusted model is used. However, if we adjust for the measured baseline (Y_0^*) when there is confounding, we have the additional backdoor path $E \leftarrow C \rightarrow Y_0 - - - \epsilon_1 \rightarrow \Delta Y^*$ which will add bias as well. This shows that not only will adjusting for the measured baseline produce bias if there is measurement error, but if confounding is present, adjusting for the measured baseline may increase the bias due to confounding. However, it is also possible that these forms of bias may cancel each other out to a degree if the biases are in different directions.

From DAG's we can conclude that in non-randomized settings where the exposure affects the baseline outcome, a change model should not adjust for the measured baseline and should attempt to adjust for potential confounders [Lepage et al., 2015]. This result can be generalized to mixed models in the more common environmental epidemiological study context where there are more than one follow-up time per subject. We therefore prefer a mixed model controlling for modeled baseline rather than controlling for the measured baseline.

B Simulation Framework

In this section, we describe the outcome model used in simulations and its parameterization. We assume the modeled baseline model is the true model. We also provide details on the exposure and confounder frameworks, fitted models and parameter choices, as well as discuss some of the limitations of the simulations. In the Supplement Section E.1 we provide simulation R code. In Table 5 of Supplement Section C, we provide a guide for which population and study design features we explore, and the relevant sections in the main paper and Supplement.

B.1 Outcome Model

In our simulations, we consider a simplified model for our outcome with a single confounder and without transient effects,

$$Y_{iv} = (\alpha_0 + \alpha_1 E_{i0} + \alpha_2 X_{i0} + a_i) + (\beta_0 + \beta_1 E_{iv} + \beta_2 X_{iv} + b_i) t_{iv} + \epsilon_{iv}$$

Here, Y_{iv} is the outcome for subject i at follow-up visit v. The α's and β's are fixed effects, $\begin{pmatrix} a_i \\ b_i \end{pmatrix} \sim \text{MVN}(0, \Sigma_{rand})$ are random effects and $\epsilon_{iv} \sim N(0, \sigma^2_\epsilon)$ is an error term. For simplicity, the exposure, E_{it} and time since baseline, t_{iv}, are the only variables included in the model in most simulations. The covariate X_{iv} is considered in confounding simulations in Section C.9 but otherwise is ignored by setting its effects to be zero. The three fitted models used in the simulations are described in Section B.4. The measured outcome used in the fitted model has classical measurement error: $Y^m_{iv} = Y_{iv} + e_m$, where $e_m \sim N(0, \sigma^2_{e_m})$. However, this is independent of and indistinguishable from the outcome residual error, ϵ_{iv}.

B.2 Exposure Framework

The exposure variable is modeled by three separate components,

$$E = E_{trend} + E_{subject} + E_{time}.$$

The term E_{trend} is an underlying temporal trend in the exposure that is the same for the entire population. We set the trend as a linear function of time, and in most simulations have a linear decline in exposure for the entire population. The term $E_{subject} \sim N(0, \sigma^2_{E_{subject}})$ gives the deviation of the subject-specific average exposure levels from the population average. This term is time-constant and varies across subjects. The term $E_{time} \sim N(0, \sigma^2_{E_{time}})$ is subject-specific random temporal variation in the exposure. This contributes to the across-subject variation at any given measurement time, as well as to the within-subject variation of an individual. In the case of a constant population average exposure (i.e. $E_{trend} = 0$), then $\sigma^2_{E_{time}}$ is the within-subject variance of the exposure. This random temporal variation is different for each subject, but the average temporal decline (E_{trend}) is the same for all subjects.

Figure 5 provides an example of how these components of the exposure make up the exposure levels of 4 subjects. A plot of 150 of the simulated subjects' exposures compared to a plot of 150 subjects from the Baltimore cohort of the MESA Air study are provided in Figure 6. Note that the entire simulated population is followed for exactly 5 years. The subjects in MESA Air were on average followed for 4.8 years, but many were followed for more or less than this length of time (maximum of 11.2 years, IQR of 1.4 to 9.3 years).

In practice, an exposure may have more complex patterns of variation not captured by our framework, such as an autocorrelation structure in the within-subject temporal variation, and complex relationships with other relevant covariates in the model. With our simulation framework we are still able to learn how mixed model estimates of the

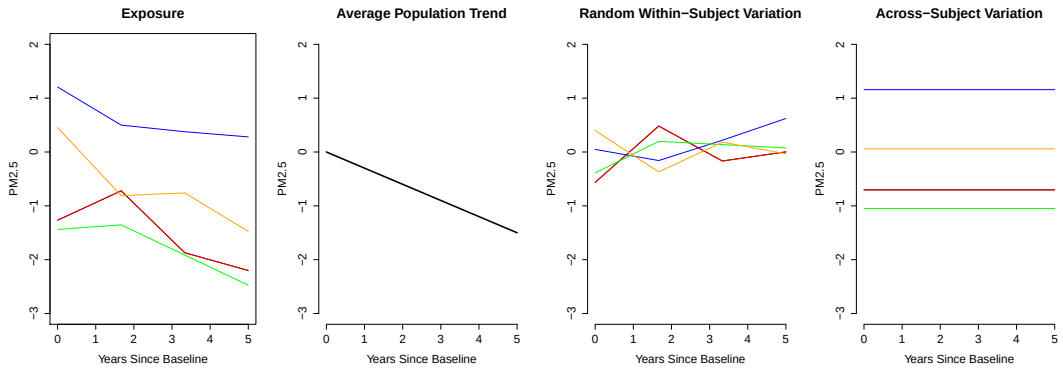

Figure 5: Example of simulated time-varying exposure for four individuals, broken up into its components. The leftmost graph shows the exposure of four different subjects over five years of follow-up. The three graphs on the right show the components that make up this exposure. From left to right: (1) the average exposure decline in the population (same for each subject), (2) random temporal exposure variation for each subject, and (3) the subject-specific average exposure levels (time-constant).

exposure effect parameters of interest are influenced by characteristics of the exposure, study design and population.

In Section C.10, we consider more complex exposure models. We consider how different exposure trends and time-scales of the true and fitted exposure effect (e.g. short-term, long-term and cumulative effects) influence the modeled baseline rate of change exposure effect estimates.

B.3 Study Design and Parameter Values

Each study design feature and parameter value has a "default" value used in most simulations. However, many of these features were investigated in simulations studies and had their values varied. Table 5 in Supplement Section C provides a summary of these default values, the range of values considered, and the relevant section in which each feature was considered.

For most simulations a sample size of 1000 was simulated, each with 4 measurement times equally spaced over 5 years. Parameter values in the model were chosen so that the outcome roughly represents coronary artery calcium (CAC). Analysis of

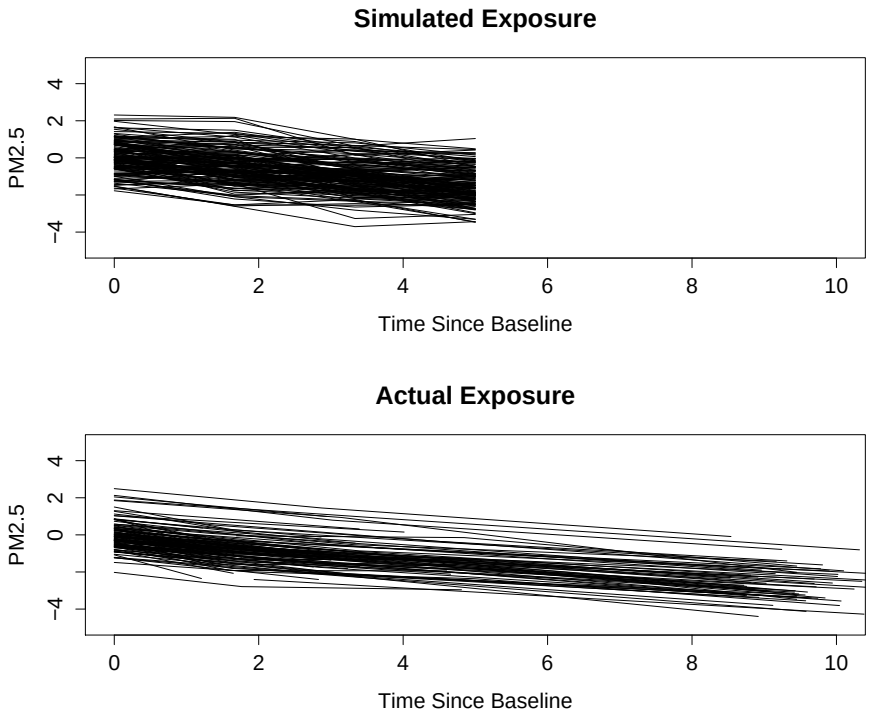

Figure 6: Simulated and actual (mean-centered) fine particulate matter ($PM_{2.5}$) for samples of 150 subjects from the simulated population and actual Baltimore cohort from the MESA Air study. Note that the simulated exposure is only "measured" for 5 years. The actual data has an average follow-up of 4.8 years, but there are some subjects followed for up to around 10 years.

the MESA Air Study was used to determine these values, and then many of these values were varied in different simulations. The intercept was chosen to be $\alpha_0 = 190$, which represents the average CAC at baseline for subjects with exposure levels equal to the average of the population. The rate of change parameter, $\beta_0 = 40$ represents the average change in CAC per year for subjects with the average exposure levels.

The exposure roughly represents $PM_{2.5}$ exposure. The population average decline (E_{trend}) was set to -0.3 $\frac{\mu g}{m^3}$ per year, consistent with the average particulate matter decline in the MESA Air Study. The across-subject variation in average exposure levels was set to $\sigma^2_{E_{subject}} = 0.7$, which is based on the Baltimore sample variance of the subject-average exposures after the population average trend was subtracted. The random within-subject variation was set to $\sigma^2_{E_{time}} = 0.15$, which was based on the sample variance of the exposure once the population average trend and subject-specific average exposure levels were subtracted. The exposure effect on baseline was chosen to be $\alpha_1 = 50$. The exposure effect on rate of change was set to $\beta_1 = 5$, which is the target of our estimates. These values are close to the exposure effect estimates from analyses in the Baltimore dataset. The value of $\beta_1 = 5$ is consistent with the estimated 5.2 Agatston unit change per year for a 1 $\mu g/m^3$ effect of $PM_{2.5}$ reported in Keller et al. [2017], which considered the Baltimore dataset. This is larger than the entire MESA study effect reported in Kaufman et al. [2016], which is 0.82 Agatston unit change per year for a 1 $\mu g/m^3$ effect of $PM_{2.5}$.

The outcome residual variance is set to $\sigma^2_\epsilon = 9,000$, and the random effects variance covariance matrix is $\Sigma_{rand} = \begin{pmatrix} 200,000 & 20,000 \\ 20,000 & 3,000 \end{pmatrix}$. In most simulations the outcome measurement error has a variance of $\sigma^2_{e_m} = 11,000$, which makes up approximately 5% of the total outcome variance at baseline.

Model
Modeled Baseline Mixed Model $\widehat{\mathrm{E}[Y_{iv}^m]} = (\hat{\alpha}_0 + \hat{\alpha}_1 E_{i0} + \hat{\alpha}_2 X_{i0} + a_i) + (\hat{\beta}_0 + \hat{\beta}_1 E_{iv} + \hat{\beta}_2 X_{iv} + b_i)t$
Measured Baseline Mixed Model $\widehat{\mathrm{E}[Y_{iv}^m]} = (\hat{\alpha}_0^* + \hat{\alpha}_1^* Y_{i0}^m + a_i) + (\hat{\beta}_0^* + \hat{\beta}_1^* E_{iv} + \hat{\beta}_2^* X_{iv} + b_i)t$
Repeated Scaled Change Mixed Model $\widehat{\mathrm{E}[\frac{Y_{iv}^m - Y_{i0}^m}{t}]} = (\hat{\beta}_0^{**} + \hat{\beta}_1^{**} E_{iv} + \hat{\beta}_2^{**} X_{iv} + b_i)$

Table 4: Fitted models used in the simulations studies. Description of the notation is provided in the Supplement Section B.1.

B.4 Fitted Models

We fit the three mixed models described in the main paper. The forms of the models used in the simulations are given in Table 4. We compare the exposure effects on rate of change (i.e. $\hat{\beta}_1$, $\hat{\beta}_1^*$ and $\hat{\beta}_1^{**}$), to the true value of $\beta_1 = 5$. The cross-sectional estimate from the modeled baseline model, $\hat{\alpha}_1$, is estimating the true value of $\alpha_1 = 50$.

B.5 Model Assumptions and Limitations

Several assumptions were made in the simulations primarily for the purpose of simplicity. These assumptions lead to several limitations in the simulation results.

As mentioned above, the outcome model is very simplistic. It follows a modeled baseline linear mixed model framework and only includes the exposure and time as covariates, as well as one additional covariate in some of the simulations. All of the variables are simulated, rather than using actual data, which ignores potential characteristics of the covariates such as outliers and skewness. The outcome also avoids skewness, even though real measures of CAC are positively skewed. The purpose of the model framework is to understand how the model components influence estimator precision and accuracy, and is not intended to be a completely accurate representation of the true mechanisms influencing CAC.

In this framework the exposure has a fairly small contribution to the total outcome

variation. In most simulations the exposure is responsible for approximately 1% of the outcome variance at baseline. In some of the simulations the exposure has a larger influence, but in general these simulations are representative of a context where there are many other factors influencing the outcome variable. In a scenario where the exposure is responsible for a much larger amount of the outcome variation (e.g. if there are small random effects in Section C.5), we might expect several factors to have more drastic influences on the exposure effect estimates.

Several assumptions were also made regarding the study design. All subjects are followed for the same length of time and have the same number of follow-up visits. The visits are equally spaced throughout the follow-up time and no data are missing. It is not investigated in this paper, but would be of interest to consider how variation in number of follow-up visits and timing of follow-up visits would effect estimation. This would especially be of interest if the follow-up timing is related to the outcome, as in practice how long a subject is followed may be associated with their outcome status.

C Additional Simulation Results

In this section, we provide results and discussion of how design, population and exposure characteristics influence the exposure effect estimates. Table 5 provides a summary of which factors are varied and the sections in which they are considered. We compare the modeled baseline, measured baseline and repeated scaled change estimates of the exposure effect on rate of change. We also consider the differences of the cross-sectional and longitudinal exposure effect estimates from the modeled baseline model. Note that since the modeled baseline is the correctly specified model, the exposure effect estimates are in general unbiased, and so we only present the bias of the cross-sectional and longitudinal exposure effect estimates from the modeled baseline model when we consider bias due to unadjusted confounding in Section C.9.

In Section C.10 we consider a more in-depth exploration of exposure specification, and in Section C.11 we consider how the parameterization of the cross-sectional effect influences the two modeled baseline exposure effects.

Parameter	"Default" Value	Range of Values & Section Where Varied		Notes
Study Design Characteristics				
Cross-sectional exposure effect (α_1)	50	0 (rand.), 50 (non-rand.)	3.1	
Length of Follow-up	5	3-18	C.1	
Sample Size	1000	500-6,000	C.3	
Frequency of Visits (total # of visits)	4	4-12	C.4	Length of follow-up held constant
Outcome Characteristics				
Outcome Residual Variance (σ^2_ϵ)	11,000	1-100,000	3.1 and C.2	
Random Intercept Variance (σ^2_a)	200,000	10,000-400,000	C.5	Correlation of a_i,b_i held constant
Random Slope Variance (σ^2_b)	3,000	1,000-6,000	C.5	Correlation of a_i,b_i held constant
Exposure Characteristics				
Linear Trend Slope	-0.3	(-0.5)-(0.5)	C.6	Also related to Section C.10
Random Temporal Variance ($\sigma^2_{E_{time}}$)	0.15	0-3	C.7	
Across-Subject Variance ($\sigma^2_{E_{subject}}$)	0.7	0-3	C.8	
Other Simulations				
Unadjusted Confounding	(None)	-	C.9	Effect sizes and correlation varied
Exposure Specification	See Section B.2	-	C.10	Also Figures 2, 3 in Section 3.2.2
Cross-Sectional Parameterization	(Baseline)	-	C.11	

Table 5: Summary of simulations in the main paper and Supplement. "Default" values used in most simulations are provided, as well as the range of values considered in specific simulation studies considering the factor. The sections in which the specific simulations discussed are provided as well, along with additional notes.

C.1 Length of Follow-up

We vary the number of years each patient is followed for. Although the length of follow-up changes, in each simulation all subjects have 4 visits equally spaced over the study time. In all models we find that the longer the follow-up, the more precise the rate of change estimates are. Intuitively, this makes sense because having a longer time to observe change will allow one to better estimate the rate of change and how an exposure affects this rate.

In Figures 7 and 8 we see the estimated exposure effects (estimating a true effect of $\beta_1 = 5$) and standard errors, respectively. The modeled baseline and repeated scaled change models are unbiased across different lengths of follow-up. The measured baseline estimate is biased when there is a short follow-up. This bias may be explained by how with shorter follow-up, there is less temporal variation in the exposure and outcome, and so the outcome measurement error is more influential. Outcome measurement error contributes to the bias of the measured baseline estimate due to controlling for the measured baseline.

Each estimate has improved precision as the length of follow-up increases (seen by the standard errors in Figure 7). The precision of the repeated scaled change estimate is highly influenced by the length of follow-up, having comparable precision to the modeled baseline model when there is a very long study, but much less precise estimates in a shorter study.

In Figure 8, we see that the precision of the longitudinal estimate from the modeled baseline model is highly affected by the length of follow-up. The cross-sectional estimate is also influenced, but to a lesser extent relative to the longitudinal effect. This is in part because a longer follow-up time will increase the measured temporal variation in the exposure caused by the population-level decrease in exposure over time. With a longer follow-up, there is also more information on the outcome change, and better inference may be made on how factors effect this rate of change.

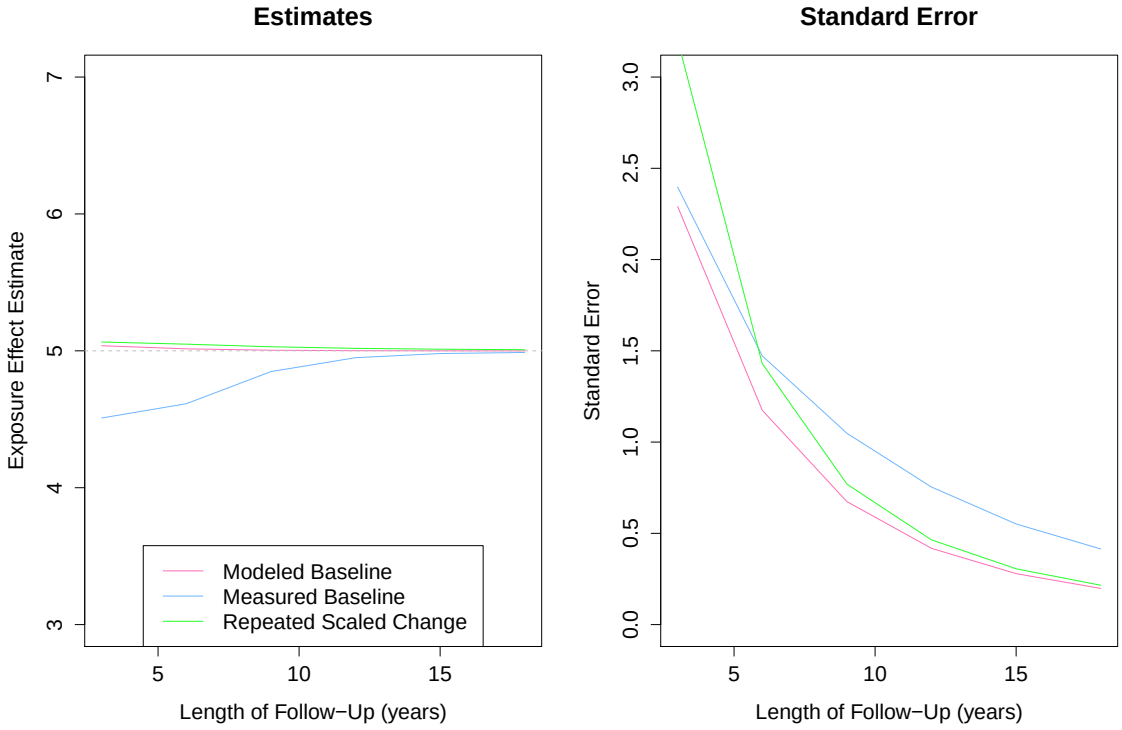

Figure 7: Influence of the length of follow-up on the bias and precision of the rate of change exposure effect estimates. Average exposure effect estimates (left plot) and standard error (right) from the modeled baseline, measured baseline and repeated scaled change mixed models. The horizontal dashed gray line gives the true value of the exposure effect ($\beta_1 = 5$).

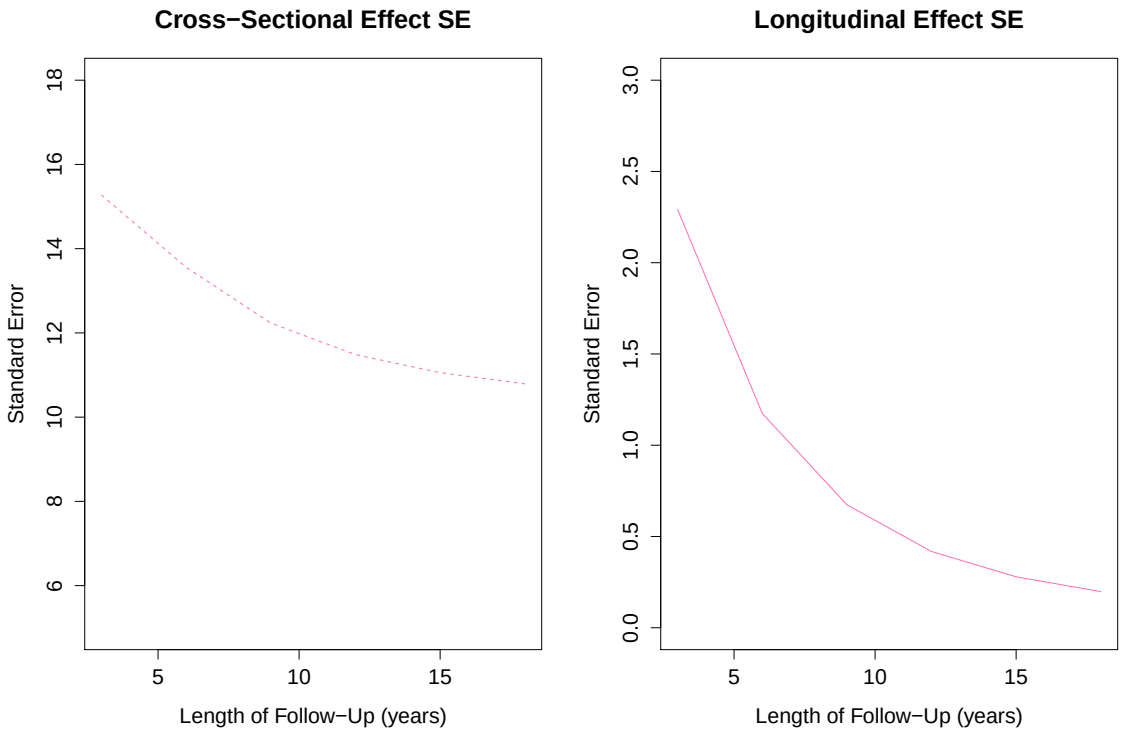

Figure 8: Influence of the length of follow-up on the precision of the cross-sectional and longitudinal exposure estimates from modeled baseline model. Standard error of the cross-sectional effect estimate (left plot) and longitudinal effect estimate (right) are provided.

C.2 Outcome Error

In our simulation framework the outcome residual error term ($\epsilon_{iv} \sim N(0, \sigma^2_\epsilon)$) and outcome measurement error term ($e_m \sim N(0, \sigma^2_{e_m})$) have the same form and are indistinguishable by the fitted models. They therefore influence the estimates similarly in our simulations, and so we only present results from varying the outcome residual variance.

In Figures 9, we see that increasing the outcome error leads to less precise exposure effect estimates. The modeled baseline and repeated scaled change models are unbiased across different levels of the outcome residual variance, but the measured baseline model experiences bias. In Section 3.1, we provide further graphs and discussion of why this bias has the shape it does in the graph.

In Figure 10, we see that the precision of both exposure effects from the modeled baseline model are influenced by the outcome error.

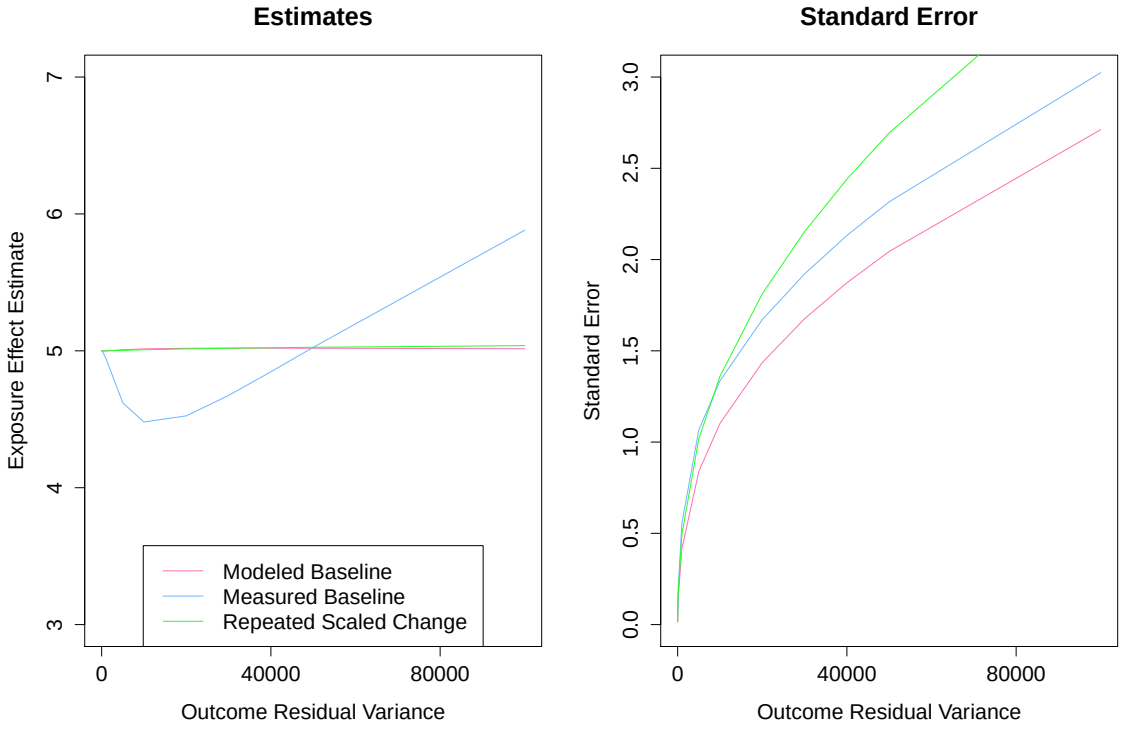

Figure 9: Influence of outcome error (either residual error or measurement error) on the bias and precision of the rate of change exposure effect estimates. Average exposure effect estimates (left plot) and standard error (right) from the modeled baseline, measured baseline and repeated scaled change mixed models. The horizontal dashed gray line gives the true value of the exposure effect ($\beta_1 = 5$).

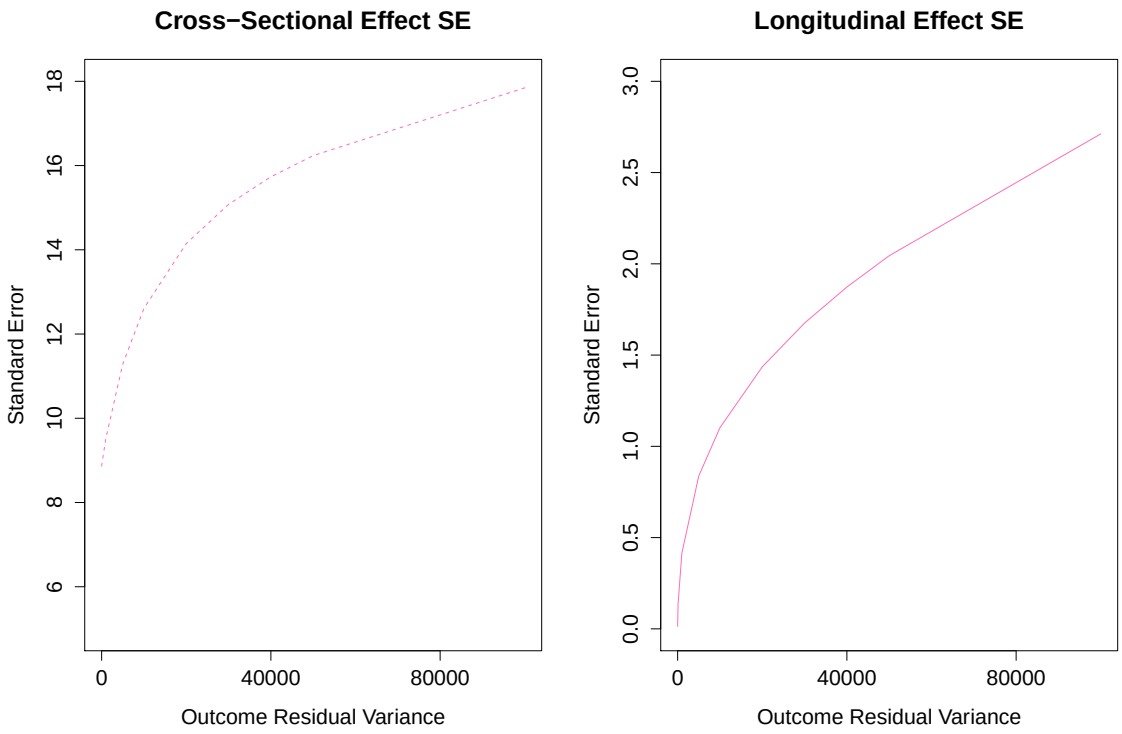

Figure 10: Influence of outcome error (either residual error or measurement error) on the precision of the cross-sectional and longitudinal exposure estimates from modeled baseline model. Standard error of the cross-sectional effect estimate (left plot) and longitudinal effect estimate (right) are provided.

C.3 Sample Size

We vary the number of subjects, while holding the number of follow-up measures and timing the same for all subjects.

In Figure 11, we see that the sample size does not noticeably affect the bias of any of the estimates, but does have a large influence on precision. As the sample size increases, the precision of each model's estimates improves, with the influence of additional subjects becoming less and less as the sample size increases.

In Figure 12, we see that the precision of the cross-sectional and longitudinal exposure effects are both highly influenced by sample size. This makes sense because additional subjects will increase the number of baseline observations as well as the number of follow-up observations, providing more information about both the outcome baseline and change in outcome over time.

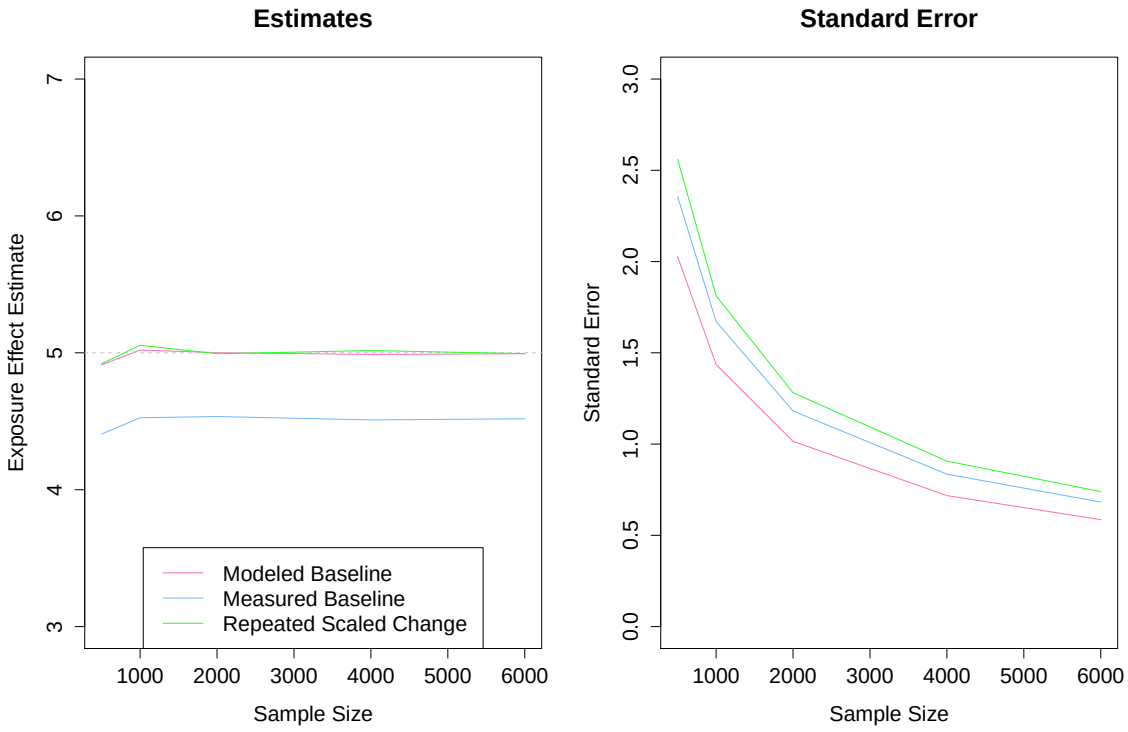

Figure 11: Influence of sample size on the bias and precision of the rate of change exposure effect estimates. Average exposure effect estimates (left plot) and standard error (right) from the modeled baseline, measured baseline and repeated scaled change mixed models. The horizontal dashed gray line gives the true value of the exposure effect ($\beta_1 = 5$).

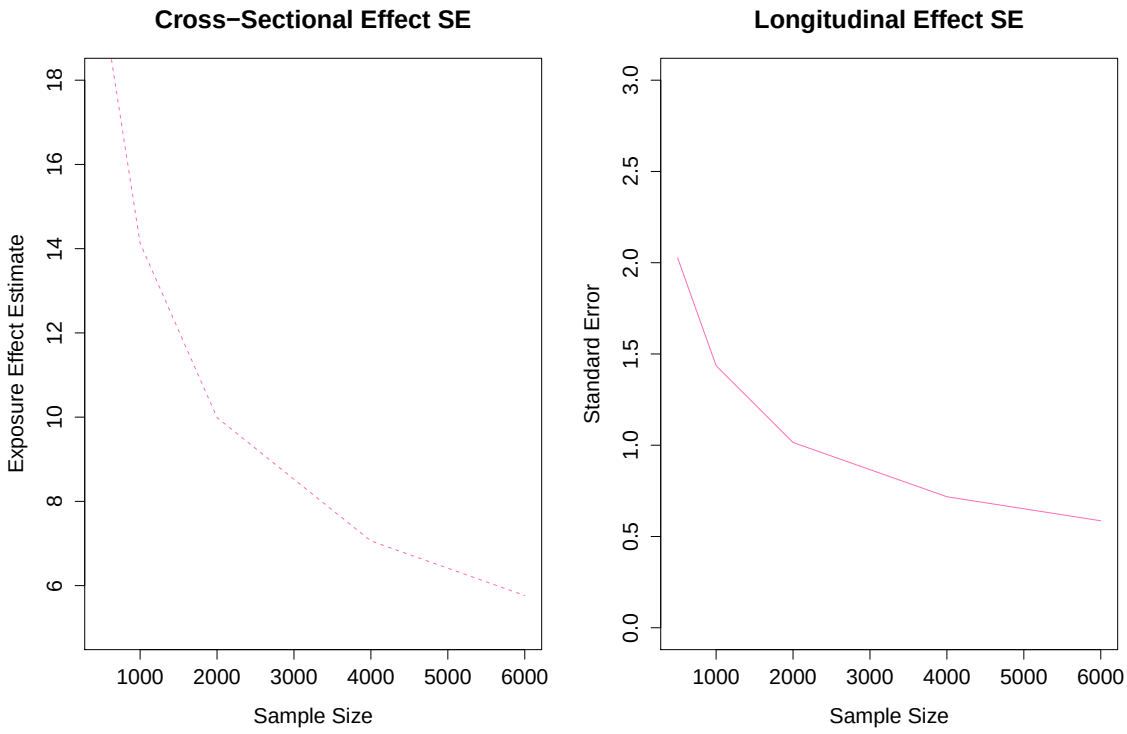

Figure 12: Influence of sample size on the precision of the cross-sectional and longitudinal exposure estimates from modeled baseline model. Standard error of the cross-sectional effect estimate (left plot) and longitudinal effect estimate (right) are provided.

C.4 Frequency of Follow-Up

We vary the number of visits over a five year follow-up. In each case the visits are equally spaced in the follow-up period (e.g. 4 visits would be at 0, 1.67, 3.33 and 5 years). We vary the total number of visits (i.e. including baseline visit) from 4 to 12. We did not consider 2 total visits (baseline and follow-up) or 3 visits because not all of the models could be fit with the same random effects structure if there were that few measurements per subject.

In Figures 13 and 14 we see that the there is relatively small influence of visit frequency on the estimates. However, there are several limitations we must consider. Our simulations assume there is no temporal autocorrelation structure in the random temporal variation in exposure between subjects, E_{time}, meaning that more frequent follow-ups will not reduce the random within-subject exposure variance ($\sigma^2_{E_{time}}$). In practice, some factors may in fact influence measurements the same regardless of the time between measurements (e.g. measurement error due to inaccuracy of measurement devices). However, we also expect that in many settings the exposure may have some sort of temporal autocorrelation structure.

We also assume that each subject has the same temporal trend of the exposure, E_{trend}, and that it is a linear decline. Furthermore, we assume the time-scale of the relevant exposure history (e.g. short-term vs. long-term exposure effect) is correctly specified. If we were to relax some of these assumptions, we may see that the frequency is more influential on the bias or precision of the exposure effect estimates. We look into aspects of the exposure specification in Section C.10.

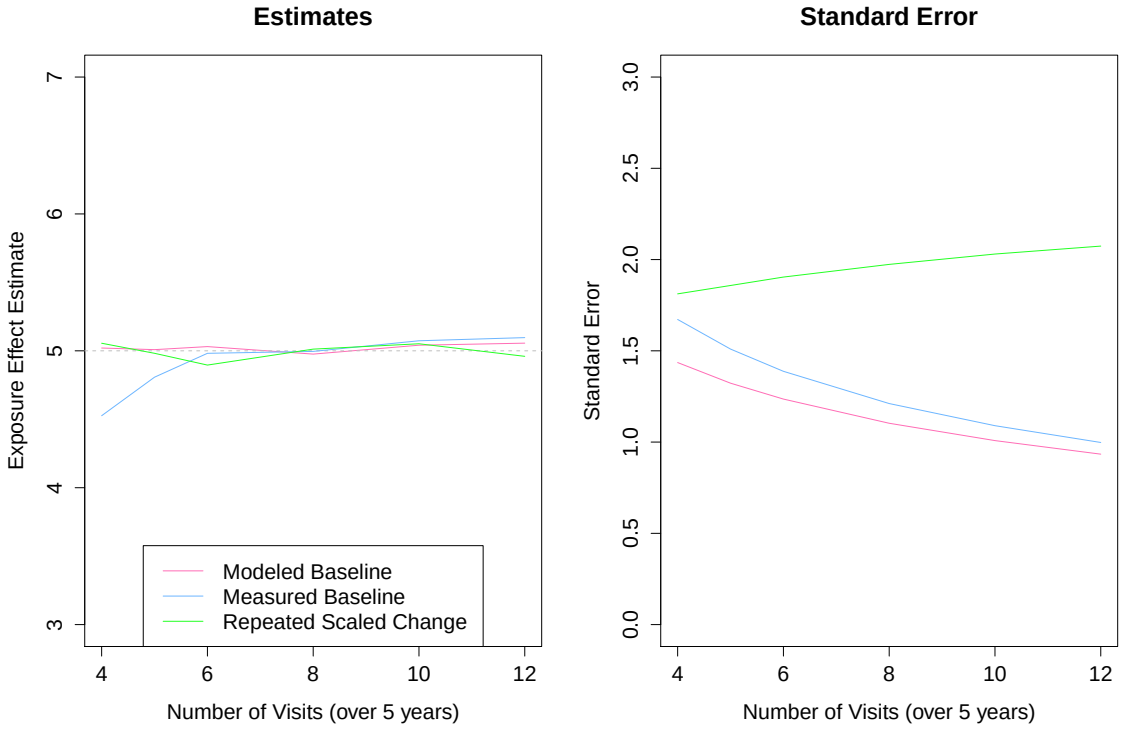

Figure 13: Influence of frequency of follow-up on the bias and precision of the rate of change exposure effect estimates. Average exposure effect estimates (left plot) and standard error (right) from the modeled baseline, measured baseline and repeated scaled change mixed models. The horizontal dashed gray line gives the true value of the exposure effect ($\beta_1 = 5$).

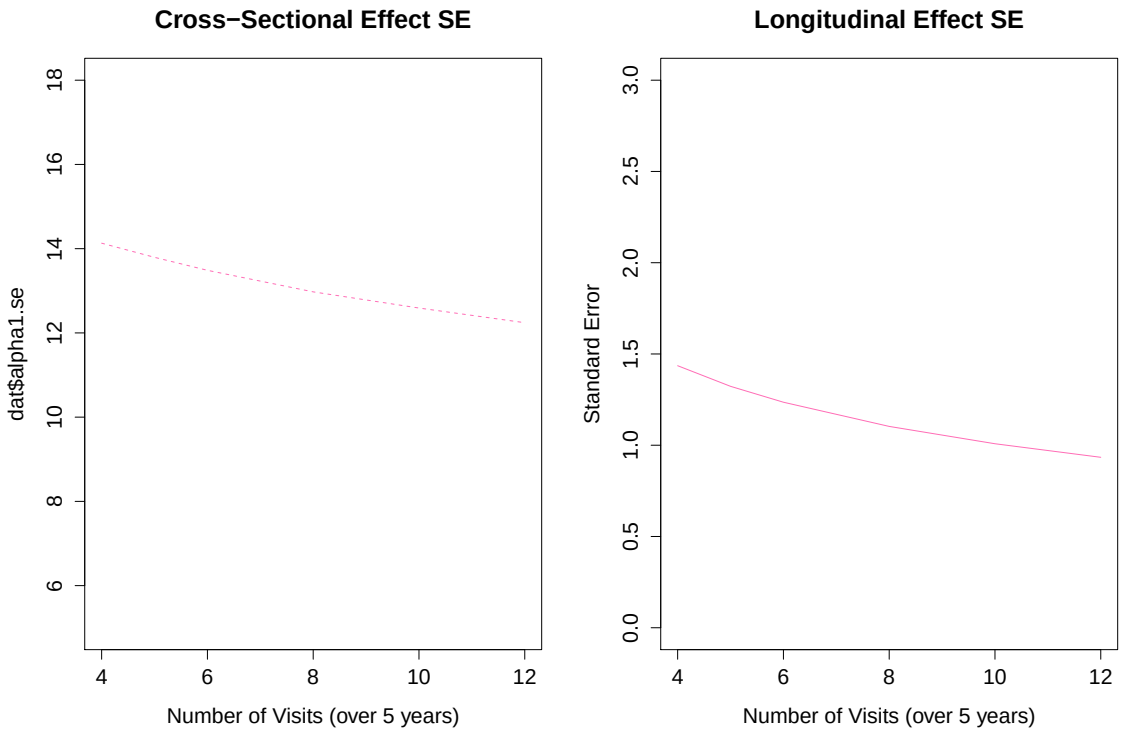

Figure 14: Influence of frequency of follow-up on the precision of the cross-sectional and longitudinal exposure estimates from modeled baseline model. Standard error of the cross-sectional effect estimate (left plot) and longitudinal effect estimate (right) are provided.

C.5 Variance of Random Effects

Random effects allow us to account for differences by subject in our fitted models that cannot be explained by the covariates in the models. In our simulations, we may use random effects in the outcome generating model as a tool to capture between subject differences. By varying the size of the random effects, we vary the amount of information that cannot be captured by the measured covariates. In this sense, random effects can be thought of as similar to adding unmeasured covariates to the model, with a few important restrictions:

- Random effects are independent of other covariates, including the exposure.
- Subject-specific random effects are constant for each subject (in our framework), similar to a time-constant variable.
- Random effects have a defined variance-covariance structure,

$$\Sigma_{rand} = \begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{bmatrix}.$$

In simulations we vary the variance of the random intercept and slope separately, while adjusting the covariance to hold the correlation constant. Varying the covariance of the random effects did not have an effect on the estimates.

In Figures 15 and 16, we see that the variance of the random intercept does not have a strong influence on the precision of the rate of change exposure estimates. The random intercept variance also does not bias the modeled baseline or repeated scaled change estimates, but does influences the bias of the measured baseline estimate. This is likely a result of the exposure and measurement error making up larger proportions of the outcome variance at baseline when there is a small random intercept variance. The random intercept is also highly influential on the precision of the cross-sectional

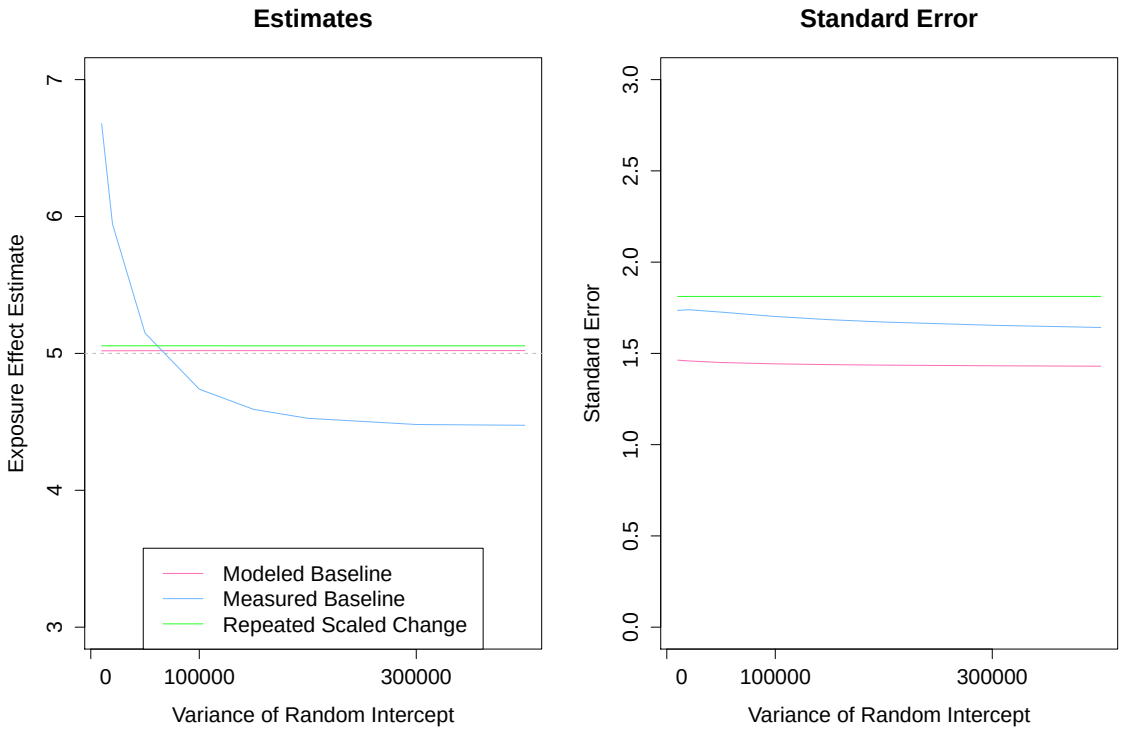

Figure 15: Influence of random intercept variance on the bias and precision of the rate of change exposure effect estimates. Average exposure effect estimates (left plot) and standard error (right) from the modeled baseline, measured baseline and repeated scaled change mixed models. The horizontal dashed gray line gives the true value of the exposure effect ($\beta_1 = 5$).

exposure estimate from the modeled baseline model.

In Figures 17 and 18, we see that the random slope affects the precision of the rate of change estimates. We do not see a strong influence, but this may be due to the conditions (i.e. follow-up time, exposure variation etc.) being otherwise fairly good for rate of change estimation. The bias of the three rate of change estimates is not largely influenced by the random slope.

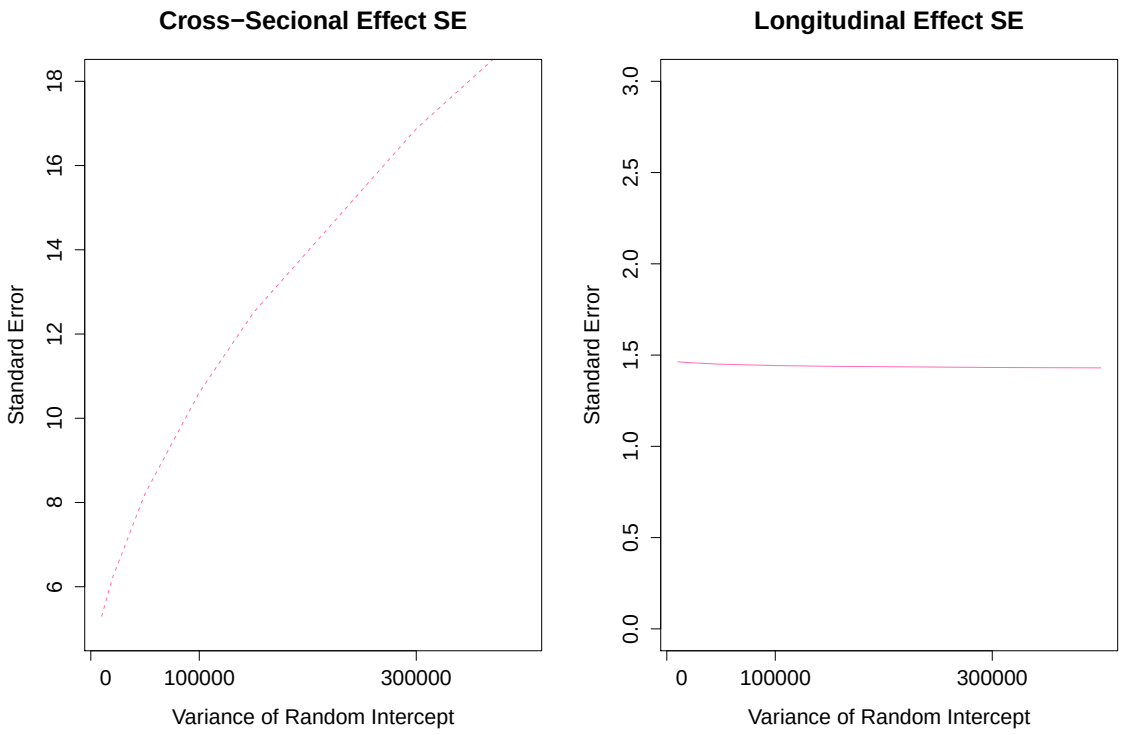

Figure 16: Influence of random intercept variance on the precision of the cross-sectional and longitudinal exposure estimates from modeled baseline model. Standard error of the cross-sectional effect estimate (left plot) and longitudinal effect estimate (right) are provided.

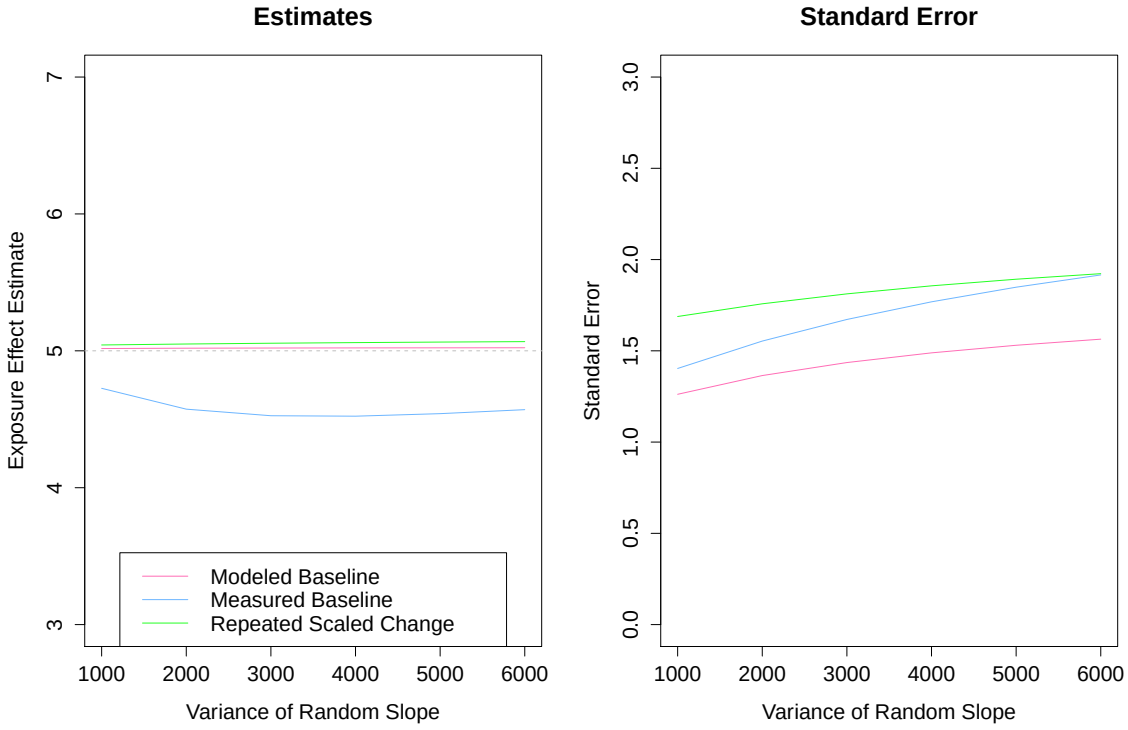

Figure 17: Influence of random slope variance on the bias and precision of the rate of change exposure effect estimates. Average exposure effect estimates (left plot) and standard error (right) from the modeled baseline, measured baseline and repeated scaled change mixed models. The horizontal dashed gray line gives the true value of the exposure effect ($\beta_1 = 5$).

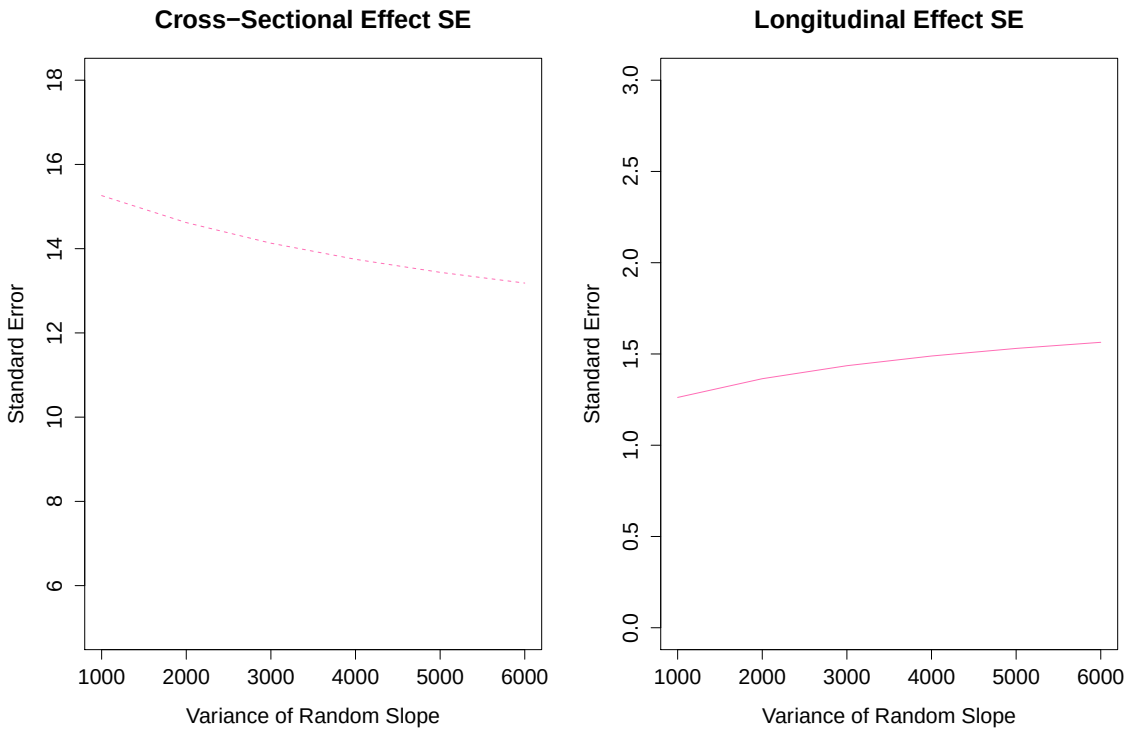

Figure 18: Influence of random slope variance on the precision of the cross-sectional and longitudinal exposure estimates from modeled baseline model. Standard error of the cross-sectional effect estimate (left plot) and longitudinal effect estimate (right) are provided.

C.6 Exposure Variation: Trend Correlated with Time

Often in practice, the average levels of an exposure in the population may increase or decrease over time. For example, in the MESA Air study, Baltimore experienced a decrease in particulate matter over the study time. In our simulations we assume the population trend is a linear function of time, and vary the slope of this trend. In Section C.10 we consider more complex exposure trends, as well as the relevant time scale of the exposure (i.e. short-term, long-term and cumulative effects).

In Figure 19, we see that the precision of the rate of change estimates improves as the slope of the population trend increases. This may be explained by more temporal variation in the exposure when there is a steeper linear trend. When the exposure either declines or increases over time, the standard error of the estimates may be slightly inflated by the correlation between the exposure with time. However, this is overpowered by the decrease in standard error resulting from increased exposure variability. The direction of the slope (i.e. increase or decrease in exposure) does not make a difference for the precision of estimates. The repeated scaled change model is influenced most drastically, having a much higher standard error with little or no population trend. The precision of the measured baseline estimate is not influenced as much as the other two estimates. The slope of the exposure trend does not have a large influence on the bias of the rate of change exposure effect estimates. In the modeled baseline model (Figure 20), the cross-sectional estimate is barely affected by the slope, while the precision of the longitudinal term is affected.

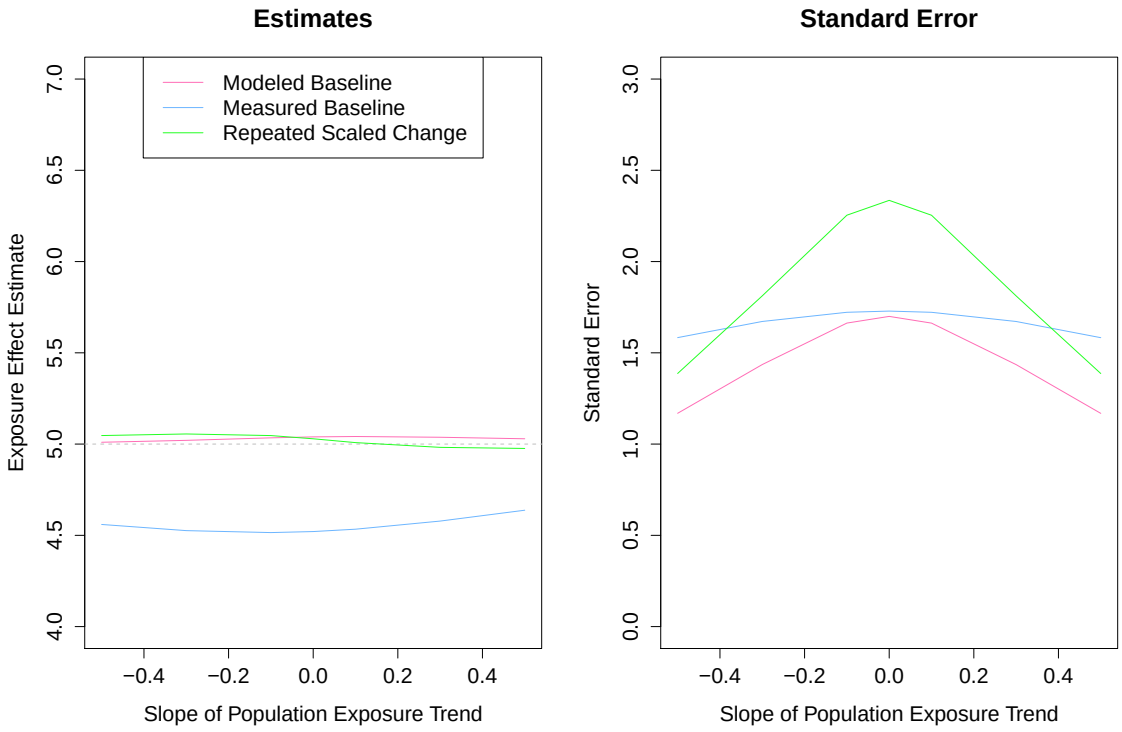

Figure 19: Influence of the slope of the population temporal exposure trend on the bias and precision of the rate of change exposure effect estimates. Average exposure effect estimates (left plot) and standard error (right) from the modeled baseline, measured baseline and repeated scaled change mixed models. The horizontal dashed gray line gives the true value of the exposure effect ($\beta_1 = 5$).

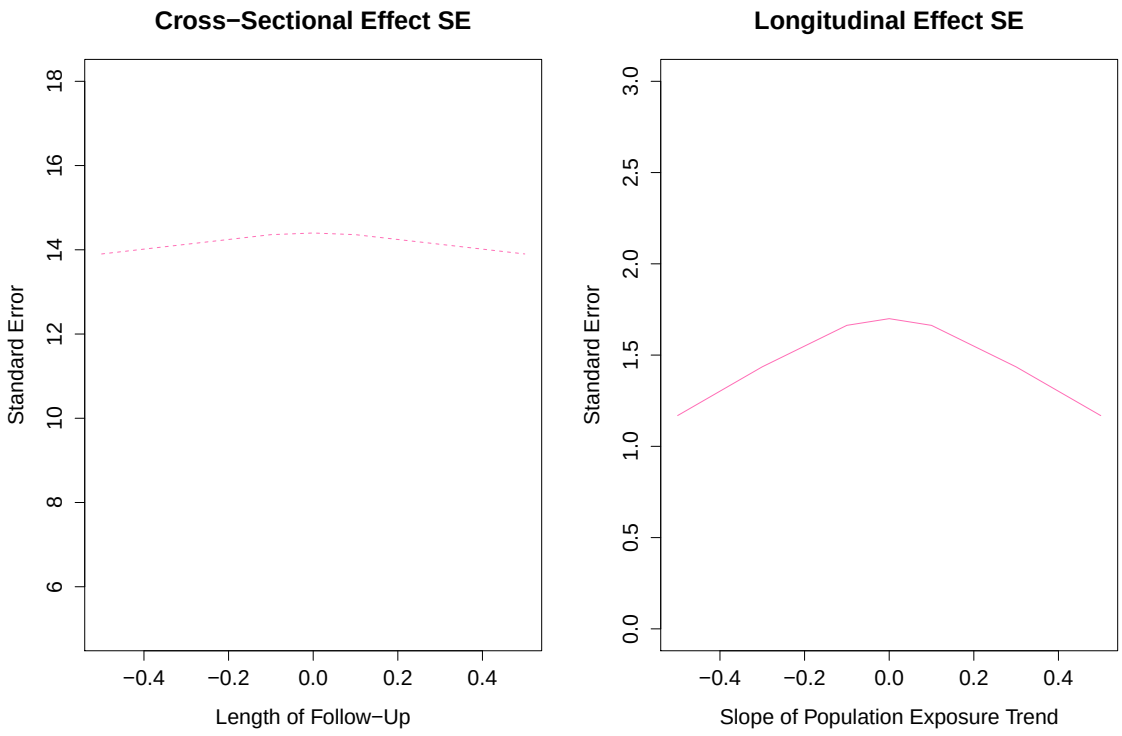

Figure 20: Influence of the slope of the population temporal exposure trend on the precision of the cross-sectional and longitudinal exposure estimates from modeled baseline model. Standard error of the cross-sectional effect estimate (left plot) and longitudinal effect estimate (right) are provided.

C.7 Exposure Variation: Random Within-Subject Temporal Variation

We vary the amount of random within-subject temporal variation of the exposure. This variation can represent changes in individual exposure levels over time not accounted for by the average population trend. For example, employment location or changes in activities during the study period may result in changes in exposure levels. Variability in factors such as these, which we assume to be random, contributes to the across-subject variation at any given measurement time, as well as to the within-subject (temporal) variation of an individual. Note that in these simulations we do not clearly define the time-scale of the exposure at each measurement time. That is, we do not specify whether the exposure measurement represents a short-term average, long-term average, or cumulative measure (e.g. average since baseline). We explore these characteristics in Section C.10. However, without more clearly defining the time-scale and by assuming the scales of the fitted models are correctly specified, we may still learn how this random temporal variation influences exposure effect estimates in general.

In Figure 21, we see that when there is more random temporal exposure variation, all rate of change exposure estimates are more precise. The measured baseline estimate experiences bias with low amounts of variation, and is less biased when there is more variation.

In Figure 22, we see that both of the modeled baseline exposure estimates have improved precision when there is more exposure variation. This makes sense because the variation contributes to the across-subject exposure variation at baseline as well as the temporal exposure variation within-subjects. The cross-sectional estimate is strongly influenced by the exposure variation at baseline, and the longitudinal estimate is affected by both the baseline and temporal variation in the exposure.

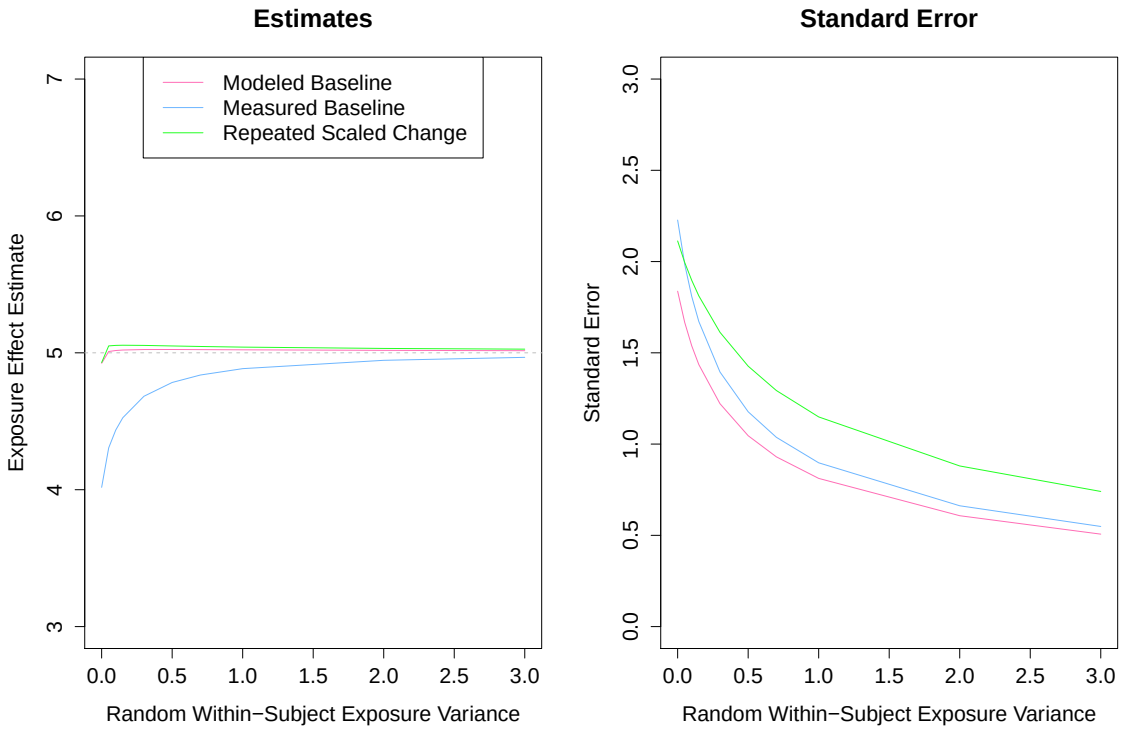

Figure 21: Influence of the random temporal exposure variation on the bias and precision of the rate of change exposure effect estimates. Average exposure effect estimates (left plot) and standard error (right) from the modeled baseline, measured baseline and repeated scaled change mixed models. The horizontal dashed gray line gives the true value of the exposure effect ($\beta_1 = 5$).

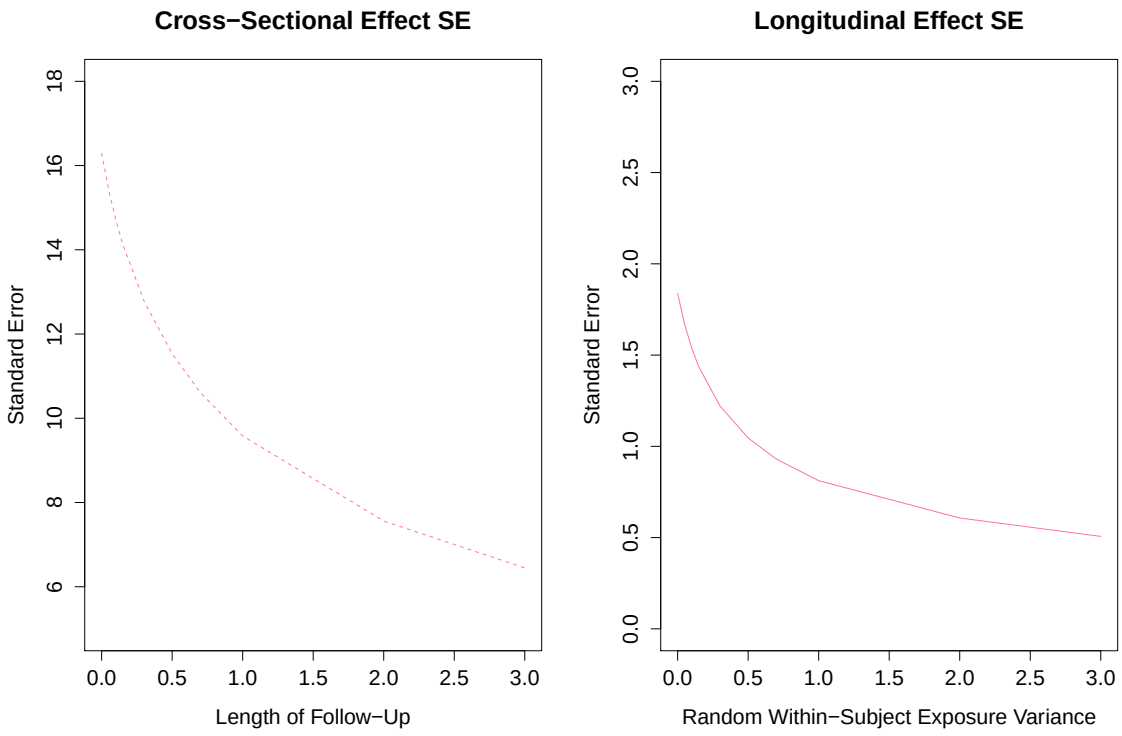

Figure 22: Influence of the random temporal exposure variation on the precision of the cross-sectional and longitudinal exposure estimates from modeled baseline model. Standard error of the cross-sectional effect estimate (left plot) and longitudinal effect estimate (right) are provided.

C.8 Exposure Variation: Across-Subject Variation

We consider the across-subject exposure variation, $\sigma^2_{subject}$. This is the variance of the subject-specific average deviations from the population average exposure level, $E_{subject}$.

In Figure 23, we see that the rate of change exposure effect estimates are more precise when there is larger across-subject exposure variation. The bias of the measured baseline estimate is larger when there is more across-subject variation.

In Figure 24, we see that the across-subject variation is incredibly influential on the precision of the modeled baseline cross-sectional exposure effect estimate. The across-subject variation also has some effect on the precision of the longitudinal exposure effect estimate, but it is not as influential as on the cross-sectional effect. This makes sense because the variation only contributes to across-subject exposure variation and not to temporal variation.

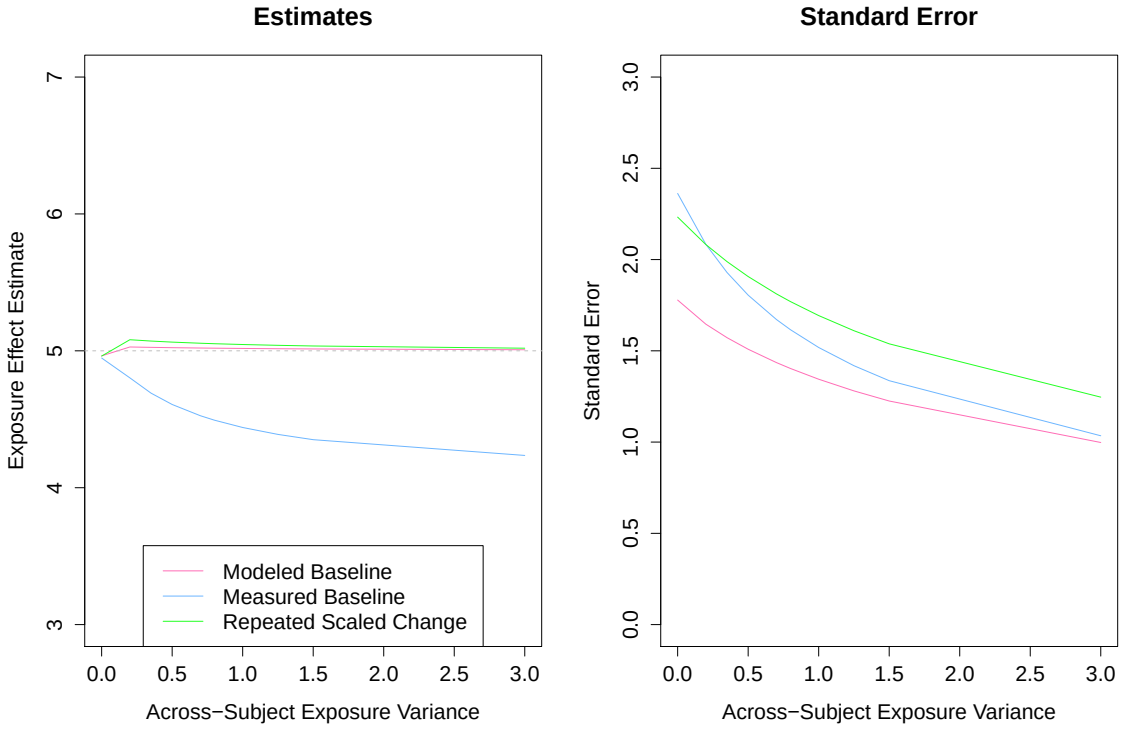

Figure 23: Influence of the across-subject exposure variation on the bias and precision of the rate of change exposure effect estimates. Average exposure effect estimates (left plot) and standard error (right) from the modeled baseline, measured baseline and repeated scaled change mixed models. The horizontal dashed gray line gives the true value of the exposure effect ($\beta_1 = 5$).

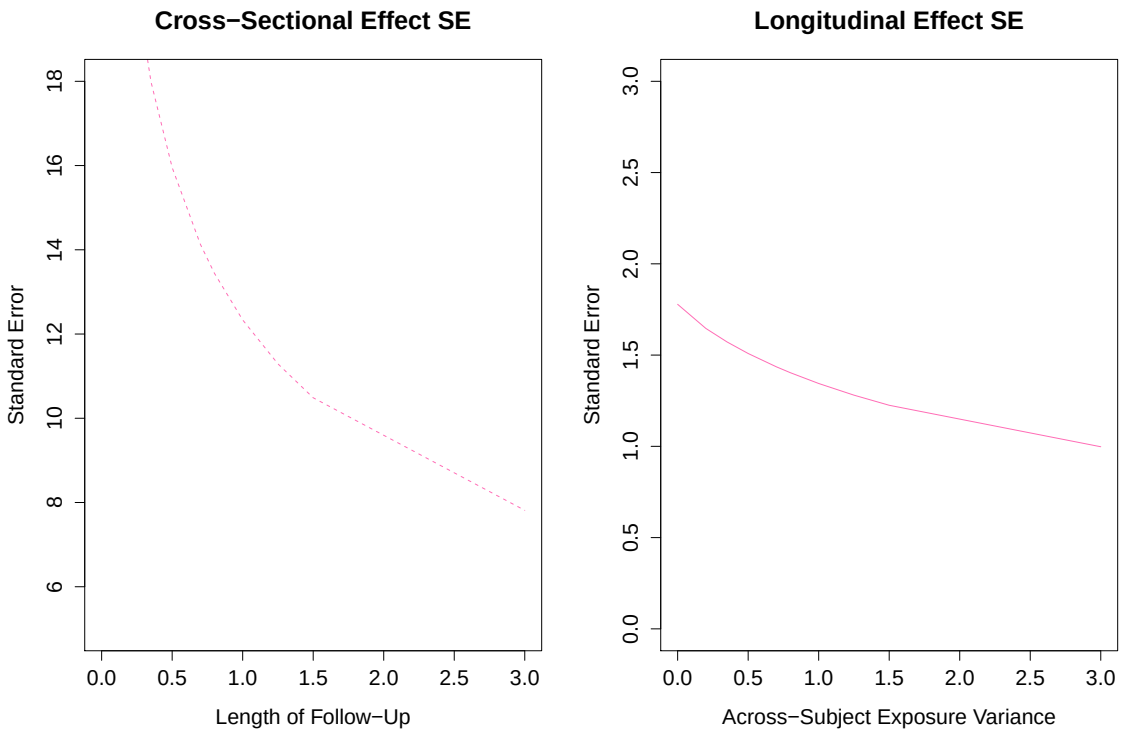

Figure 24: Influence of the across-subject exposure variation on the precision of the cross-sectional and longitudinal exposure estimates from modeled baseline model. Standard error of the cross-sectional effect estimate (left plot) and longitudinal effect estimate (right) are provided.

C.9 Unadjusted Confounding

We consider the effect of unadjusted confounders on our estimates.

A confounder X_{iv} is included in the true outcome model as show in Section B.1, but is not adjusted for in the fitted models. Therefore, if α_2 or β_2 are non-zero and X_{iv} is correlated with the exposure, E_{iv}, there will be a confounding effect. We generated the variable X_{iv} such that the correlation between X_{iv} and E_{iv} would be ρ. The additional covariate has the form

$$X_{iv} \sim \mathrm{N}\left(\frac{\rho E_{iv}}{\sigma_E}, 1-\rho^2\right)$$

where σ_E is the total standard deviation of the exposure variable. It follows that X_{iv} has a variance of 1. The mean is 0 at baseline (because the mean of E_{i0} is 0) and changes at the same rate as the population average exposure.

We consider when the confounder effects the baseline outcome (i.e. $\alpha_2 \neq 0$) and outcome rate of change (i.e. $\beta_2 \neq 0$) separately. In either case, we set the the cross-sectional or longitudinal effect of X_{iv} (i.e. α_2 and β_2) to be equal in magnitude to the corresponding exposure effect (i.e. α_1 or β_1). We consider effects that are in the same and opposite directions of the exposure effects, and control the strength of confounding with the correlation between the exposure and confounder, ρ.

The effect of confounding on estimator precision is unimportant compared to the effect on estimator bias, and so we only present the effects on bias.

We find that all of the rate of change estimates from the three models (i.e. $\hat{\beta}_1$,$\hat{\beta}_1^*$ and $\hat{\beta}_1^{**}$) are similarly affected by longitudinal confounding (Figure 25). The exposure effect is overestimated (positive bias) if either the confounder has an effect in the same direction as the exposure (e.g. both increase the rate of CAC progression) and they are positively correlated, or if the effects are in opposite directions but the correlation is negative. In the case of the measured baseline model, the bias

from confounding may either cancel out or add to the bias due to controlling for the measured baseline in the presence of measurement error. These two causes of bias do not depend on each other. However, the measured baseline model is also biased from cross-sectional confounding, whereas the other rate of change estimates are only biased by longitudinal confounding. This shows that controlling for the measured baseline outcome not only adds bias when measurement error is present, but can potentially have more bias due to confounding than models that don't adjust for the measured baseline. This is supported by the directed acyclic graphs (DAG's) in Section A which are based off of similar DAG's in Lepage et al. [2015].

In Figure 26, we see that the modeled baseline cross-sectional estimate is only biased by cross-sectional confounding and the longitudinal estimate is only biased by longitudinal confounding. This is a potentially advantageous property of the model, since we are primarily interested in the longitudinal estimate, and in general we expect cross-sectional confounding to be more of a concern since the baseline outcome may be influenced by many long-term, unmeasured or unaccounted for factors. Our simulations suggest that even if the cross-sectional exposure estimate is biased from confounding, it does not necessarily mean the longitudinal estimate of interest is. However, we expect that most variables that affect the outcome baseline may also affect the outcome rate of change, and so confounding is still a legitimate concern.

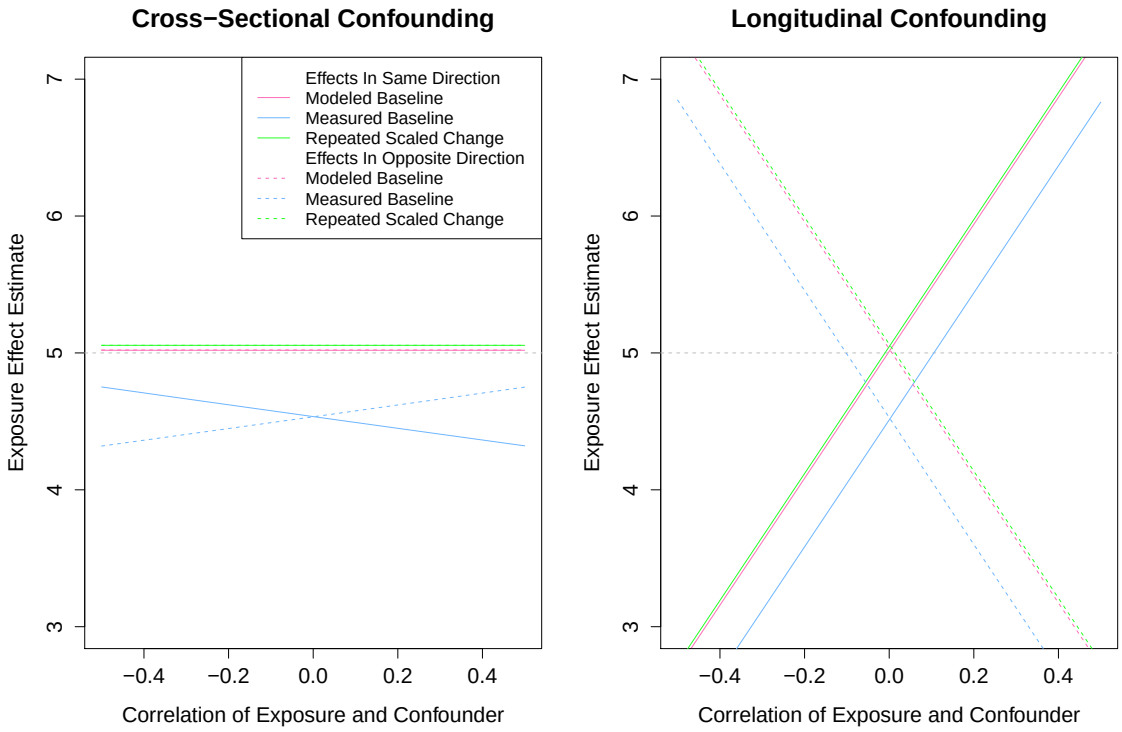

Figure 25: Influence of unmeasured confounding on the bias of rate of change exposure effect estimates from the modeled baseline, measured baseline and repeated scaled change mixed models (i.e. $\hat{\beta}_1$,$\hat{\beta}_1^*$ and $\hat{\beta}_1^{**}$). The left plot shows cross-sectional confounding and the right plot shows longitudinal confounding (determined by whether the confounder affects the baseline outcome or outcome rate of change). The horizontal gray lines show the true value of exposure effect on rate of change being estimated ($\beta_1 = 5$) Lines are given for when the confounder and exposure have effects in the same direction as well as when the effects are in opposite directions.

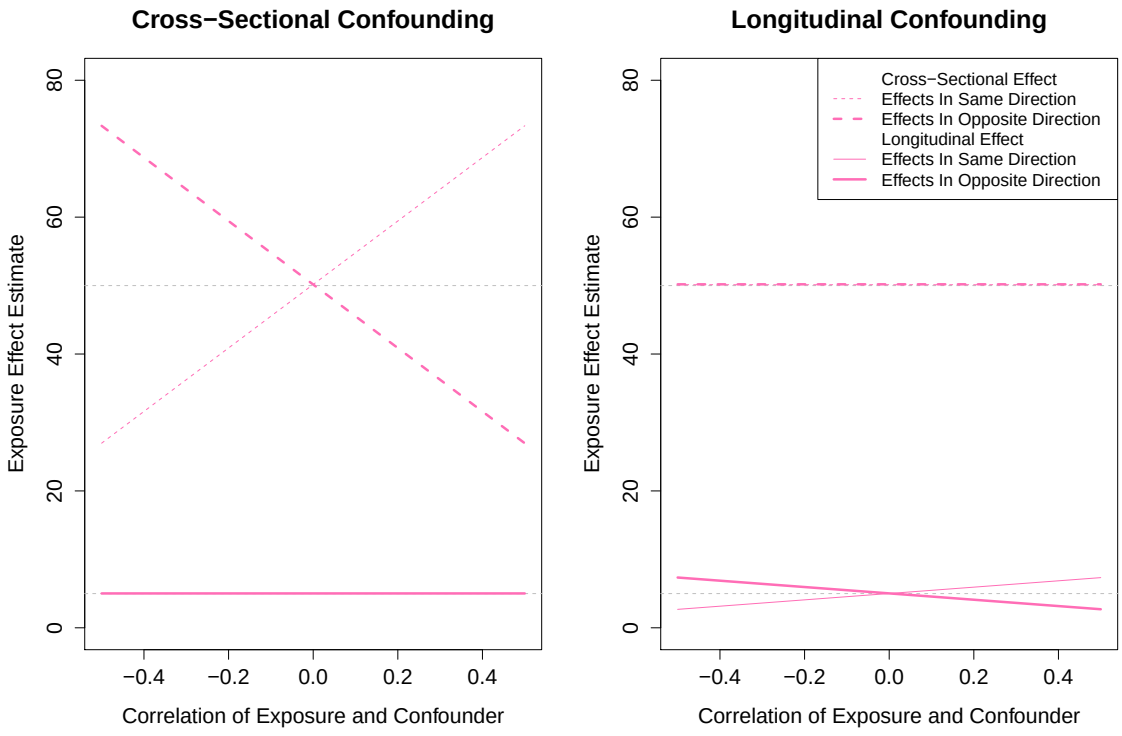

Figure 26: Influence of unmeasured confounding on the bias of the modeled baseline cross-sectional and longitudinal effect estimates (i.e. $\hat{\alpha}_1$ and $\hat{\beta}_1$). The left plot shows cross-sectional confounding and the right plot shows longitudinal confounding (determined by whether the confounder affects the baseline outcome or outcome rate of change). The horizontal gray lines show the true value of exposure effects being estimated ($\alpha_1 = 50$ and $\beta_1 = 5$) Lines are given for when the confounder and exposure have effects in the same direction as well as when the effects are in opposite directions.

C.10 Exposure Specification

In the Supplement Sections C.6, C.7 and C.8, we considered how the characteristics of the exposure affected the estimates from correctly specified models. In all the simulations of the main paper and Supplement, the exposure variable had the framework $E = E_{trend} + E_{subject} + E_{time}$, with the terms E_{trend}, $E_{subject}$, and E_{time} representing a population average trend, subject-specific deviation from the population trend, and subject-specific random temporal variation, respectively. We now expand on this model framework, exploring different population and subject-specific temporal trends as well as considering new mechanisms of how the exposure affects the outcome. We also consider the specification of the exposure in the fitted model, and how misspecification of the exposure may influence estimator precision and accuracy.

Our model and parameter values are the same as previous simulations with the exception of how exposure is modeled and the follow-up length and visit timing. We simulate a study that follows patients for 10 years with a visit every 2 years (6 total visits). Exposure values are generated every 6 months starting 10 years before the study and throughout the study period (20 total exposure times). These values may be interpreted as the average exposure over the previous 6 month period. The exposure variable which affects the outcome at each measurement time is some average of these 6 month average exposure measurements.

Four different exposure trends will be used: constant, linear decline, quadratic decline and linear decline with varied subject-specific slopes. Figure 27 illustrates the four exposure trend scenarios with lines showing examples of exposure values for specific subjects. In each scenario, the population average exposure trend will have a mean of zero across the 20-year period starting 10 years prior to baseline, and ending after 10 years of follow-up. For the constant trend, $E_{trend} = 0$ for all subjects. The linear decline represents a decline of $0.3\frac{\mu g}{m^3}$ in $PM_{2.5}$ per year, representing a constant reduction in air pollution that is the same for all subjects. The quadratic decline has

the same decrease in 20 years, but 75% of the decline occurs in the 10 years prior to the study and 25% of the decline occurs during the 10 year study period. This may represent a study in a geographical location which has experienced a large decline in air pollution recently, but with the rate of air quality improvement slowing down. The linear decline with varying slopes has a population average exposure trend equivalent to the linear decline mentioned above (decline of $6\frac{\mu g}{m^3}$ in $PM_{2.5}$ over 20 years), but each subject also has their own specific rate of decline. The subject-specific rates of decline were determined by adding a multiplier generated from the uniform distribution so that the declines varied between no decline and twice as large of a decline (a range of average declines between 0 to 12 $\frac{\mu g}{m^3}$ in $PM_{2.5}$ over 20 years). Across simulations, we maintain the same framework for the variance of the subject-specific deviation from the population trend ($E_{subject} \sim N(0, \sigma^2_{E_a})$) and the subject-specific random temporal trend ($E_{time} \sim N(0, \sigma^2_{E_w})$). However, note that for the linear decline with subject-specific slopes, the total across-subject variation will now also depend on the measurement time, since subjects with different rates of exposure decline will be farther apart the farther away the time is from baseline.

In addition to modeling the exposure levels for each subject, we must define the mechanism by which the exposure affects the outcome. In the previous simulations in the main paper and Supplement, we did not clearly define the time-scale of the exposure, and whether each measurement time represents a short term average, long term average, or cumulative measure. We simply were exploring how the different exposure variation characteristics of these measurements influenced our estimates. However, we now have a more defined framework of exposure measurements which can be interpreted as a 6 month average, and we can explore different time-scales in which the exposure affects the outcome. Depending on the scientific context, there may be reason to believe that the exposure effect will depend on a short or long exposure history. We consider three different exposure mechanisms: a short-term

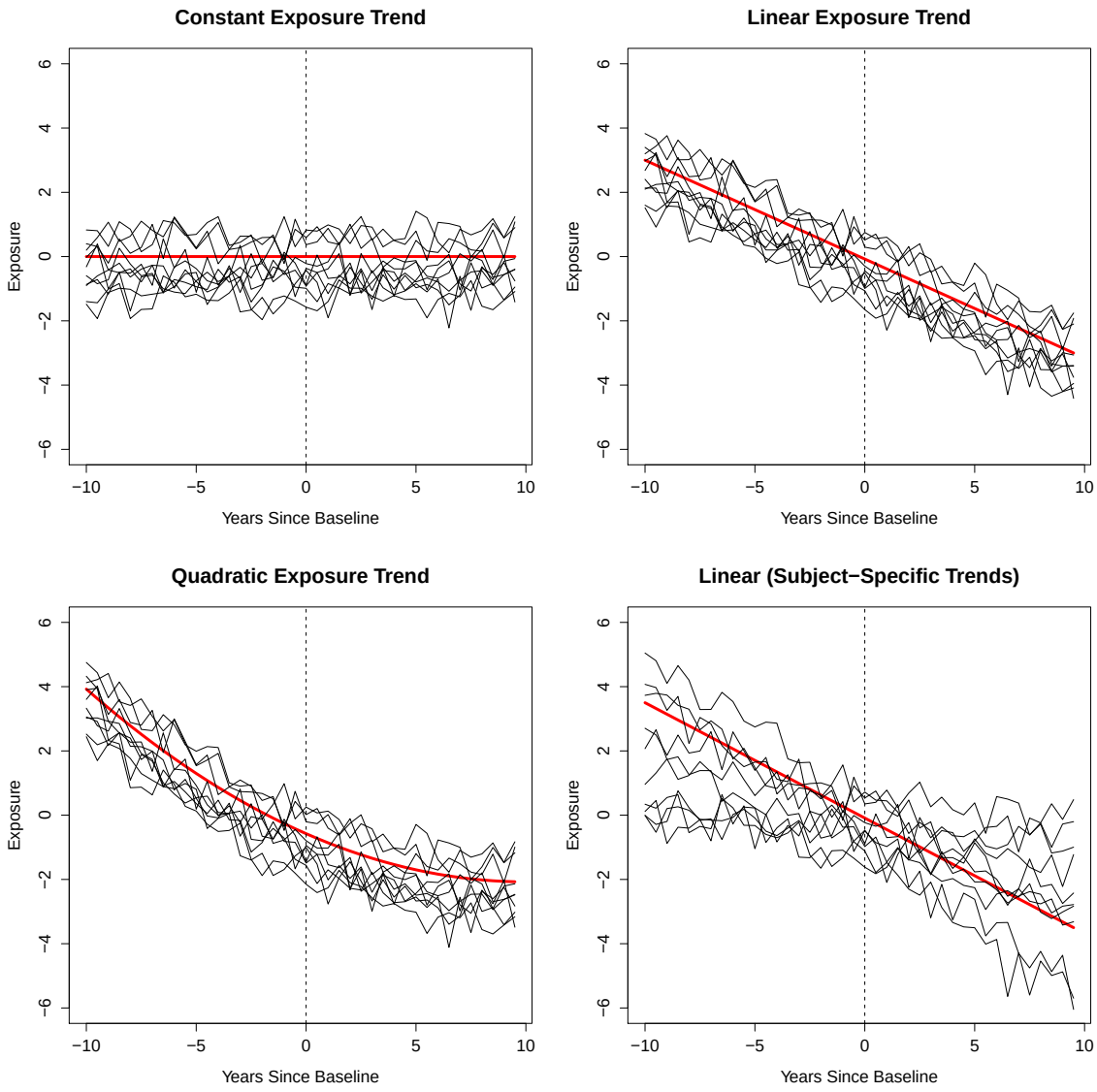

Figure 27: Example simulated exposure values for the simulations in the Supplement Section C.10. Exposure values are simulated starting 10 years prior to the beginning of the study and throughout a 10 year study period (20 years total). The red line gives the population average trend, and the black lines give exposure levels for specific subjects. The plots show population average trends that are constant (top left), a linear decline (top right), a quadratic decline (bottom left), and a linear decline with subject-specific rates of decline (bottom right). The vertical dashed line marks the beginning of the study.

moving average (average exposure levels in the 2 years prior to the measurement time), a long-term moving average (average exposure levels in the 10 years prior to the measurement time), and a cumulative mechanism (average exposure levels from 10 years prior to the beginning of the study and the measurement time). An illustration of these three mechanisms is given in Figure 28. Note that for the long-term mechanisms which average over more than 2 years of the exposure history, there will be overlap in the time frame that exposure is averaged over, and the exposure levels at a particular time may influence multiple outcome measurement times. This matters as it can give exposure levels from certain time periods more influence, as we discuss later.

Finally, we consider how the exposure is specified in the fitted models. We will consider five exposure specifications:

- **Time-constant/Baseline** (cross-sectional and longitudinal exposure are both specified as the average exposure levels in the two years prior to the beginning of the study). This type of variable may be used in a study with minimal exposure measurements/modeling, and will ignore any variation in exposure over time. We also considered an average of the 10 years prior to baseline, which produced fairly similar longitudinal exposure estimates.

- **"6 month moving average"**. (cross-sectional and longitudinal exposure are both the average exposure levels in the 6 months prior to the visit time) This method is the extreme for a short-term exposure. This is the framework used in the other simulations, although in that case the exposure measurements did not necessarily represent a 6 month average.

- **2 year moving average** (cross-sectional and longitudinal exposure are measured by the average exposure levels of the 2 years prior to the measurement time).

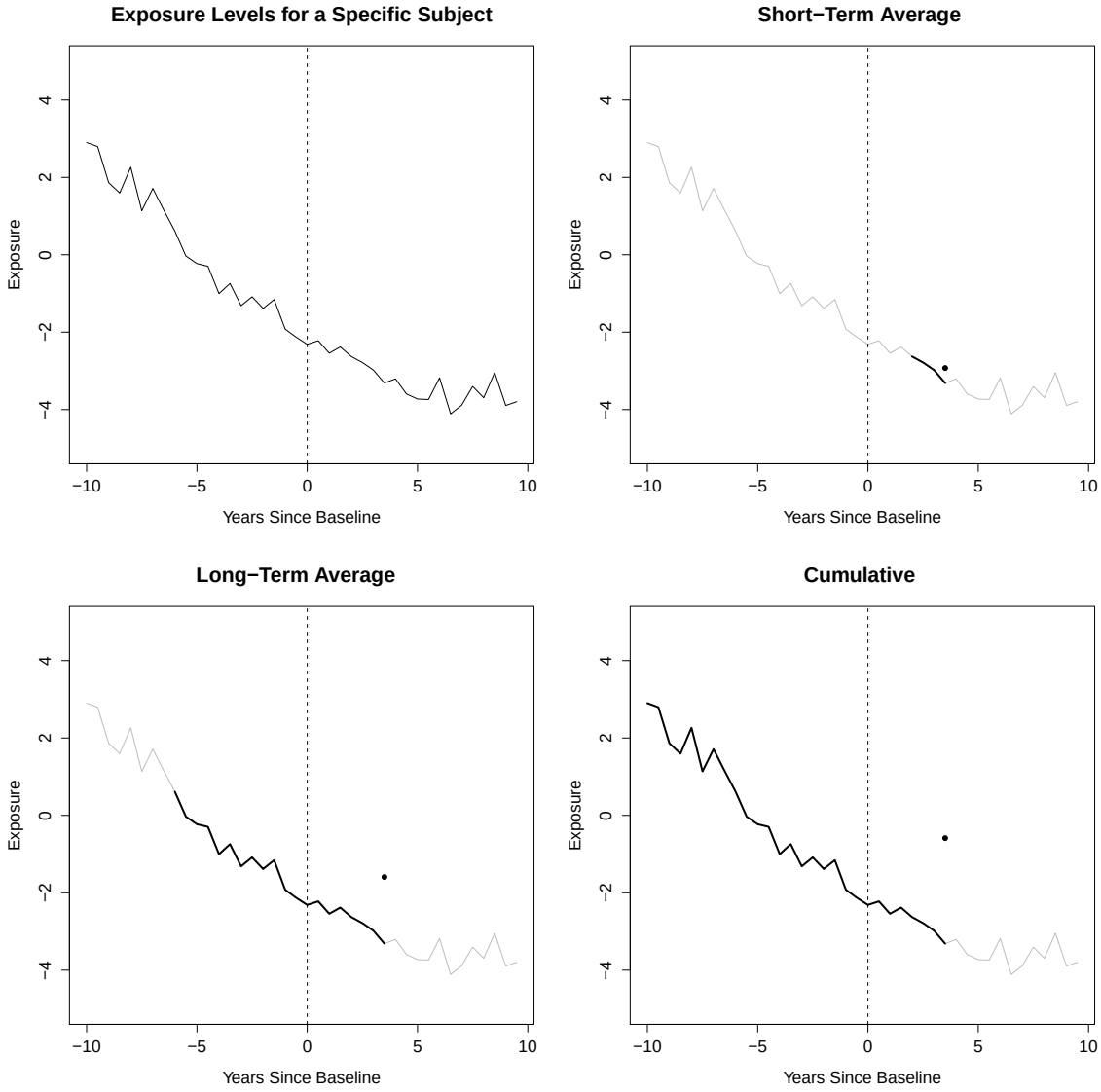

Figure 28: Illustration of exposure mechanisms. An example of exposure levels for a particular subject are given in the top left figure. The three other figures illustrate how the three exposure mechanisms use the exposure history in defining the exposure variable at a specific time (4 years after baseline). The dark portion of the line is the exposure history which the exposure mechanism depends on. For a short-term average, this is the 2 year period prior to the measurement time. For the long-term average, a 10 year period is used, and for the cumulative mechanism, the whole period from 10 years prior to the beginning of the study to the measurement time is used. The points shows the average exposure level in the respective time periods, and these average values are the exposure values used in the outcome model. The vertical dashed line shows the beginning of the study period.

- **10 year moving average** (cross-sectional and longitudinal exposure are measured by the average exposure levels of the 10 years prior to the measurement time).

- **Average since baseline** (longitudinal exposure is measured by the average exposure levels of the period between the baseline and measurement time, and the cross-sectional (baseline) exposure value is the average exposure levels of the 2 years prior to baseline). This method has exposure measurements that are interpreted as the cumulative exposure since baseline. Although this model is cumulative, it is a different cumulative specification than the cumulative specification used in the true outcome model (cumulative since 10 years prior to the study). We also considered an similar model with the cross-sectional (baseline) exposure being the average of the 10 years prior to baseline. This model produced fairly similar longitudinal exposure estimates.

From Tables 6 and 7, we see that when the exposure trend is constant, all fitted models perform similarly for the most part. This is because with little variation over time, the main contrast in exposure comes from across-subject comparisons, which all exposure specifications will capture similarly. If there were larger amounts of within-subject variation over time, we would expect to see more differences in the model performances as the longer-term averages will have less variability. When we consider the other three models with exposure trends correlated with time, the within-subject variation becomes more influential, and we find that much of the differences in model performance can be explained by how the underlying effect of exposure on the outcome and its time scale give different weight to the exposure levels from different time periods.

First, we will consider the fitted models with time-constant baseline exposure specification (First column of Tables 6 and 7). In general, the time-constant fitted models have worse precision than the time-varying models (Table 7), as they rely

only on contrasts across-subjects, and do not benefit from information about the exposure variation over time. If there is an exposure decline over time but it is the same for each subject (i.e. the same constant, linear or quadratic trend for each subject), the time-constant exposure models have little bias (Table 6) as they are able to capture the across-subject variation. However, they have a larger standard error than time-varying models, as they ignore the within-subject variation over time. When the temporal exposure trend is different for each subject in the linear trend with subject-specific slopes simulations, then the baseline exposure model estimates are more biased. More complex temporal exposure structures will likely also lead to more bias in the baseline exposure models. The time-varying exposure models are even more precise than the time-constant model when each subject has a different exposure trend since the time-varying models may benefit from additional information on both within-subject and across-subject contrasts in exposure.

Among the fitted models which allow the exposure to vary over time, there are still many differences which may influence our estimates. One important question to consider is how a short-term exposure specification in the fitted model will perform if the true exposure effect is based on a long-term history of exposure. Environmental exposures such as air pollution are often hypothesized to have long-term effects, but many studies may still use short term exposure measurements if longer-term data is unavailable or if old measurements are not trusted or incompatible to recent measurements with improved technology. If the true exposure mechanisms are long term (Tables 6B and 7B), the short term model will be biased in some cases. When there are linear declines, there is not as much of a difference in the fitted short term and long term models, since the decline is a similar rate for the most recent 2 years and the previous 10 years. On the other hand, when there is a quadratic decline, the rate of decline in the past 10 years will be faster than the rate from the previous 2 years, and so the misspecified short term model is biased. When there is no exposure trend,

the random temporal variation, E_{time}, makes up all of the within-subject variation, and failure to account for a longer exposure history may also result in bias.

Although it may not be as important of an issue in practice, we could also have a situation in which we specify an exposure as long term when really the effect is short term. In Tables 6A and 7A, we see that a misspecified long term model performs similarly to a correctly specified short term model with a linear decline and different with a quadratic decline.

Finally we will consider the differences between the moving average and cumulative exposure specifications. Cumulative exposures are usually longer term, and so it makes sense to compare the cumulative and 10 year moving average models. We find that if the true exposure mechanism is a long-term moving average (Tables 6B and 7B), the average since baseline (cumulative) model performs poorly. Similarly, if the true mechanism is cumulative (Tables 6C and 7C), the short and long-term moving average models perform worse. These differences may be explained by how each mechanism gives different weight to the exposure levels of certain time periods. With a cumulative exposure mechanism, the early exposure levels have the most weight as they contribute to the exposure variable at each measurement time. The 10 year moving average, on the other hand, will give most weight to exposure levels around the beginning of the study, and less weight to exposure times that are either many years prior to the study, or near the end of the study. The short-term averages give equal weight to the time periods during the study since in this setting the exposure levels contributing to the value used for each measurement do not overlap. The short-term specification also gives no weight to a large portion of the time prior to the study.

The differences between how the cumulative and moving average exposure models each give weight to certain time periods matters most when the exposure levels are correlated with time. For example, if there is an exposure decline, all exposure measures from the cumulative model will be influenced by the high levels early on.

The moving average model will only be influenced by high levels of exposure in the early measurements during the study, and less so as the study continues. This will cause the moving average model to have an exposure variable that declines over time faster than the cumulative model. Therefore, a fitted model with a moving average exposure mechanism will associate a larger change in exposure with the effect seen in the outcome, and estimate a weaker exposure effect. This was seen in the simulations by how the moving average fitted models had negative bias if the true exposure mechanism was cumulative (Table 6C). Negative bias makes sense because the exposure effect on outcome is positive, and so a weaker effect will be closer to zero which is in the negative direction. On the other hand, if the true exposure mechanism is a moving average, the average since baseline fitted model, which is a form of cumulative model, is positively biased (Table 6B).

The average since baseline model tended to be more affected by misspecification than the moving average models. There is even bias in the average since baseline model when the underlying mechanism is cumulative, because the point at which the accumulation begins is different for the fitted model (baseline) as for the true model (10 years prior to the study). However, these claim should not be accepted without further investigation. The decision between a cumulative and moving average exposure should in large part be decided by the scientific context. In many cases a moving average may be appropriate if the outcome only depends on "recent" exposure, and after a certain amount of time an exposure has little influence. On the other hand, a cumulative exposure may be more appropriate in other contexts, such as if an exposure does irreversible damage to some biological function influencing the outcome. However, as mentioned above, although an average since baseline model is a form of cumulative model, there are many options of how to specify cumulative exposure. This has nice interpretability for the exposure effect on rate of change, since the baseline outcome is dependent on exposure prior to the study and the rate

of change (an instantaneous quantity) is dependent on the cumulative exposure since baseline. These exposure levels being averaged since baseline may be averages themselves, and be short-term or long-term averages depending on the context. This sort of structure may be more realistic than a cumulative exposure since baseline where each exposure level being averaged is the exposure from a single point in time (as is done in the simulations), in which case the rate of change would not depend on the exposure prior to baseline. If we take the cumulative average of single exposure measurements, in many situations we may believe a scientifically reasonable exposure model is cumulative since some other point prior to the study, since the beginning of the study is a rather arbitrary time when considering an ongoing exposure effect. There may be difficulty in deciding when the cumulative effect starts, as some effects may plausibly start long before a study, in some cases even at birth. The available data may also limit this decision and make a "starting point" somewhat arbitrary from a scientific perspective. In reality, an exposure effect may be a combination of a cumulative and moving average or have a more complex framework. Further research on exposure mechanisms and specification is required to have a more complete understanding, and this is a topic which is very important for proper specification of longitudinal analysis models.

A. Short-Term Underlying Exposure Mechanism (2 Year Moving Average)

Bias

	Exposure Framework of Fitted Model				
Exposure Trend	Baseline	6mth Ave.	2yr Ave.	10yr Ave.	Ave. Since Baseline
Constant	-0.33	-3.50	-0.04	0.32	1.11
Linear	-0.33	-1.04	0.01	0.03	4.44
Quadratic	-0.33	-2.62	-0.00	-2.30	1.89
Linear (Varying Slopes)	-2.28	-0.66	0.00	0.02	4.31

B. Long-Term Underlying Exposure Mechanism (10 Year Moving Average)

Bias

	Exposure Framework of Fitted Model				
Exposure Trend	Baseline	6mth Ave.	2yr Ave.	10yr Ave.	Ave. Since Baseline
Constant	-0.31	-4.54	-3.38	-0.09	-0.74
Linear	-0.31	-1.35	-0.38	0.01	4.16
Quadratic	-0.31	-2.38	0.93	0.01	5.61
Linear (Varying Slopes)	-1.45	-0.84	-0.21	0.01	4.22

C. Cumulative Underlying Exposure (Ave. Since 10 Years Prior to Study)

Bias

	Exposure Framework of Fitted Model				
Exposure Trend	Baseline	6mth Ave.	2yr Ave.	10yr Ave.	Ave. Since Baseline
Constant	-0.30	-4.59	-3.53	-0.61	-0.93
Linear	-0.30	-3.11	-2.61	-2.41	-0.11
Quadratic	-0.30	-3.45	-1.50	-2.13	1.52
Linear (Varying Slopes)	-0.41	-2.88	-2.56	-2.50	-0.23

Table 6: Bias of the exposure effect on outcome rate of change, $\hat{\beta}_1$, from fitted models with different specifications of the exposure variable, under several frameworks for the true fitted model. Note that the parameter being estimated is $\beta_1 = 5$. In **A**, the outcome is affected by the average exposure levels in the two years prior to the measurement time. In **B**, the outcome is affected by the average exposure levels in the ten years prior to the measurement time. In **C**, the outcome is affected by the average exposure levels between the measurement time and 10 years prior to the beginning of the study (a cumulative exposure mechanism). In each case, the population average exposure trend is either constant ($E_{trend} = 0$), a linear decline (common to all subjects), a quadratic decline (common to all subjects), or a linear decline (with subject-specific rates of decline). There are five exposure specifications in the fitted models: (1) a time-constant exposure (each subject has one exposure value which is the average of the 2 years prior to baseline), (2) a short-term 6 month moving average, (3) a short-term 2 year moving average, (4) a long-term 10 year moving average, and (5) an average since baseline (with the baseline exposure being the average of the two years prior to baseline).

A. Short-Term Underlying Exposure Mechanism (2 Year Moving Average)

Standard Error

	Exposure Framework of Fitted Model				
Exposure Trend	Baseline	6mth Ave.	2yr Ave.	10yr Ave.	Ave. Since Baseline
Constant	2.06	0.61	1.09	1.93	1.88
Linear	2.06	0.33	0.37	0.39	0.74
Quadratic	2.06	0.55	0.75	0.50	1.19
Linear (Varying Slopes)	2.04	0.27	0.29	0.32	0.55

B. Long-Term Underlying Exposure Mechanism (10 Year Moving Average)

Standard Error

	Exposure Framework of Fitted Model				
Exposure Trend	Baseline	6mth Ave.	2yr Ave.	10yr Ave.	Ave. Since Baseline
Constant	2.06	0.61	1.09	1.93	1.88
Linear	2.06	0.34	0.38	0.39	0.73
Quadratic	2.06	0.55	0.75	0.50	1.19
Linear (Varying Slopes)	2.03	0.27	0.29	0.32	0.55

C. Cumulative Underlying Exposure (Ave. Since 10 Years Prior to Study)

Standard Error

	Exposure Framework of Fitted Model				
Exposure Trend	Baseline	6mth Ave.	2yr Ave.	10yr Ave.	Ave. Since Baseline
Constant	2.06	0.61	1.09	1.93	1.88
Linear	2.06	0.33	0.37	0.39	0.73
Quadratic	2.06	0.55	0.75	0.50	1.19
Linear (Varying Slopes)	2.02	0.27	0.29	0.32	0.55

Table 7: Average standard error of the exposure effect on outcome rate of change, $\hat{\beta}_1$, from fitted models with different specifications of the exposure variable, under several frameworks for the true fitted model. Note that the parameter being estimated is $\beta_1 = 5$. In **A**, the outcome is affected by the average exposure levels in the two years prior to the measurement time. In **B**, the outcome is affected by the average exposure levels in the ten years prior to the measurement time. In **C**, the outcome is affected by the average exposure levels between the measurement time and 10 years prior to the beginning of the study (a cumulative exposure mechanism). In each case, the population average exposure trend is either constant ($E_{trend} = 0$), a linear decline (common to all subjects), a quadratic decline (common to all subjects), or a linear decline (with subject-specific rates of decline). There are five exposure specifications in the fitted models: (1) a time-constant exposure (each subject has one exposure value which is the average of the 2 years prior to baseline), (2) a short-term 6 month moving average, (3) a short-term 2 year moving average, (4) a long-term 10 year moving average, and (5) an average since baseline (with the baseline exposure being the average of the two years prior to baseline).

C.11 Correlation of Estimates and Parameterization of the Cross-Sectional Term

We further explore the relationship of the cross-sectional and longitudinal exposure effect estimates from the modeled baseline model (i.e. $\hat{\alpha}_1$ and $\hat{\beta}_1$). We consider how the time corresponding to the cross-sectional term may influence the exposure estimates. In practice, the only parameterization that makes sense is using the baseline ($t = 0$) as the time corresponding to the cross-sectional term. This is for several reasons. The cross-sectional estimate will be estimating a different quantity if we have the cross-sectional term corresponding to a time after baseline, and this estimate will be less interpretable. Often studies will also have varied follow-up timing across subjects, and so the only common time that all subjects share is at baseline ($t = 0$). We also run into problems if the exposure is time-varying, and we parameterized the cross-sectional term to correspond to a time other than baseline, since it could model an outcome at an earlier time being based on covariate information from a future visit.

Setting aside the fact that we would never use a parameterization other than the cross-sectional term corresponding to baseline, we consider the influence of time parameterization to better understand the relationship of the exposure effect estimates. We set the exposure to be time-constant in order to avoid issues of basing a model on future data. Since our simulations have identical follow-up timing for each patient, the main issue left with changing the cross-sectional term's time parameterization is the change in estimate interpretation.

In Figure 29, we provide scatterplots of $\hat{\alpha}_1$ and $\hat{\beta}_1$ when the cross-sectional term corresponds to baseline, the 2nd measurement time, 3rd measurement time, and 4th (final) measurement time. We see that the longitudinal effect remains the same, but the cross-sectional term is shifted. The shift in the cross-sectional term is equal to the product of longitudinal estimate, and the time since baseline that the cross-

sectional term corresponds to. For example, when we use the 2nd measurement time parameterization (1.67 years after baseline), then $\hat{\alpha}_1^{t=1.67} = \hat{\alpha}^{t=0} + 1.67(\hat{\beta}_1^{t=0})$ and $\hat{\beta}_1^{t=1.67} = \hat{\beta}_1^{t=0}$. We would expect the cross-sectional estimate to change since in the new parameterization we are no longer estimating the baseline effect, α_1, but the effect on outcome at measurement time two, which is $\alpha_1 + 1.67\beta_1$. In Figure 29 we also see that this change in the cross-sectional estimate results in stronger correlated exposure effect estimates.

Although these simulations are an artificial scenario to develop understanding of the properties of the two effect estimates, the results further suggest that the cross-sectional component is more susceptible to being influenced by study design and analysis choices. For this reason as well as the reasons mentioned in the discussion of the main document, we recommend focusing on the modeled baseline model's rate of change estimate of interest, $\hat{\beta}_1$, and treating the cross-sectional effect, $\hat{\alpha}_1$, primarily as a tool for better estimating this longitudinal effect.

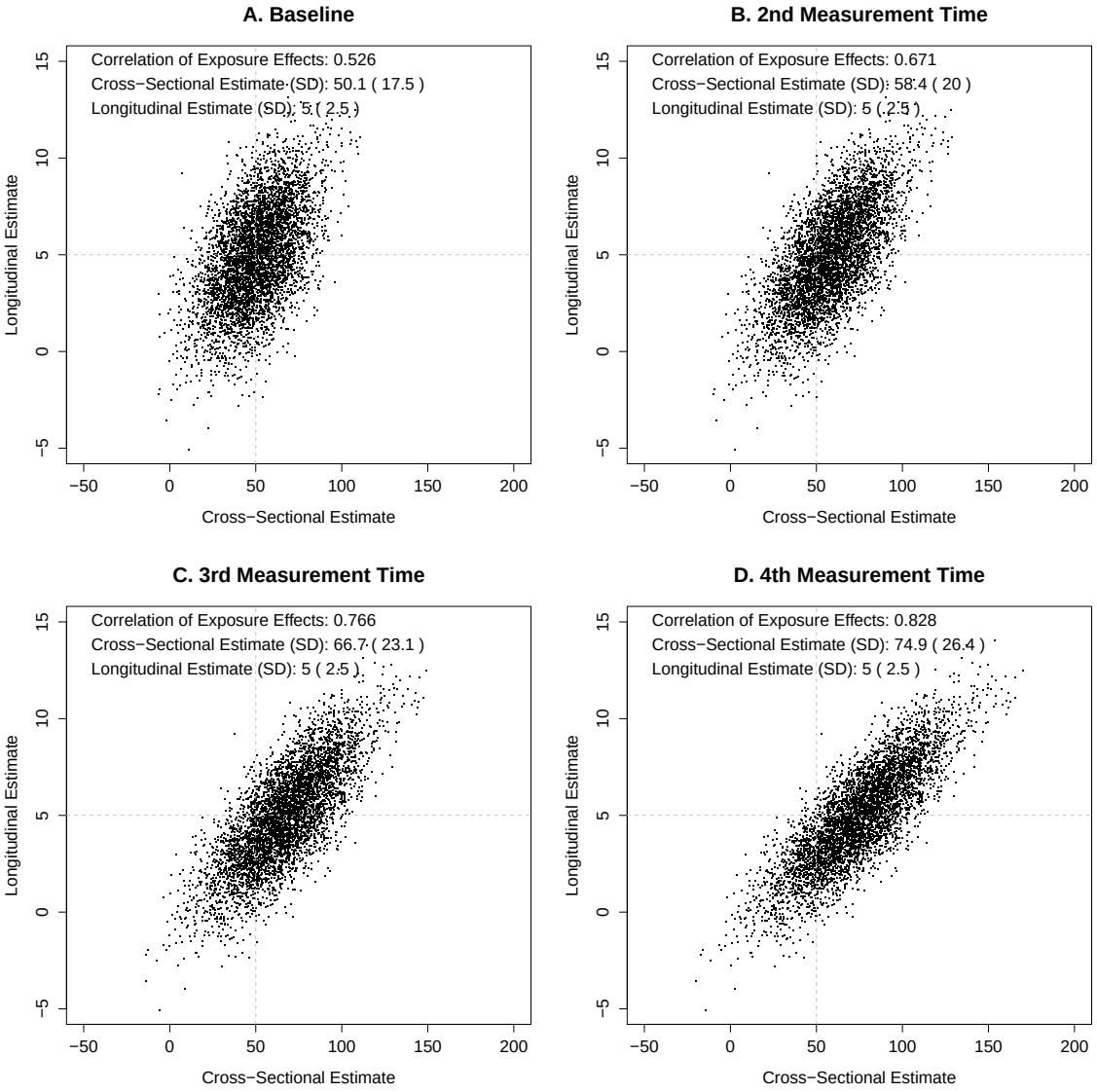

Figure 29: Modeled baseline mixed model exposure effect estimates under different parameterizations of the time corresponding to the cross-sectional term. **A** plots estimates from the usual modeled baseline mixed model where the cross-sectional term corresponds to the baseline ($t = 0$). **B**, **C** and **D** give estimates when the cross-sectional term corresponds to the 2nd, 3rd and 4th (last) follow-up visit times respectively. The averages and standard deviations of the cross-sectional and longitudinal exposure effect estimates are provided in each plot, along with the correlation of the two estimates. The horizontal and vertical dashed lines show the exposure effect parameter values from the outcome-generating model ($\alpha_1 = 50$ and $\beta_1 = 5$). Note that when a different time than baseline is specified for the cross-sectional effect, the cross-sectional effect estimate is no longer estimating $\alpha_1 = 50$.

D Analysis of MESA Air Study

In this section we provide additional details on the analysis of the MESA Air study presented in Section 4.

D.1 Adjustment Covariates

We adjust for the same set of covariates as in Keller et al. [2017], which is very similar to the set of covariates adjusted for in Kaufman et al. [2016]. We adjust for variables in the same way, which is described in detail in the Supplement of Kaufman et al. [2016].

In both the modeled baseline model and repeated scaled change model, we adjust for age, sex, race/ethnicity, income, employment outside of the home, smoking status, packyears, physical activity, adiposity measurements, total cholesterol, high density lipoprotein (HDL), triglycerides, statin use, neighborhood socio-economic status (SES) index and education. These covariates are adjusted for in both the cross-sectional and longitudinal components of the modeled baseline model. Age, income, packyears, total cholesterol, HDL, triglycerides, neighborhood SES index were adjusted for as continuous variables. Smoking status is categorized as: Never, no second-hand smoke (SHS) exposure; Never, any SHS exposure; Former, no SHS exposure; Former, SHS exposure; and Current. Physical activity was included by indicators for quartile of intentional exercise minutes. Five adiposity measurements were used: weight, 1/height, 1/height^2, waist circumference, and 1/(hip circumference). Education was categorized as: less than high school; high school; some college or technical school; bachelor's degree or more.

With the modeled baseline model, we adjust for CAC scanner type in the transient component of the model. We fit a repeated scaled change model that is unadjusted for scanner type, as well as a repeated scaled change model in which the outcome was

Exam Number	Scanner Model 1	Scanner Model 2
1	0	894
2	0	391
3	158	215
4	135	0
5	366	0

Table 8: Distribution of CAC scanner types used at each exam. Scanner model 1 is the Aquilion model and scanner model 2 is the volume zoom model.

pre-adjusted by the modeled baseline scanner type effect estimates.

D.2 Descriptives

A table of descriptives similar to that in Kaufman et al. [2016] is provided in Table 9. There are small differences between Table 9 and the Baltimore column of the table in Kaufman et al. [2016] since a slightly different set of subjects were used in the analyses. The set of individuals included in our analyses matches up with the cohort analyzed in Keller et al. [2017].

In the Baltimore subset of the MESA Air study, subjects have an average of 2.2 visits (range of 0 to 4) and an average of 4.8 years of follow-up (range of 0 to 11.2).

We also provide descriptives of CAC scanner type in Table 8. We see that one scanner type was used initially, and then phased out for a different model for later visits.

Year 2000 PM2.5 (μg/m³)	15.9 (0.8)
n	
Baseline	894
Follow-up	837
Coronary Artery Calcium	
Baseline (Agatston Score)	187 (468)
Progression (Agatstons per year)	22 (60)
Demographics	
Age	63 (10)
Male	438 (49%)
Ethnicity	
White	443 (49%)
Black	451 (50%)
Education Level	
Less than High School	96 (11%)
High School	173 (19%)
Some College or Technical	267 (30%)
University or Graduate	358 (40%)
Smoking Status	
Never Smoker	379 (42%)
Former Smoker	380 (42%)
Present Smoker	135 (15%)
General Health Characteristics	
Body mass index (kg/m²)	29 (6)
Systolic blood pressure (mm Hg)	128 (21)
HDL cholesterol (mmol/L))	1.33 (0.39)
Total cholesterol (mmol/L))	4.98 (0.94)
LDL cholesterol (mmol/L))	3.04 (0.82)
Hypertension Medication	401 (45%)
Statin Use	161 (18%)

Table 9: Participant characteristics at baseline and unadjusted longitudinal outcomes, for participants included in the analysis of coronary artery calcium. Data are mean (SD) or n (%). All percentages are calculated using the baseline as the denominator. PM2.5= fine particulate matter less than 2.5 μm in diameter (average of 2-week predictions from January, 2000, to December, 2010, generated at each participant's baseline address. HDL=high density lipoprotein. LDL=low density lipoprotein. Progression of coronary artery calcium is based on individual slopes.

E Examples of R Code

In this section we provide R code used for the simulations studies (Supplement Section E.1) and analysis of the MESA Air Study (Supplement Section E.2).

E.1 Simulation Code

We provide the code for the simulations presented in the main paper and Supplement (summarized in Table 5 in Supplement Section C).

E.1.1 Primary Simulation Code

The following function code was the main code used to produce the modeled baseline, measured baseline, and repeated scaled change model estimates. This is the code used for all simulations other than the exposure specification simulations and cross-sectional parameterization simulations. This function produces one iteration, which was replicated 1000 times for specified inputs. This process was repeated varying the inputs in order to address the desired scientific question.

```
comparison.sim<-function(
n,            #number of subjects
m,            #number of measures per subject
var.resid,    #variance of error term
var.measerror, #Variance of outcome measurement error
var.ew,       #random temporal within-subject variation in exposure
var.ea,       #across-subject variation in exposure
slope.E,      #ave. yearly decrease in exposure (1 unit ~1.4 u mg/m^3 PM_2.5)
lfu,          #length of follow-up
rand.matrix,  #random effects variance/covariance matrix (input a 2x2 matrix)
fixed.effects, #vector of (intercept, E baseline, time, E rate of change)
```

```
X.effects.cs , #cross sectional effect of additional covariate X
X.effects.l , #longitudinal effect of additional covariate X
X.cor         #correlation of additional covariate X and exposure E
){
require(lme4)
require(MASS)

# ------------------------------------------------------------ #
# Generate Exposure and  Outcome #
# ------------------------------------------------------------ #

#subject ID
sid<-rep(1:n,each = m)

#time since baseline
t<-rep( seq(from = 0,to = lfu,by = lfu/(m-1)) , times = n)

#Across-subject variation in exposure
E.a<-rep(rnorm(n,0,sqrt(var.ea)),each=m)

#Subject-specific exposure temporal trend
E.w<-rnorm(n*m,0,sqrt(var.ew))

#Exposure variable
E<-E.a+E.w+slope.E*t
E.b<-rep(E[t==0],each=m)
E.fu<-ifelse(t==0,0,E)
```

```
#record empirical variance of exposure
var.E.emp<-var(E)

#random effects
random.effects<-matrix( rep( mvrnorm(n,c(0,0),rand.matrix),each=m), ncol=2)

#Outcome residual error
epsilon<-rnorm(n*m,0,sqrt(var.resid) )

#Define additional covariate (with correlation X.cor with exposure)
var.E.true<-slope.E^2*var(t)+var.ew+var.ea
X<-rnorm(n*m,E*X.cor/sqrt(var.E.true),sqrt(1-X.cor^2))
X.b<-rep(X[t==0],each=m)
X.fu<-ifelse(t==0,0,X)

#Generate outcome
Y<-(fixed.effects[1]+fixed.effects[2]*E.b+X.b*X.effects.cs+random.effects[,1]) +
(fixed.effects[3]+fixed.effects[4]*E.fu+X.fu*X.effects.l+random.effects[,2] )*t +
 epsilon

#record empirical correlation of exposure and confounder
cor.E.cac<-cor(Y[t==0],E[t==0])

#measured outcome
Y.m<-Y+rnorm(n*m,0,sqrt(var.measerror))

# ------------------------------------------------------------ #
```

```
# Fit Linear Mixed Models #
# ------------------------------------------------------------ #

#Fit Modeled Baseline Model
mod<-lmer(Y.m~E.b+t+E.fu:t+(1+t|sid))
b1<-summary(mod)$coef[4,1]
b1.se<-summary(mod)$coef[4,2]
alpha1<-summary(mod)$coef[2,1]
alpha1.se<-summary(mod)$coef[2,2]

#Define Follow-up data for Measured Baseline and Repeated Scaled Change Models
Y.m.b<-rep(Y.m[t==0],each=(m))
Y.m.b.new<-rep(Y.m[t==0],each=(m-1))
Y.m.new<-Y.m[t!=0]
t.new<-t[t!=0]
E.new<-E[t!=0]
sid.new<-sid[t!=0]
E.fu.new<-E.fu[t!=0]

#Fit Measured Baseline Model
mod.meas<-lmer(Y.m.new~Y.m.b.new+t.new+E.fu.new:t.new+(1+t.new|sid.new))
#mod.meas<-lmer(Y.m~Y.m.b+E.b+t+E.fu:t+(1+t|sid))
meas.b1<-summary(mod.meas)$coef[4,1]
meas.b1.se<-summary(mod.meas)$coef[4,2]

#Fit Repeated Scaled Change Model
Y.sc<-(Y.m.new-Y.m.b.new)/t.new
```

```
mod.sc<-lmer(Y.sc~E.new+(1|sid.new))
sc.b1<-summary(mod.sc)$coef[2,1]
sc.b1.se<-summary(mod.sc)$coef[2,2]

# ------------------------------------------------------------ #
# Return function inputs and outputs #
# ------------------------------------------------------------ #

return(  c(n=n,m=m,var.resid=var.resid,var.ew=var.ew,var.ea=var.ea,
slope.E=slope.E,var.E.emp=var.E.emp,var.a=rand.matrix[1,1],
var.b=rand.matrix[2,2],cov.ab=rand.matrix[1,2],lfu=lfu,b1=b1,b1.se=b1.se,
alpha1=alpha1,alpha1.se=alpha1.se,meas.b1=meas.b1,meas.b1.se=meas.b1.se,
sc.b1=sc.b1,sc.b1.se=sc.b1.se,X.effects.cs=X.effects.cs,
X.effects.l=X.effects.l,X.cor=X.cor,cor.E.cac=cor.E.cac,
var.measerror=var.measerror)  )
}
```

E.1.2 Exposure Specification Code

For the exposure specification simulations (Supplement Section C.10), we use similar code to the primary code provided in Supplement Section E.1.1. However, there are several changes to the study design and exposure variable framework. The length of follow-up is intended to always be 10 years, as the exposure is simulated from 10 years prior to baseline until the end of a 10 year study (20 years total). We add inputs for "E.type" (the exposure time-scale) and "trend.type" (the average temporal exposure trend). Note that there are baseline models and average since baseline models that have the baseline being 2 year averages and 10 year averages. In the Supplement Section C.10 we only consider the models that use a 2 year average, since the model performance was fairly similar, but here we provide code to do a longer-term baseline as well.

```
exp.spec.sim<-function(
n,            #number of subjects
m,            #number of measures per subject
var.resid,    #variance of error term
var.measerror, #Variance of outcome measurement error
var.ew,       #random temporal within-subject variation in exposure
var.ea,       #across-subject variation in exposure
slope.E,      #ave. yearly decrease in exposure (1 unit= ~1.4 u mg/m^3 PM_2.5)
lfu,          #length of follow-up
rand.matrix,  #random effects variance/covariance matrix (input a 2x2 matrix)
fixed.effects, #vector of (intercept, E baseline, time, E rate of change)
X.effects.cs , #cross sectional effect of additional covariate X
X.effects.l , #longitudinal effect of additional covariate X
X.cor,         #correlation of additional covariate X and exposure E
trend.type,    # 1=constant, 2=linear, 3=quadratic, 4=linear (subj.-spec. slopes)
```

```
E.type        #E time-scale (1=6mth ave ,2=10yr ave, 3=2yr ave, 4=cum. since t=0)
){
require(lme4)
require(MASS)

# ------------------------------------------------------------ #
# Generate Exposure and  Outcome #
# ------------------------------------------------------------ #

#subject ID
sid<-rep(1:n,each = m)

#time since baseline
t<-rep( seq(from = 0,to = lfu,by = lfu/(m-1)) , times = n)

#Across-subject variation in exposure
E.a<-rep(rnorm(n,0,sqrt(var.ea)),each=40)

#Subject-specific exposure temporal trend
E.w<-rnorm(n*40,0,sqrt(var.ew))

##############################################
#Population Exposure Trend

if (trend.type==1){
#constant
E.trend<-rep(0,40)
```

```
E.mat<-matrix(rep(E.trend,times = n)+E.w+E.a,ncol=n)
}else if (trend.type==2) {
#linear decline (Decreases 0.3 ug/m^3 per year)
E.trend<-seq(-1*slope.E*10,slope.E*10,length.out = 40)
E.mat<-matrix(rep(E.trend,times = n)+E.w+E.a,ncol=n)
}else if (trend.type==3){
#quadratic decline (Note: same total decline as linear,
#but 75% of decline comes in first 10 years, 25% is last 10 years)
E.trend<-((seq(0:39)-21)/2)^2*-.05*slope.E-((seq(0:39)-21)/2)*-slope.E -
mean(((seq(0:39)-21)/2)^2*-.05*slope.E-((seq(0:39)-21)/2)*-slope.E)
E.mat<-matrix(rep(E.trend,times = n)+E.w+E.a,ncol=n)
}else {
#Linear decline with subject specific slopes
#Note: slopes vary uniformly from 0 to -.6 (twice the observed ave. decline)
E.trend<-seq(-1*slope.E*10,slope.E*10,length.out = 40)*rep(runif(n,0,2),each=40)
E.mat<-matrix(E.trend+E.w+E.a,ncol=n)
}



##############################################
#Exposure  visit time (FOR TRUE MODEL)

if (E.type==1){
#"6month average"
E<-c( (E.mat)[c(20,24,28,32,38,40),]  )
}else if (E.type==2) {
#10yr average
```

```
E<-c(  rbind(   apply( E.mat[1:20,] ,2, mean ) ,
apply( E.mat[5:24,] ,2, mean ) ,
apply( E.mat[9:28,] ,2, mean ) ,
apply( E.mat[13:32,] ,2, mean ),
apply( E.mat[17:36,] ,2, mean ),
apply( E.mat[21:40,] ,2, mean )   )      )
}else if (E.type==3) {
#2yr average
E<-c(  rbind(   apply( E.mat[17:20,] ,2, mean ) ,
apply( E.mat[21:24,] ,2, mean ) ,
apply( E.mat[25:28,] ,2, mean ) ,
apply( E.mat[29:32,] ,2, mean ) ,
apply( E.mat[33:36,] ,2, mean ) ,
apply( E.mat[37:40,] ,2, mean )   )      )
}else
#Cumulative
E<-c(  rbind(   apply( E.mat[1:20,] ,2, mean ) ,
apply( E.mat[1:24,] ,2, mean ) ,
apply( E.mat[1:28,] ,2, mean ) ,
apply( E.mat[1:32,] ,2, mean ) ,
apply( E.mat[1:36,] ,2, mean ) ,
apply( E.mat[1:40,] ,2, mean )    )      )



#Baseline and follow-up exposure variables
E.b<-rep(E[t==0],each=m)
E.fu<-ifelse(t==0,0,E)
```

```
#random effects
random.effects<-matrix( rep( mvrnorm(n,c(0,0),rand.matrix),each=m), ncol=2)

#Outcome residual error
epsilon<-rnorm(n*m,0,sqrt(var.resid) )

#Define additional covariate (with correlation X.cor with exposure)
var.E.true<-slope.E^2*var(t)+var.ew+var.ea
X<-rnorm(n*m,E*X.cor/sqrt(var.E.true),sqrt(1-X.cor^2))
X.b<-rep(X[t==0],each=m)
X.fu<-ifelse(t==0,0,X)

#Generate outcome
Y<-(fixed.effects[1]+fixed.effects[2]*E.b+random.effects[,1])+
(fixed.effects[3]+fixed.effects[4]*E.fu+random.effects[,2] )*t + epsilon

#record empirical correlation of exposure and confounder
cor.E.cac<-cor(Y[t==0],E[t==0])

#measured outcome
Y.m<-Y+rnorm(n*m,0,sqrt(var.measerror))

# ------------------------------------------------------------ #
# Fit Mixed Models (only Modeled Baseline) #
# ------------------------------------------------------------ #
```

```
#Baseline Model (2yr average)
E.b2<- rep(apply( E.mat[17:20,] ,2, mean ) ,each=m)
mod.b2<-lmer(Y~E.b2*t+(1+t|sid))
b1.b2<-summary(mod.b2)$coef[4,1]
b1.se.b2<-summary(mod.b2)$coef[4,2]
alpha1.b2<-summary(mod.b2)$coef[2,1]
alpha1.se.b2<-summary(mod.b2)$coef[2,2]

#Baseline Model (10yr average)
E.b10<- rep(apply( E.mat[1:20,] ,2, mean ) ,each=m)
mod.b10<-lmer(Y~E.b10*t+(1+t|sid))
b1.b10<-summary(mod.b10)$coef[4,1]
b1.se.b10<-summary(mod.b10)$coef[4,2]
alpha1.b10<-summary(mod.b10)$coef[2,1]
alpha1.se.b10<-summary(mod.b10)$coef[2,2]

#"Real-time" model (6 month Average)
E.rt<-c( (E.mat)[c(20,24,28,32,36,40),]  )
E.rt.b<-rep(E.rt[t==0],each=m)
mod.rt<-lmer(Y~E.rt.b+E.rt:t+t+(1+t|sid))
b1.rt<-summary(mod.rt)$coef[4,1]
b1.se.rt<-summary(mod.rt)$coef[4,2]
alpha1.rt<-summary(mod.rt)$coef[2,1]
alpha1.se.rt<-summary(mod.rt)$coef[2,2]

#10yr Average
E.10<-c(  rbind(   apply( E.mat[1:20,] ,2, mean ) ,
```

```
apply( E.mat[5:24,] ,2, mean ) ,
apply( E.mat[9:28,] ,2, mean ) ,
apply( E.mat[13:32,] ,2, mean ),
apply( E.mat[17:36,] ,2, mean ),
apply( E.mat[21:40,] ,2, mean ) )      )
E.10.b<-rep(E.10[t==0],each=m)
mod.10<-lmer(Y~E.10.b+E.10:t+t+(1+t|sid))
b1.10<-summary(mod.10)$coef[4,1]
b1.se.10<-summary(mod.10)$coef[4,2]
alpha1.10<-summary(mod.10)$coef[2,1]
alpha1.se.10<-summary(mod.10)$coef[2,2]

#2yr Average
E.2<-c(  rbind(   apply( E.mat[17:20,] ,2, mean ) ,
apply( E.mat[21:24,] ,2, mean ) ,
apply( E.mat[25:28,] ,2, mean ) ,
apply( E.mat[29:32,] ,2, mean ) ,
apply( E.mat[33:36,] ,2, mean ) ,
apply( E.mat[37:40,] ,2, mean ) )      )
E.2.b<-rep(E.2[t==0],each=m)
mod.2<-lmer(Y~E.2.b+E.2:t+t+(1+t|sid))
b1.2<-summary(mod.2)$coef[4,1]
b1.se.2<-summary(mod.2)$coef[4,2]
alpha1.2<-summary(mod.2)$coef[2,1]
alpha1.se.2<-summary(mod.2)$coef[2,2]

#Average Since Baseline (baseline is 2yr ave)
```

```
E.asb2<-c(  rbind(   apply( E.mat[17:20,] ,2, mean ) ,
apply( E.mat[21:24,] ,2, mean ) ,
apply( E.mat[21:28,] ,2, mean ) ,
apply( E.mat[21:32,] ,2, mean ) ,
apply( E.mat[21:36,] ,2, mean ) ,
apply( E.mat[21:40,] ,2, mean ) )      )
E.asb2.b<-rep(E.asb2[t==0],each=m)
mod.asb2<-lmer(Y~E.asb2.b+E.asb2:t+t+(1+t|sid))
b1.asb2<-summary(mod.asb2)$coef[4,1]
b1.se.asb2<-summary(mod.asb2)$coef[4,2]
alpha1.asb2<-summary(mod.asb2)$coef[2,1]
alpha1.se.asb2<-summary(mod.asb2)$coef[2,2]

#Average Since Baseline (baseline is 10yr ave)
E.asb10<-c(  rbind(   apply( E.mat[1:20,] ,2, mean )  ,
apply( E.mat[21:24,] ,2, mean ) ,
apply( E.mat[21:28,] ,2, mean ) ,
apply( E.mat[21:32,] ,2, mean ) ,
apply( E.mat[21:36,] ,2, mean ) ,
apply( E.mat[21:40,] ,2, mean ) )      )
E.asb10.b<-rep(E.asb10[t==0],each=m)
mod.asb10<-lmer(Y~E.asb10.b+E.asb10:t+t+(1+t|sid))
b1.asb10<-summary(mod.asb10)$coef[4,1]
b1.se.asb10<-summary(mod.asb10)$coef[4,2]
alpha1.asb10<-summary(mod.asb10)$coef[2,1]
alpha1.se.asb10<-summary(mod.asb10)$coef[2,2]
```

```
# ---------------------------------------------------------- #
# Return function inputs and outputs #
# ---------------------------------------------------------- #

return(  c(n=n,m=m,var.resid=var.resid,var.ew=var.ew,var.ea=var.ea,
var.a=rand.matrix[1,1],var.b=rand.matrix[2,2],cov.ab=rand.matrix[1,2],
lfu=lfu,b1.b2=b1.b2,b1.se.b2=b1.se.b2,alpha1.b2=alpha1.b2,
alpha1.se.b2=alpha1.se.b2,b1.b10=b1.b10,b1.se.b10=b1.se.b10,
alpha1.b10=alpha1.b10,alpha1.se.b10=alpha1.se.b10,b1.rt=b1.rt,
b1.se.rt=b1.se.rt,alpha1.rt=alpha1.rt,alpha1.se.rt=alpha1.se.rt,b1.10=b1.10,
b1.se.10=b1.se.10,alpha1.10=alpha1.10,alpha1.se.10=alpha1.se.10,b1.2=b1.2,
b1.se.2=b1.se.2,alpha1.2=alpha1.2,alpha1.se.2=alpha1.se.2,b1.asb2=b1.asb2,
b1.se.asb2=b1.se.asb2,alpha1.asb2=alpha1.asb2,alpha1.se.asb2=alpha1.se.asb2,
b1.asb10=b1.asb10,b1.se.asb10=b1.se.asb10,alpha1.asb10=alpha1.asb10,
alpha1.se.asb10=alpha1.se.asb10,trend.type=trend.type,E.type=E.type)  )
}
```

E.1.3 Cross-Sectional Parameterization Code

The following code was used for the simulation presented in Supplement Section C.11 which considered the parameterization of the cross-sectional term of the modeled baseline model. This code was also used to look at the baseline vs. time-varying exposure specification considered in Figure 3 of Section 3.2.2. The only difference between this code and the primary simulation code given in Supplement Section E.1.1 is the fitted models and what is returned from the function, so we only present those components to avoid repetition.

```
# ------------------------------------------------------------ #
# Fit Linear Mixed Models #
# ------------------------------------------------------------ #

#################################
#Model Parameterization 1: Usual Parameterization for Modeled Baseline

#Time Varying Exposure
mod<-lmer(Y.m~E.b+t+E.fu:t+(1+t|sid))
b1<-summary(mod)$coef[4,1]
b1.se<-summary(mod)$coef[4,2]
alpha1<-summary(mod)$coef[2,1]
alpha1.se<-summary(mod)$coef[2,2]



#Time Constant (baseline) Exposure
mod.base<-lmer(Y.m~E.b+t+E.b:t+(1+t|sid))
base.b1<-summary(mod.base)$coef[4,1]
base.b1.se<-summary(mod.base)$coef[4,2]
```

```
base.alpha1<-summary(mod.base)$coef[2,1]
base.alpha1.se<-summary(mod.base)$coef[2,2]



##################################
#Model 2: Cross-Sectional Component Specified from 2nd Visit Time

meastime2<-t[2]
E.mt2<-rep(E[t==meastime2],each=m)
E.fu.mt2<-ifelse(t==meastime2,0,E)
t.mt2<-t-meastime2
mod.base.mt2<-lmer(Y.m~E.mt2+t.mt2+E.mt2:t.mt2+(1+t.mt2|sid))
base.b1.mt2<-summary(mod.base.mt2)$coef[4,1]
base.b1.se.mt2<-summary(mod.base.mt2)$coef[4,2]
base.alpha1.mt2<-summary(mod.base.mt2)$coef[2,1]
base.alpha1.se.mt2<-summary(mod.base.mt2)$coef[2,2]

##################################
#Model 3: Cross-Sectional Component Specified from 3rd Visit Time

meastime3<-t[3]
E.mt3<-rep(E[t==meastime3],each=m)
E.fu.mt3<-ifelse(t==meastime3,0,E)
t.mt3<-t-meastime3
mod.base.mt3<-lmer(Y.m~E.mt3+t.mt3+E.mt3:t.mt3+(1+t.mt3|sid))
base.b1.mt3<-summary(mod.base.mt3)$coef[4,1]
base.b1.se.mt3<-summary(mod.base.mt3)$coef[4,2]
```

```
base.alpha1.mt3<-summary(mod.base.mt3)$coef[2,1]
base.alpha1.se.mt3<-summary(mod.base.mt3)$coef[2,2]

################################
#Model 4: Cross-Sectional Component Specified from 4th (last) Visit Time

meastime4<-t[4]
E.mt4<-rep(E[t==meastime4],each=m)
E.fu.mt4<-ifelse(t==meastime4,0,E)
t.mt4<-t-meastime4
mod.base.mt4<-lmer(Y.m~E.mt4+t.mt4+E.mt4:t.mt4+(1+t.mt4|sid))
base.b1.mt4<-summary(mod.base.mt4)$coef[4,1]
base.b1.se.mt4<-summary(mod.base.mt4)$coef[4,2]
base.alpha1.mt4<-summary(mod.base.mt4)$coef[2,1]
base.alpha1.se.mt4<-summary(mod.base.mt4)$coef[2,2]

# ------------------------------------------------------------ #
# Return function inputs and outputs #
# ------------------------------------------------------------ #

return(c(n=n,m=m,var.resid=var.resid,var.ew=var.ew,var.ea=var.ea,
slope.E=slope.E,var.E.emp=var.E.emp,var.a=rand.matrix[1,1],
var.b=rand.matrix[2,2],cov.ab=rand.matrix[1,2],lfu=lfu,b1=b1,b1.se=b1.se,
alpha1=alpha1,alpha1.se=alpha1.se,base.b1=base.b1,base.b1.se=base.b1.se,
X.effects.cs=X.effects.cs,X.effects.l=X.effects.l,X.cor=X.cor,
cor.E.cac=cor.E.cac,var.measerror=var.measerror,base.alpha1=base.alpha1,
base.alpha1.se=base.alpha1.se,base.b1.mt2=base.b1.mt2,
```

```
base.b1.se.mt2=base.b1.se.mt2,base.alpha1.mt2=base.alpha1.mt2,
base.alpha1.se.mt2=base.alpha1.se.mt2,base.b1.mt3=base.b1.mt3,
base.b1.se.mt3=base.b1.se.mt3,base.alpha1.mt3=base.alpha1.mt3,
base.alpha1.se.mt3=base.alpha1.se.mt3,base.b1.mt4=base.b1.mt4,
base.b1.se.mt4=base.b1.se.mt4,base.alpha1.mt4=base.alpha1.mt4,
base.alpha1.se.mt4=base.alpha1.se.mt4)   )
}
```

E.2 MESA Air Analysis Code

We provide examples of R code used to fit the modeled baseline and repeated scaled change models to study the effect of particulate matter on rate of CAC progression (Results presented in Section 4.2). We also provide code for the repeated scaled change model pre-adjusted for CAC scanner type. Note that in the modeled baseline model, we use $*$ to specify a baseline and longitudinal (time interaction) when the variable is time-constant. However, for the time-varying variables such as exposure, we must have a separate term for the baseline and longitudinal term.

```
# ------------------------------------------------------------ #
#  Primary Analyses:  Fit mixed models                         #
# ------------------------------------------------------------ #

#################################
#modeled baseline

#fit model
mod.fit<-lmer(agatpc~pm_bl+ctmodel+pm_fu:fu_yr+factor(statin)*fu_yr+
age.c*fu_yr+factor(male)*fu_yr+factor(race1c)*fu_yr+factor(employ)*fu_yr+
factor(smknew)*fu_yr+pkyrs1c*fu_yr+factor(exercm_cat)*fu_yr+weight*fu_yr+
height.inv*fu_yr+height.inv2*fu_yr+waist*fu_yr+hip.inv*fu_yr+chol.c*fu_yr+
nses.c*fu_yr+income.c*fu_yr+factor(education)*fu_yr+factor(hghchol1)*fu_yr+
hdl1.c*fu_yr+trig1.c*fu_yr+(1+fu_yr|ppt_id), data = baltimore)

#################################
#repeated scaled change

#Define follow-up data
```

```
baltimore.fu<-baltimore[baltimore$fu_yr!=0,]

#Define term for outcome scaled change
baltimore.fu$cac.sc<-(baltimore.fu$agatpc-baltimore.fu$agatpc1) /
baltimore.fu$fu_yr

#Fit model
rsc.fit<-lmer(cac.sc~pm_fu+factor(statin)+age.c+factor(male)+factor(race1c)+
factor(employ)+factor(smknew)+pkyrs1c+factor(exercm_cat)+weight+height.inv+
height.inv2+waist+hip.inv+chol.c+nses.c+income.c+factor(education)+
factor(hghchol1)+hdl1.c+trig1.c+(1|ppt_id), data = baltimore.fu)

# ------------------------------------------------------------ #
# Exploratory Analysis: Pre-adjusted Repeated Scaled Change    #
# (Pre-adjust CAC by modeled baseline scanner type effect)     #
# ------------------------------------------------------------ #

#Adjust CAC by modeled baseline effect estimate
baltimore$cac.adj<-ifelse(baltimore$ctmodel=="11: Volume Zoom",
baltimore$agatpc-summary(mod.fit)$coef["ctmodel11: Volume Zoom",1],
baltimore$agatpc)

#define adjusted baseline CAC variable
#and define follow-up dataset for repeated scaled change
baltimore$cac.b.adj<-baltimore$agatpc1 -
summary(mod.fit)$coef["ctmodel11: Volume Zoom",1]
baltimore.fu2<-baltimore[baltimore$fu_yr!=0,]
```

```
baltimore.fu2$cac.sc.adj<-(baltimore.fu2$cac.adj-baltimore.fu2$cac.b.adj) /
baltimore.fu2$fu_yr

#Fit pre-adjusted repeated scaled change model
rsc.fit2<-lmer(cac.sc.adj~pm_fu+factor(statin)+age.c+factor(male)+
factor(race1c)+factor(employ)+factor(smknew)+pkyrs1c+factor(exercm_cat)+
weight+height.inv+height.inv2+waist+hip.inv+chol.c+nses.c+income.c+
factor(education)+factor(hghchol1)+hdl1.c+trig1.c+
(1|ppt_id),data = baltimore.fu2)
```