A simulation platform to quantify survival bias: an application to research on determinants of cognitive decline

Elizabeth Rose Mayeda,[1] Eric J. Tchetgen Tchetgen,[2] Melinda C. Power,[3] Jennifer Weuve,[4] Hélène Jacqmin-Gadda,[5] Jessica R. Marden,[6] Eric Vittinghoff, [1] Niels Keiding,[7] M. Maria Glymour[1]

[1] Department of Epidemiology and Biostatistics, University of California, San Francisco

[2] Departments of Biostatistics and Epidemiology, Harvard School of Public Health

[3] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health

[4] Department of Internal Medicine, Rush Institute for Healthy Aging

[5] INSERM, U897, ISPED, Universite de Bordeaux

[6] Department of Social and Behavioral Sciences, Harvard School of Public Health

[7] Section of Biostatistics, University of Copenhagen

Abbreviations: DAG=directed acyclic graph, SD=standard deviation, RMSE=root-mean-square error

ABSTRACT

Bias due to selective mortality is a potential concern in many studies and is especially relevant in cognitive aging research because cognitive impairment strongly predicts subsequent mortality. Biased estimation of the effect of an exposure on rate of cognitive decline can occur when mortality is a common effect of exposure and an unmeasured determinant of cognitive decline and similar settings. This potential is often represented as collider-stratification bias in directed acyclic graphs, but it is difficult to anticipate the magnitude of bias. We present a flexible simulation to quantify the expected bias in longitudinal studies of determinants of cognitive decline. We evaluated potential survival bias in naïve analyses under several selective survival scenarios, assuming that exposure had no effect on cognitive decline for anyone in the population. Compared with the situation with no collider bias, the magnitude of bias was higher when exposure and an unmeasured determinant of cognitive decline interacted on the hazard ratio scale to influence mortality or when both exposure and rate of cognitive decline influenced mortality. Bias was, as expected, larger in high-mortality situations. This simulation platform provides a flexible tool for evaluating biases in studies with high mortality, as is common in cognitive aging research.

Keywords: cognitive decline, dementia, selective survival, selection bias, survival bias, collider-stratification bias, simulation, truncation by death

Selective survival presents an important potential source of bias in longitudinal research with truncation by death. Survival bias is especially relevant in research on determinants of cognitive decline because cognitive decline predicts death (1-4), many exposures of interest also influence death, and death rates are high in older adult populations. Hernán *et al.* represented this challenge using Directed Acyclic Graphs, conceptualizing survival bias under the sharp null (no effect of the exposure on the outcome) as a type of collider-stratification bias that arises when survival is a common effect of the exposure and outcome process (5). For cognitive aging research, situations where cognitive outcomes directly influence survival and where a determinant of cognitive decline influences survival are both plausible.

Although in some situations researchers may be confident they have appropriately measured and accounted for the selection processes, the potential for survival bias is widely recognized in numerous settings when many determinants of survival are not known or not measured. Simulation studies provide the opportunity to systematically assess likely magnitude of bias under an array of assumptions about the data generating process, including the causal structure and effect sizes. Simulation studies may also provide useful guidance to researchers about which assumptions about the selection process matter most for their research question. We have developed a flexible simulation platform to quantify expected magnitude of bias in studies of determinants of rate of cognitive decline. Using this simulation platform, we evaluated potential survival bias when implementing naïve analyses of cognitive decline under selective survival. We consider several causal scenarios, with a binary baseline exposure and continuous cognitive function measured at repeated occasions, as is typical in cohort studies of cognitive aging. The binary exposure could represent many different exposures of interest to epidemiologists. For example, a treatment in a hypothetical randomized controlled trial, such as hormone replacement therapy or intensive glycemic control, or an exposure in a hypothetical cohort study, such as a genetic variant or smoking.

Simulating data to model cognitive aging entails incorporating complex correlation structures that prevail across successive cognitive assessments of a person and error in typical neuropsychological measurements. We therefore begin by describing our overall approach to generating cognitive data. We then describe the specific causal scenarios we evaluate and report effect estimates from naïve regression models (standard regression models that condition on a person being alive) under each scenario. Our objective was to simulate a study of cognitive aging that corresponds with assumptions of substantive researchers and captures the most important drivers of survivor bias, but omits superfluous data features in order to maintain transparency and constrain the number of input assumptions. We therefore adopted several simplifying assumptions for clarity of exposition, most notably, the sharp null (no effect of the exposure on rate of cognitive change in any person). Simple modifications to the simulation code could make it relevant to a range of research questions related to diverse outcomes, for example, performance of alternative analytic approaches in the presence of selective survival; effects of early lifecourse factors that influence survival for decades prior to study enrollment; or gender differences in dementia incidence.

METHODS

Hypothetical study

We considered a study of the effect of a binary exposure on rate of cognitive decline in adults aged ≥60 years. The hypothetical study sample was followed for up to 9 years with cognitive assessments every 1.5 years (*i.e.*, up to 7 assessments). To focus discussion on survivor bias, we assumed no other form of attrition was present and that exposure was effectively randomized at baseline (no confounding).

Generation of repeated cognitive measures and survival times

Each person in each simulated data set is assigned: exposure (from Bernoulli(0.50)); baseline age (baseline_age), defined as 60 years plus a random variable drawn from a Beta(2,4) distribution, which we scaled to allow a possible 40-year age range (see **Web Figure 1**); and an unmeasured continuous covariate, U~N(0,1), which can represent a single variable or a set of variables that influence rate of cognitive change. We adopt a growth curve framework for generating cognitive measures since it easily incorporates within person correlation, as well as other sources of variation between people, such as random slopes. For each person *i* at wave *j*, we generated a value of **cognitive function** C_{ij} following an autoregressive linear model with a random intercept (ζ_{0i}) and random slope (ζ_{1i}) drawn from a bivariate normal distribution with mean 0, variances $\sigma^2_{\zeta_0}$ and $\sigma^2_{\zeta_1}$, and covariance $\sigma_{\zeta_{01}}$ (coefficient definitions in **Box 1**). Although we use the term $time_{ij}$, the value of time at wave *j* does not vary between people in the present simulations.

$$C_{ij} = \beta_{00} + \beta_{01}exposure_i + \beta_{02}baseline\_age_i + \beta_{03}U_i + (\beta_{10} + \beta_{11}exposure_i + \beta_{12}baseline\_age_i + \beta_{13}U_i)time_{ij} + \zeta_{0i} + \zeta_{1i}time_{ij} + \varepsilon_{ij} \quad (1)$$

In this model, ε_{ij} represents unexplained variation in C_{ij}, where $\varepsilon_{ij} = \rho\varepsilon_{ij-1} + \alpha_{ij},\ \varepsilon_{ij} \sim N(0, {\sigma_\varepsilon}^2),\ \alpha_{ij} \sim N(0, (1-\rho^2){\sigma_\varepsilon}^2), and\ \alpha_{ij} \coprod \varepsilon_{ij-1}$. This structure creates an autoregressive model within a person where the variance of ε_{ij} is constant across waves (see **Web Table 1**).

Our data generating model for cognitive function allows for random intercepts and random slopes and an autoregressive within-person covariance structure. Most researchers evaluating determinants of rate of cognitive change adopt either a linear mixed effects model with random intercepts and random slopes *or* a generalized estimating equation (GEE) approach assuming an autoregressive within-person covariance structure (*i.e.*, marginalizing over random effects). Our

simulation platform can accommodate a variety of structures. For example, the data generating model for cognitive function can be specified without random effects (by setting $\sigma^2_{\zeta_0} = 0$, $\sigma^2_{\zeta_1} = 0$, and $\sigma_{\zeta_{01}} = 0$) or the data generating model can be set to correspond with a strictly random effects model (by setting $\rho = 0$).

For conceptual clarity, note that equation 1 can be written more succinctly as:

$$C_{ij} = X_i'\beta_0 + (X_i'\beta_1)time_{ij} + \zeta_{0i} + \zeta_{1i}time_{ij} + \varepsilon_{ij} \quad (2)$$

where β_0 is a vector of fixed effects for cognitive intercept and β_1 is a vector of fixed effects for cognitive slope (rate of cognitive change), and *X'* represents the vector of covariates including exposure, baseline age, and U.

After generating true values of cognitive function for each person *i* at each wave *j*, we generated measured values of cognitive function (C^*_{ij}) by adding random measurement error, $\delta_{ij} \sim N(0, \sigma^2_\delta)$, to true values of cognitive function (C_{ij}) described in equation 1:

$$C^*_{ij} = C_{ij} + \delta_{ij} \quad (3)$$

The simulation code can be modified to eliminate measurement error by setting $\sigma^2_\delta = 0$.

We generated **survival time** for each person as a function of exposure (time-constant), U (time-constant), age (time-varying), rate of cognitive change (time-varying), and level of cognitive function (time-varying). At each cognitive assessment wave *j*, a "time to death" value is generated for each person who has remained alive up to that wave. The time to death variable is generated based on the person's covariate history up to that time point as a random variable drawn from an exponential survival distribution (based on the hazard function specified in equation 4 below). If the randomly generated time to death exceeds the length of the interval between waves *j* and *j+1* (1.5 years), the person is considered alive at wave *j+1* and a new survival time is generated for the next interval conditional on history up to the start of the interval. The process is repeated until the person's survival time falls within

a given interval or the end of the study (7 waves), whichever comes first. Each person's hazard function at time *t* in the *jth* interval is defined as (coefficient definitions in **Box 1**):

$$h(t_{ij}|x) = \lambda \exp(\gamma_1 exposure_i + \gamma_2 age_{ij} + \gamma_3 U_i + \gamma_4 exposure_i * U_i + \gamma_5(slope_{ij}) + \gamma_6 C_{ij}) \quad (4)$$

In this model, $slope_{ij}$ represents person *i*'s rate of cognitive change for the interval starting at time *j* and is defined as:

$$slope_{ij} = (X_i'\beta_1 + \zeta_{1i}) + \left(\frac{\varepsilon_{ij+1} - \varepsilon_{ij}}{time_{ij+1} - time_{ij}}\right) \quad (5)$$

where$(X_i'\beta_1 + \zeta_{1i})$ equals person *i*'s average slope throughout follow up and $\left(\frac{\varepsilon_{ij+1} - \varepsilon_{ij}}{time_{ij+1} - time_{ij}}\right)$ represents person *i*'s deviation from this long-term average slope during the current time interval. This slope does not incorporate measurement error because it represents the person's true, rather than measured, rate of cognitive change. We then used the inverse cumulative hazard function transformation formula described by Bender *et al.* (6) (formula in **Web Appendix 1**) to generate each person's survival time for a given time interval at risk based on the above hazard function.

Causal scenarios guiding data generation

We carried out simulations under several causal scenarios (**Figure 1**). We chose an initial scenario (Scenario A) with no anticipated survivor bias. Although the exposure influences mortality, no correlate of cognitive decline affects mortality in this scenario, so conditioning on survival should not give rise to collider-stratification bias. We then considered scenarios under which selective survival was expected to induce collider-stratification bias. In the second scenario (Scenario B), mortality is influenced by exposure and U, an unmeasured determinant of rate of cognitive decline. Because interaction between causes is known to influence the magnitude of collider-stratification bias (5, 7, 8), we consider a situation in which exposure and U influence mortality hazard multiplicatively, with no interaction on the hazard ratio scale in determining mortality (Scenario B1) and also a situation where exposure and U interact on the hazard ratio scale in determining mortality (Scenario B2). In Scenario C1,

mortality is influenced by exposure and current rate of cognitive change, and in Scenario C2, level of cognitive function at the beginning of the time interval also influences mortality.

For each causal scenario, we simulated B=1,000 samples of N=1,500 people with the archetype parameter inputs displayed in **Table 1**. We selected "archetype" values that represented large associations between variables to assess survival bias when the associations between variables were on the higher bound of the spectrum of what we considered likely in real data. Specifically, every one-unit increase in U doubles rate of cognitive decline (β_{13}=β_{10}=-0.05). Exposure doubles the hazard of death (exp(Υ_1)=2.0). A one-unit increase in U increases the hazard of death by a factor of 5.0, either overall (Scenario B1, exp(Υ_3)=5.0) or among exposed only (Scenario B2 exp(Υ_4)=5.0). In Scenario B2 this is essentially a silencing interaction, such that if exposure is not present, U has no effect. Each 0.10-unit faster annual rate of cognitive decline increases the hazard of death by 34% (exp(Υ_5*-0.10)=1.34), and every one-unit decrease in level of cognitive function (C_{ij}) doubles the hazard of death (exp(Υ_6*-1)=2.0).

To assess bias with more moderate effect sizes, we examined alternative parameter inputs for selected scenarios, changing one parameter value at a time, keeping the other parameters at archetype values. We varied: the effect of U on annual rate of cognitive change (β_{13}); the effect of the exposure on the log hazard rate of mortality (γ_1); the effect of U on the log hazard rate of mortality (γ_3); the interaction between exposure and U in determining the log hazard rate of mortality (γ_4); and the effect of rate of cognitive change on the log hazard rate of mortality (γ_5). We did not vary the distributions of random effects, within-person covariance structure, or measurement error for cognitive function. These distributions are described in **Table 2**.

Because the proportion of the study population that dies throughout follow-up may influence the degree of selection bias, we also varied cumulative mortality by the end of hypothetical follow-up. For each causal scenario, we generated the data under low, intermediate, and high mortality conditions

(cumulative mortality 25%, 50%, and 75%, respectively). To achieve this across causal scenarios, we adjusted the baseline hazard such that cumulative mortality equaled the target rate.

Assessment of survival bias in estimated exposure-cognitive change associations

As previously stated, we assumed the sharp null hypothesis that there is no person in the population for whom the exposure affects rate of cognitive change ($\beta_{11} = 0$). Therefore, we attributed any non-null association of the exposure on rate of cognitive change to survival bias. We produced estimates of the effect of exposure on rate of cognitive change using measured values of cognitive function (C_{ij}^{*}) among people who were alive at each cognitive assessment based on linear mixed effects models and population average effects models estimated using generalized estimating equations (GEE), the two modeling approaches for repeated measures most commonly used in practice. In all modeling approaches, a person was censored at her last cognitive assessment prior to death or at the end of follow up (9 years), whichever came first. Consistent with the data generating process, we used study time as the timescale and adjusted for baseline age. All models adopted a naïve analysis approach without adjusting for any measure of U to reflect the fact that in practice, such a measure may not be available.

We estimated linear mixed effects models with random intercepts and slopes allowing for possible correlation of the random intercepts and random slopes and no additional within-person covariance structure, as is common practice for standard linear mixed effects models. For the GEE approach, we specified an autoregressive within-person correlation structure, where $Corr\left(C_{ij}, C_{ij+k}\right) = \rho^{k}$ for all *j* and *k*. The GEE approach is often adopted because it is considered robust to misspecifications of the covariance matrix.

Neither of these commonly-used models is completely consistent with our data generating model, which incorporated an autoregressive within-person covariance structure in addition to random intercepts and slopes; thus, as a sensitivity analysis, we also estimated linear mixed effects models with

an autoregressive within-person covariance structure, where $Corr(e_{ij}, e_{ij+k}) = \rho^k$ for all *j* and *k*. These models more closely approximate the data generating model.

Across the B=1,000 simulated samples, $\beta_{11} = 0 =$ true value of the effect of exposure on rate of cognitive decline and $\hat{\beta}_{11} =$ estimated effect of exposure on rate of cognitive decline. Because effect sizes for annual rate of cognitive change are relatively small, we expressed effect sizes in increments over 10 years for clarity. We estimated the mean of the estimate over all simulations, $\overline{\hat{\beta}}_{11} = \sum_{k=1}^{B} \frac{\hat{\beta}_{11_k}}{B}$. We assessed accuracy as root-mean-square error, which is the square root of the mean squared deviation of the estimated effect of exposure on rate of cognitive decline ($\hat{\beta}_{11}$) from the true value ($\beta_{11} = 0$). We estimated the 95% confidence interval coverage as the proportion of simulations in which the 95% confidence interval for $\hat{\beta}_{11}$ included $\beta_{11} = 0$. We do not report percent bias because data generation is under the null.

In Scenarios B1 and B2, conditioning on survival induces collider-stratification bias because U is an unmeasured common cause of mortality and rate of cognitive change. To illustrate the association between exposure and U induced by selective survival in these scenarios, we examined the difference in mean value of U between exposed and unexposed survivors at each wave of the study for Scenario A (where no bias was anticipated) and Scenarios B1 and B2 with low, intermediate, and high mortality. Scenarios C1 and C2 are omitted because selective survival induces negligible associations between exposure and U in these scenarios; the bias in these scenarios does not arise from U.

We used Stata SE version 13.1 (StataCorp LP, College Station, Texas) for all data generation and analyses. The outline of steps for running the code (**Web Figure 2**), code books for the simulation code (**Web Tables 2-3**), and simulation code (**Web Appendix 2-3**) are available online).

RESULTS

Recall that in all scenarios, exposure has no effect on rate of cognitive change ($\beta_{11} = 0$), so any association between exposure and rate of cognitive change ($\hat{\beta}_{11} \neq 0$) represents survival bias. In Scenario A, the estimated effect of exposure and rate of cognitive change was unbiased, as expected, with estimated effects centered around the null with approximately correct coverage under low, intermediate, and high mortality for both linear mixed effects (**Table 3**, top panel) and the GEE approach (**Table 3**, bottom panel). The estimated effect of exposure on rate of cognitive change was biased to some degree in all other causal scenarios (**Tables 3-6**). The extent of bias was similar for linear mixed effects models and the GEE approach, so we discuss only the mixed model results in detail.

With archetype parameter values, in Scenario B1 (exposure and U affect mortality), we would estimate that exposure reduces rate of cognitive decline by 0.05 and 0.09 standard deviation per decade under low (25%) and high (75%) mortality, respectively (**Table 2**). Coverage for Scenario B1 ranged from 91% under low mortality to 88% under high mortality. The magnitude of bias was much larger for Scenario B2, where exposure and U have more than multiplicative effects on mortality hazard. Scenario C1 (exposure and rate of cognitive change affect mortality) also entailed substantial bias. We would estimate that exposure reduces rate of cognitive decline by 0.09 and 0.32 standard deviation units per decade under low and high mortality, respectively. Coverage for Scenario C1 ranged from 81% under low mortality to 23% under high mortality. Results for Scenario C2, in which rate of cognitive change *and* cognitive level affect mortality, were similar to results for Scenario C1. With more moderate parameter inputs in Scenarios B1, B2, and C1, the magnitude of bias and coverage improved (**Tables 4-6**).

In sensitivity analyses, the extent of bias in estimates from linear mixed effects models with autoregressive within-person covariance structure were similar to results from linear mixed effects models with no additional within-person covariance structure beyond that imposed by the random intercepts and slopes and the GEE approach (**Web Tables 4-7**).

Underlying selective survival bias in Scenarios B1 and B2 is a correlation between the exposure and U induced by restricting analyses to people who are alive. If U is correlated with both the exposure and rate of cognitive decline, in effect it becomes a confounder of this relationship. Thus, the larger the association between the exposure and U, the larger the magnitude of survival bias. **Figure 2** shows the difference in mean value of U between exposed and unexposed participants alive at each wave of the study from Scenarios A, B1, and B2 with archetype input values under A) low mortality (25%), B) intermediate mortality (50%), and C) high mortality (75%). Recall that the data generating processes for U and exposure were independent, so any association is due to selective survival. As expected, no association between exposure and U is induced in Scenario A. An association between exposure and U is induced in Scenario B1 (exposure and U affect mortality), but a much stronger association is induced in Scenario B2 (exposure and U have more than multiplicative effects on mortality hazard), and the association increases as cumulative mortality increases.

DISCUSSION

We developed and implemented a simulation platform to quantify possible bias due to selective survival in studies of cognitive aging with truncation by death. This platform allows flexible specification of duration of follow-up, cumulative mortality, distribution and magnitude of effects of determinants of mortality, and distribution of cognitive level and rate of cognitive change. This flexibility allows the simulation to be adapted to diverse studies. We implemented simulations across a range of scenarios corresponding with assumptions of substantive researchers. We observed substantial bias in data generating models in which exposure and an unmeasured determinant of cognitive decline interacted on the hazard ratio scale to influence mortality. The magnitude of bias was more modest when the exposure and unmeasured determinant of cognitive decline did not interact on the hazard ratio scale to

influence mortality. Bias consistently arose when both exposure and rate of cognitive decline directly influenced mortality. Bias was, as expected, larger in high-mortality situations.

It is widely-recognized that selective survival has the potential to bias estimated effects of exposures on rate of cognitive decline (5, 9), but there have been few tools to systematically estimate plausible magnitude of this bias when many determinants of survival are not known or not measured. We focused here on bias under the sharp null, but similar bias might impede researchers' ability to identify exposures that have protective or harmful effects on cognitive aging. We found that the magnitude of survival bias often appears large enough to obscure likely causal effects of exposures in typical cohorts. For example, in the Three-City Dijon Study (3C), carriers of an APOE-e4 allele, one of the strongest known predictors of dementia (10), declined approximately 0.14 more standard deviation units over 10 years on the Mini Mental State Examination for global cognition than non-carriers (11). In many scenarios, we observed magnitudes of bias similar to this effect size. Few exposures have been found to be consistently associated with rate of cognitive decline (12). This surprising paucity of consistent predictors of cognitive decline may be partially explained by survival bias.

As expected, cumulative mortality over follow-up substantially influenced extent of bias. Cumulative mortality in studies of cognitive decline and dementia varies substantially across studies, depending in part on the age of study participants and length of follow-up. For example, in the Atherosclerosis Risk in Communities cohort, 15% of participants (baseline age 48-70) died over 14 years of follow-up (13); in the Sacramento Area Latino Study of Aging study, 23% of participants (baseline age ≥60) died over 10 years of follow-up (14); and in the Chicago Health and Aging Project, 54% of participants (baseline age 65-109) died over 12 years of follow-up (9).

Survival bias is often conceptualized as a type of collider-stratification bias (5, 8). Hernán *et al.* noted that when the exposure and a determinant of the outcome influence the collider, collider-stratification bias does not occur if the data follow a multiplicative survival model such that the two

causes of the collider have perfectly multiplicative effects on probability of remaining alive (5, 8). In other words, multiplicative effects imply the conditional probability of survival given E and U is equal to a product of functions of e and u, *i.e.*, $\Pr(S = 1|E = e, U = u) = g(e)h(u)$, where g(.) and h(.) represent arbitrary functions (Appendix A.3 in Hernán *et al.* 2004) (5, 8). Consistent with this, we observed substantial bias when exposure and U interacted on the hazard ratio scale in determining mortality (Scenario B2) and much smaller bias when the hazard function for mortality was a multiplicative function of the exposure and U (Scenario B1). This is consistent with findings from previous simulation studies, where the magnitude of collider-stratification bias was small under causal structures similar to Scenario B1 unless the effects of U on mortality were very large (7, 15, 16). None of our scenarios assessed the situation in which exposure and U were perfectly multiplicative for probability of remaining alive, the situation in which no survival bias would occur, although Scenario B1 (exposure and U are multiplicative for the hazard of death) approximates this situation if mortality is rare. We set up the simulation to specify covariate effects on the hazard of mortality, rather than on the cumulative probability of remaining alive at the end of follow-up as implied by the formulation by Hernán *et al.*, because the hazard of mortality formulation is more typically encountered in epidemiologic research.

We also observed substantially biased estimation of the exposure's effect on rate of cognitive decline when the exposure and (1) rate of cognitive change or (2) rate of cognitive change and level of cognitive function influenced mortality (Scenarios C1 and C2). While rate of cognitive decline itself may not directly influence mortality *per se*, rate of cognitive decline can be conceptualized as a surrogate for progression of underlying brain disease. The link between cognitive decline and impending mortality is consistent with the terminal decline and terminal drop hypotheses (17, 18). Empirical studies are needed to better understand whether level of cognitive function or rate of cognitive change is more

predictive of mortality among older adults, as this is important for understanding potential degree of survival bias in research on determinants of cognitive decline.

Although some of the parameters that determine the magnitude of survival bias are easily observed in a cohort study (e.g., cumulative mortality), others are more speculative. Our study highlights the importance of deriving plausible estimates for these parameters, including non-multiplicative influences on mortality, effects of cognitive decline on mortality, and the plausible magnitude of various rarely measured confounders of rate of cognitive decline and mortality. We selected archetype values that represented large associations between variables to assess survival bias when the associations were on the upper bound of what we considered likely in real data, but researchers studying a specific exposure of interest will need to apply input values that they deem plausible for the substantive question.

We made simplifying assumptions in order to maintain transparency and constrain the number of input parameters. It is our goal that our simulation platform will be modified to study extensions of the present work and improve understanding of selective survival in cognitive aging research. For example, extensions could incorporate loss to follow-up as an additional source of attrition, allow non-null effects of exposure on rate of cognitive change, or introduce left-censoring such that some mortality occurs between exposure and baseline cognitive assessment. These challenges are relevant for understanding a host of empirical findings in neuroepidemiology and aging research.

The conceptualization of survival bias has been controversial because counterfactual values of the outcome are not defined for the dead (9, 19, 20). Various conceptual approaches have been proposed in response, including a principal stratification approach defining the target of inference as the survivor average causal effect (SACE), *i.e.*, the effect of exposure on outcomes of individuals who would survive under either exposure regime (21). The simulation platform can be used to evaluate performance of an analytic approach for any counterfactual contrast of interest, including the SACE.

The field of dementia research has recognized the potential importance of selective survival and the causal structures that can theoretically give rise to survival bias. It is time for the field to move towards quantifying the potential magnitude of bias (22). We provide an accessible and flexible tool for examining biases in research of determinants of cognitive decline. This simulation platform can be modified to examine performance of various methods to address survival bias according to study characteristics such as the interval between visits or frequency of dropout before death (21, 23, 24). The platform can also be used to quantify the potential magnitude of bias arising from other methodological challenges in cognitive aging research, including time-varying exposures and confounders (25) and unequal interval scaling in measurement of cognitive function (26).

REFERENCES

1. Lavery LL, Dodge HH, Snitz B, et al. Cognitive Decline and Mortality in a Community-Based Cohort: The Monongahela Valley Independent Elders Survey. *Journal of the American Geriatrics Society* 2009;57(1):94-100.
2. Yaffe K, Lindquist K, Vittinghoff E, et al. The effect of maintaining cognition on risk of disability and death. *Journal of the American Geriatrics Society* 2010;58(5):889-894.
3. Thorvaldsson V, Hofer SM, Berg S, et al. Onset of terminal decline in cognitive abilities in individuals without dementia. *Neurology* 2008;71(12):882-887.
4. Wilson RS, Beckett LA, Bienias JL, et al. Terminal decline in cognitive function. *Neurology* 2003;60(11):1782-1787.
5. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004;15(5):615-625.
6. Bender R, Augustin T, Blettner M. Generating survival times to simulate Cox proportional hazards models. *Statistics in medicine* 2005;24(11):1713-1723.
7. Glymour MM, Vittinghoff E. Commentary: selection bias as an explanation for the obesity paradox: just because it's possible doesn't mean it's plausible. *Epidemiology* 2014;25(1):4-6.
8. Hernán MA, Robins JM. *Causal Inference*. Boca Raton: Chapman & Hall/CRC, forthcoming; 2016.
9. Weuve J, Tchetgen Tchetgen EJ, Glymour MM, et al. Accounting for bias due to selective attrition: the example of smoking and cognitive decline. *Epidemiology* 2012;23(1):119-128.
10. Corder E, Saunders A, Strittmatter W, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 1993;261(5123):921-923.
11. Vivot A, Glymour MM, Tzourio C, et al. Association of Alzheimer's related genotypes with cognitive decline in multiple domains: results from the Three-City Dijon study. *Molecular Psychiatry* 2015;20(10):1173-1178.
12. Plassman BL, Williams JW, Jr., Burke JR, et al. Systematic review: factors associated with risk for and possible prevention of cognitive decline in later life. *Annals of internal medicine* 2010;153(3):182-193.

13. Mayeda ER, Haan MN, Neuhaus J, et al. Type 2 diabetes and cognitive decline over 14 years in middle-aged African Americans and whites: the ARIC Brain MRI Study. *Neuroepidemiology* 2014;43(3-4):220-227.
14. Mayeda ER, Haan MN, Kanaya AM, et al. Type 2 diabetes and 10-year risk of dementia and cognitive impairment among older Mexican Americans. *Diabetes care* 2013;36(9):2600-2606.
15. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology* 2003;14(3):300-306.
16. Liu W, Brookhart MA, Schneeweiss S, et al. Implications of M bias in epidemiologic studies: a simulation study. *American journal of epidemiology* 2012;176(10):938-948.
17. MacDonald SW, Hultsch DF, Dixon RA. Aging and the shape of cognitive change before death: terminal decline or terminal drop? *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 2011;66(3):292-301.
18. Sliwinski MJ, Stawski RS, Hall CB, et al. Distinguishing preterminal and terminal cognitive decline. *European Psychologist* 2006;11(3):172-181.
19. Chaix B, Evans D, Merlo J, et al. Commentary: Weighing up the dead and missing: reflections on inverse-probability weighting and principal stratification to address truncation by death. *Epidemiology* 2012;23(1):129-131.
20. Tchetgen EJT, Glymour MM, Shpitser I, et al. Rejoinder: To Weight or Not to Weight?: On the Relation Between Inverse-probability Weighting and Principal Stratification for Truncation by Death. *Epidemiology* 2012;23(1):132-137.
21. Tchetgen Tchetgen EJ. Identification and estimation of survivor average causal effects. *Statistics in Medicine* 2014;33(21):3601-3628.
22. Weuve J, Proust-Lima C, Power MC, et al. Guidelines for reporting methodological challenges and evaluating potential bias in dementia research. *Alzheimer's & Dementia* 2015;11(9):1098-1109.
23. Kurland BF, Johnson LL, Egleston BL, et al. Longitudinal data with follow-up truncated by death: match the analysis method to research aims. *Statistical science: a review journal of the Institute of Mathematical Statistics* 2009;24(2):211.
24. Varadhan R, Xue Q-L, Bandeen-Roche K. Semicompeting risks in aging research: methods, issues and needs. *Lifetime data analysis* 2014;20(4):538-562.
25. Power MC, Weuve J, Sharrett AR, et al. Statins, cognition, and dementia-systematic review and methodological commentary. *Nature reviews Neurology* 2015;11(4):220-229.
26. Proust-Lima C, Dartigues JF, Jacqmin-Gadda H. Misuse of the linear mixed model when evaluating risk factors of cognitive decline. *American journal of epidemiology* 2011;174(9):1077-1088.

Figure 1. Causal scenarios. In all scenarios, the target of inference is the effect of the exposure on rate of change in cognitive function over time (rate of cognitive decline). "Cognitive Intercept" reflects cognitive level at the baseline cognitive assessment. The bidirectional arrow between "Cognitive Intercept" and "Rate of Cognitive Decline" indicates that the random intercept and random slope terms were drawn from a bivariate distribution with nonzero covariance. The box around "Survival" represents the fact that we are conditioning on survival=1, meaning analyses are limited to people who are alive at each cognitive assessment. Age is a determinant of survival, cognitive intercept, and rate of cognitive decline; the box around "Age" represents the fact that we are conditioning on age by adjusting for age in all regression models for cognitive change. A) Scenario A: no anticipated survivor bias. B) Scenarios B1 (exposure and U affect mortality) and B2 (exposure and U have more than multiplicative effects on mortality hazard): mortality is influenced by exposure and U, an unmeasured determinant of rate of cognitive change, without and with an interaction on the hazard ratio scale between the exposure and U. C) Scenario C1 (exposure and rate of cognitive change affect mortality): mortality risk is influenced by exposure and current rate of cognitive change. D) Scenario C2 (exposure, rate of cognitive change, and cognitive level affect mortality) mortality risk is influenced by exposure, current rate of cognitive change, and level of cognitive function the beginning of the time interval.

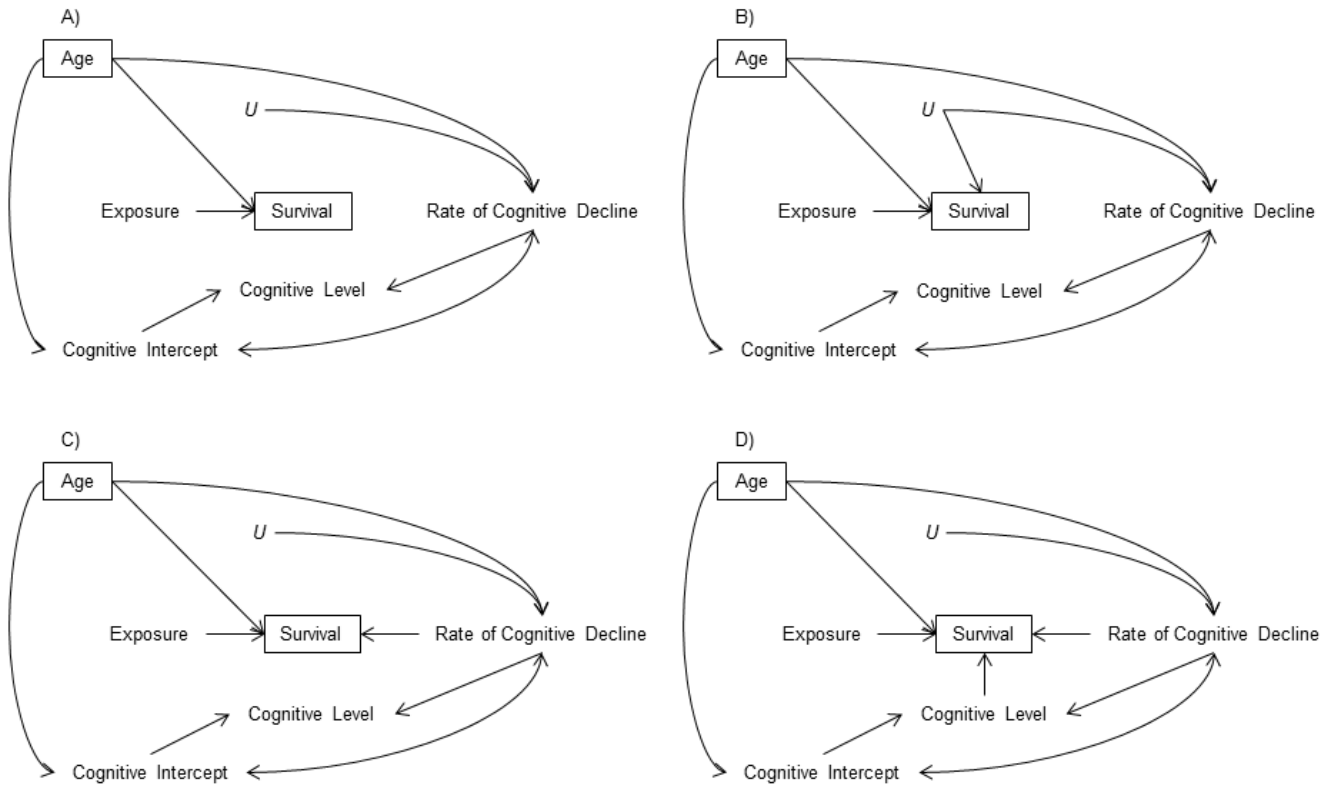
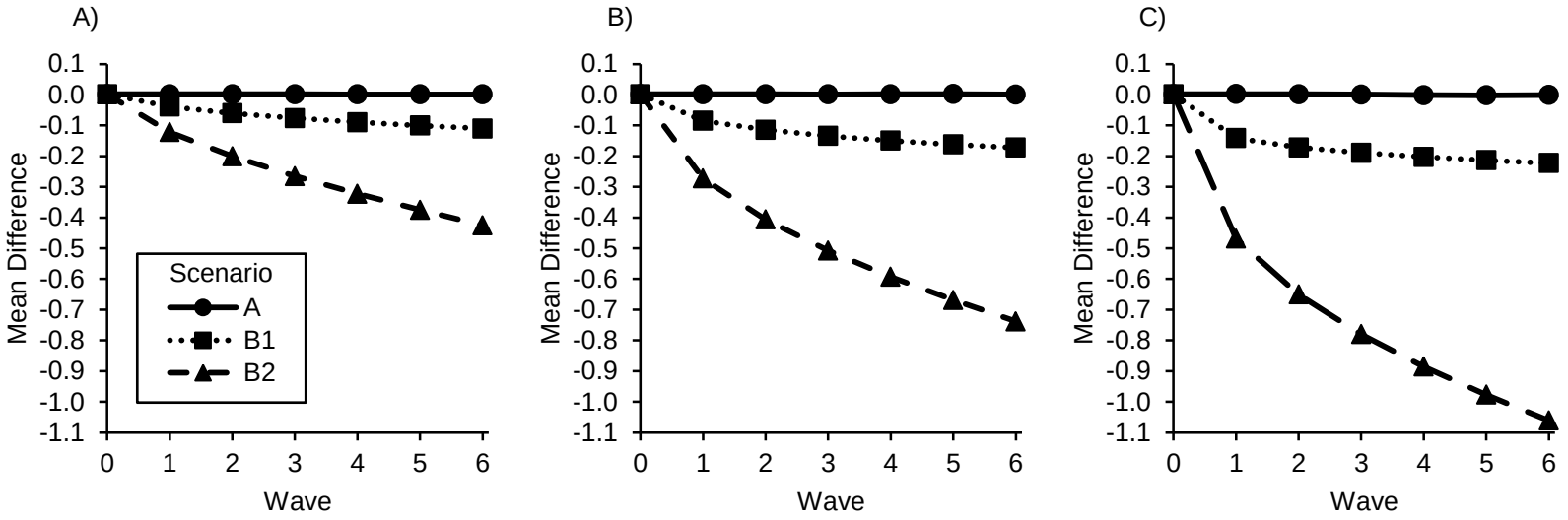


Figure 2: Difference in mean value of U between exposed and unexposed participants alive at each wave of the study, Scenarios A, B1, and B2 (archetype input values) under A) low mortality (25%), B) intermediate mortality (50%), and C) high mortality (75%). Scenarios C1 and C2 are not included because selective survival induces negligible associations between exposure and U in these scenarios; the bias does not arise from U.



Box 1. Definitions of regression parameters for generation of cognitive function and hazard of death ($h(t_{ij}|x)$).

Data generating model for cognitive function: $C_{ij} = \beta_{00} + \beta_{01}exposure_i + \beta_{02}baseline\_age_i + \beta_{03}U_i + (\beta_{10} + \beta_{11}exposure_i + \beta_{12}baseline\_age_i + \beta_{13}U_i)time_{ij} + \zeta_{0i} + \zeta_{1i}time_{ij} + \varepsilon_{ij}$ (equation 1)
β_{00} = Group mean cognitive intercept (baseline level of cognitive function) for the unexposed. β_{10} = Group mean cognitive slope (annual rate of cognitive change) for the unexposed. β_{01} = Effect of exposure on level of cognitive function at baseline. β_{02} = Effect of one year change in baseline age (in years, centered at age 60) on level of cognitive function at baseline. β_{03} = Effect of one unit change in U on level of cognitive function at baseline. β_{11} = Effect of exposure on annual rate of cognitive change. β_{12} = Effect of one year change in baseline age on annual rate of cognitive change. β_{13} = Effect of one unit change in U on annual rate of cognitive change. ζ_{0i} = Person *i*'s deviation from the group mean intercept. ζ_{1i} = Person *i*'s deviation from the group mean slope.
Data generating model for hazard of death: $h(t_{ij}\|x) = \lambda \exp(\gamma_1 exposure_i + \gamma_2 age_{ij} + \gamma_3 U_i + \gamma_4 exposure_i * U_i + \gamma_5(slope_{ij}) + \gamma_6 C_{ij})$ (equation 4)
γ_1 = Effect of the exposure on the log hazard rate of mortality. γ_2 = Effect of one year change in age on the log hazard rate of mortality. γ_3 = Effect of one unit change in U on the log hazard rate of mortality. γ_4 = Additional effect of one unit change in U on the log hazard rate of mortality among people who are exposed. γ_5 = Effect of one unit change in rate of cognitive change on the log hazard rate of mortality. γ_6 = Effect of one unit change in level of cognitive function on the log hazard rate of mortality.

Table 1. Summary table of archetype values for each causal scenario. Shaded columns represent parameters for which alternative values, which represent more moderate effect sizes, are considered.

	β_{00}	β_{01}	β_{02}	β_{03}	β_{10}	β_{11}	β_{12}	β_{13}	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6
Scenario A[a,b]	0	0	-0.05	0	-0.05	0	-0.005	-0.05	0.69	0.086	0	0	0	0
Scenario B1[a,c]	0	0	-0.05	0	-0.05	0	-0.005	-0.05	0.69	0.086	1.61	0	0	0
Scenario B2[a,d]	0	0	-0.05	0	-0.05	0	-0.005	-0.05	0.69	0.086	0	1.61	0	0
Scenario C1[a,e]	0	0	-0.05	0	-0.05	0	-0.005	-0.05	0.69	0.086	0	0	-2.9	0
Scenario C2[a,f]	0	0	-0.05	0	-0.05	0	-0.005	-0.05	0.69	0.086	0	0	-2.9	-0.69

[a]All Scenarios are variations on the general data generating structure, which is:

$$C_{ij} = \beta_{00} + \beta_{01} exposure_i + \beta_{02} baseline\_age_i + \beta_{03} U_i + (\beta_{10} + \beta_{11} exposure_i + \beta_{12} baseline\_age_i + \beta_{13} U_i) time_{ij} + \zeta_{0i} + \zeta_{1i} time_{ij} + \varepsilon_{ij}$$

$$h(t_{ij}|x) = \lambda \exp(\gamma_1 exposure_i + \gamma_2 baseline\_age_{ij} + \gamma_3 U_i + \gamma_4 exposure_i * U_i + \gamma_5 slope_{ij} + \gamma_6 C_{ij})$$

[b]Scenario A: no anticipated survivor bias. [c]Scenario B1: exposure and U affect mortality. [d]Scenario B2: exposure and U have more than multiplicative effects on mortality hazard. [e]Scenario C1: exposure and rate of cognitive change affect mortality. [f]Scenario C2: exposure, rate of cognitive change, and cognitive level affect mortality.

Table 2. Input variance and covariance values for generating C_{ij} (true cognitive function) (equation 1[a]) and C_{ij}* (measured cognitive function) (equation 3[b]):

Parameter	Definition	Value
$\sigma^2_{\zeta_0}$	Variance of ζ_{0i}, person *i*'s deviation from the group mean intercept	0.20
$\sigma^2_{\zeta_1}$	Variance of ζ_{1i}, person *i*'s deviation from the group mean slope	0.005
$\sigma_{\zeta_{01}}$	Covariance of $\zeta_{0i}$and ζ_{1i}	0.01
${\sigma_\varepsilon}^2$	Variance of ε_{ij}, unexplained variation in C_{ij}	0.70
ρ	Covariance between ε_{ij} and ε_{ij+1}	0.40
σ^2_δ	Variance of δ_{ij}, random measurement error of cognitive function	0.19

[a] $C_{ij} = \beta_{00} + \beta_{01}exposure_i + \beta_{02}baseline\_age_i + \beta_{03}U_i + (\beta_{10} + \beta_{11}exposure_i + \beta_{12}baseline\_age_i + \beta_{13}U_i)time_{ij} + \zeta_{0i} + \zeta_{1i}time_{ij} + \varepsilon_{ij}$

[b] $C^*_{ij} = C_{ij} + \delta_{ij}$

Table 3: Scenario A-C simulation results for the estimated effect of exposure on rate of cognitive change (β_{11}) per 10 years: archetype input values.

	Low mortality (25%)			Intermediate mortality (50%)			High mortality (75%)		
	$\overline{\hat{\beta}}_{11}$	RMSE	Coverage[a]	$\overline{\hat{\beta}}_{11}$	RMSE	Coverage[a]	$\overline{\hat{\beta}}_{11}$	RMSE	Coverage[a]
Linear mixed effects models									
Scenario A	0.002	0.085	0.954	0.000	0.095	0.958	-0.002	0.120	0.957
Scenario B1	0.048	0.098	0.914	0.075	0.124	0.884	0.094	0.155	0.883
Scenario B2	0.164	0.186	0.527	0.288	0.304	0.165	0.412	0.429	0.081
Scenario C1	0.094	0.126	0.809	0.186	0.209	0.518	0.316	0.335	0.229
Scenario C2	0.088	0.121	0.818	0.170	0.193	0.562	0.288	0.308	0.264
GEE approach									
Scenario A	0.001	0.088	0.964	-0.001	0.099	0.960	-0.005	0.128	0.949
Scenario B1	0.053	0.103	0.919	0.081	0.132	0.890	0.100	0.169	0.884
Scenario B2	0.189	0.210	0.464	0.326	0.342	0.122	0.458	0.476	0.061
Scenario C1	0.092	0.127	0.838	0.189	0.214	0.533	0.338	0.359	0.193
Scenario C2	0.099	0.131	0.825	0.183	0.207	0.527	0.297	0.319	0.286

Abbreviations: RMSE, root-mean-square error; GEE, generalized estimating equation. [a]Coverage = proportion of times the 95% confidence interval for $\hat{\beta}_{11}$ includes β_{11}.

Table 4: Scenario B1 (exposure and U affect mortality) simulation results for the estimated effect of exposure on rate of cognitive change (β_{11}) per 10 years: alternative input values (more moderate effect sizes).

	Low mortality (25%)			Intermediate mortality (50%)			High mortality (75%)		
	$\overline{\hat{\beta}}_{11}$	RMSE	Coverage[a]	$\overline{\hat{\beta}}_{11}$	RMSE	Coverage[a]	$\overline{\hat{\beta}}_{11}$	RMSE	Coverage[a]
Linear mixed effects models									
Under archetypical conditions[b]	0.048	0.098	0.914	0.075	0.124	0.884	0.094	0.155	0.883
Moderate effect of exposure on mortality[c]	0.029	0.091	0.934	0.044	0.108	0.923	0.056	0.135	0.922
Moderate effect of U on mortality[d]	0.034	0.092	0.938	0.062	0.114	0.912	0.089	0.150	0.902
Moderate effect of U on cognitive decline[e]	0.026	0.087	0.939	0.039	0.104	0.922	0.048	0.130	0.933
GEE approach									
Under archetypical conditions[b]	0.053	0.103	0.919	0.081	0.132	0.890	0.100	0.169	0.884
Moderate effect of exposure on mortality[c]	0.032	0.094	0.947	0.047	0.114	0.932	0.056	0.135	0.922
Moderate effect of U on mortality[d]	0.041	0.098	0.947	0.072	0.123	0.901	0.087	0.149	0.901
Moderate effect of U on cognitive decline[e]	0.027	0.090	0.952	0.041	0.109	0.940	0.047	0.130	0.937

Abbreviations: RMSE, root-mean-square error; GEE, generalized estimating equation. [a]Coverage = proportion of times the 95% confidence interval for $\hat{\beta}_{11}$ includes β_{11}; [b]Results in this row replicate results for Scenario B1 shown in Table 3 for convenience of comparisons; [c]Input value for γ_1=0.40 (HR=1.5) instead of 0.69 (HR=2.0); [d]Input value for γ_3=0.69 (HR=2.0) instead of 1.61 (HR=5.0); [e]Input value for β_{13}=-0.025 instead of -0.05.

Table 5: Scenario B2 (exposure and U have more than multiplicative effects on mortality hazard) simulation results for the estimated effect of exposure on rate of cognitive change (β_{11}) per 10 years: alternative input values (more moderate effect sizes).

	Low mortality (25%)			Intermediate mortality (50%)			High mortality (75%)		
	$\overline{\hat{\beta}}_{11}$	RMSE	Coverage[a]	$\overline{\hat{\beta}}_{11}$	RMSE	Coverage[a]	$\overline{\hat{\beta}}_{11}$	RMSE	Coverage[a]
Linear mixed effects models									
Under archetypical conditions[b]	0.164	0.186	0.527	0.288	0.304	0.165	0.412	0.429	0.081
Moderate effect of exposure on mortality[c]	0.154	0.176	0.568	0.268	0.285	0.225	0.383	0.401	0.117
Moderate effect of exposure and U on mortality[d]	0.079	0.117	0.854	0.158	0.185	0.635	0.256	0.283	0.444
Moderate effect of U on cognitive decline[e]	0.085	0.119	0.836	0.147	0.174	0.671	0.209	0.238	0.592
GEE approach									
Under archetypical conditions[b]	0.189	0.210	0.464	0.326	0.342	0.122	0.458	0.476	0.061
Moderate effect of exposure on mortality[c]	0.177	0.199	0.517	0.303	0.320	0.161	0.427	0.446	0.090
Moderate effect of exposure and U on mortality[d]	0.097	0.132	0.831	0.191	0.216	0.539	0.304	0.330	0.312
Moderate effect of U on cognitive decline[e]	0.096	0.130	0.845	0.163	0.192	0.637	0.228	0.260	0.562

Abbreviations: RMSE, root-mean-square error; GEE, generalized estimating equation. [a]Coverage = proportion of times the 95% confidence interval for $\hat{\beta}_{11}$ includes β_{11}; [b]Results in this row replicate results for Scenario B2 shown in Table 3 for convenience of comparisons; [c]Input value for γ_1=0.40 (HR=1.5) instead of 0.69 (HR=2.0); [d]Input value for γ_4=0.69 (HR=2.0) instead of 1.61 (HR=5.0); [e]Input value for β_{13}=-0.025 instead of -0.05.

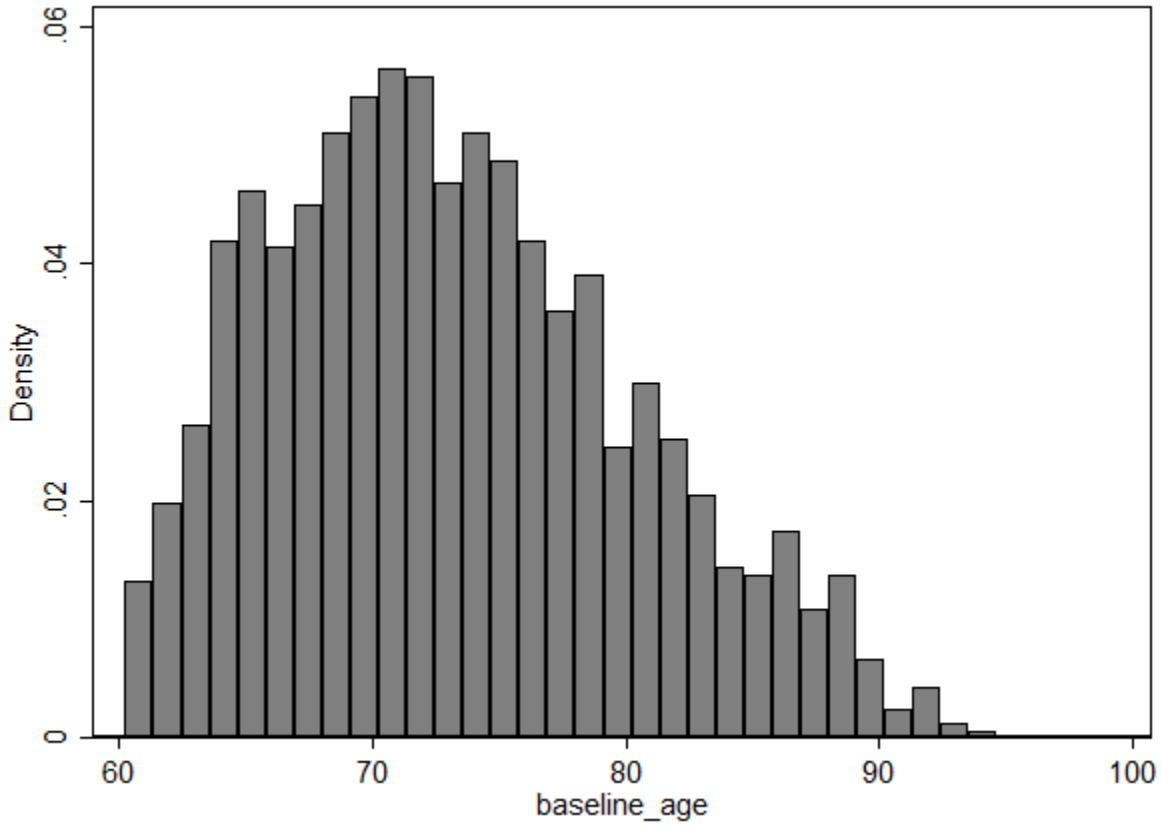
Table 6: Scenario C1 (exposure and rate of cognitive change affect mortality) simulation results for the estimated effect of exposure on rate of cognitive change (β_{11}) per 10 years: alternative input values (more moderate effect sizes).

	Low mortality (25%)			Intermediate mortality (50%)			High mortality (75%)		
	$\overline{\hat{\beta}}_{11}$	RMSE	Coverage[a]	$\overline{\hat{\beta}}_{11}$	RMSE	Coverage[a]	$\overline{\hat{\beta}}_{11}$	RMSE	Coverage[a]
Linear mixed effects models									
Under archetypical conditions[b]	0.094	0.126	0.809	0.186	0.209	0.518	0.316	0.335	0.229
Moderate effect of exposure on mortality[c]	0.055	0.100	0.902	0.107	0.142	0.819	0.180	0.212	0.673
Moderate effect of rate of cognitive change on mortality[d]	0.043	0.095	0.928	0.097	0.137	0.828	0.189	0.225	0.664
GEE approach									
Under archetypical conditions[b]	0.092	0.127	0.838	0.189	0.214	0.533	0.338	0.359	0.193
Moderate effect of exposure on mortality[c]	0.053	0.102	0.924	0.108	0.147	0.819	0.191	0.225	0.649
Moderate effect of rate of cognitive change on mortality[d]	0.041	0.097	0.942	0.098	0.141	0.840	0.207	0.244	0.624

Abbreviations: RMSE, root-mean-square error; GEE, generalized estimating equation. [a]Coverage = proportion of times the 95% confidence interval for $\hat{\beta}_{11}$ includes β_{11}; [b]Results in this row replicate results for Scenario C1 shown in Table 3 for convenience of comparisons; [c]Input value for γ_1=0.40 (HR=1.5) instead of 0.69 (HR=2.0); [d]Input value for γ_5=-1.0 (HR=0.4) instead of -2.9 (HR=0.1).

Web Materials

Web Figure 1. Baseline age (baseline_age) defined as 60 plus a random variable drawn from a Beta(2,4) distribution which we scaled to allow a possible 40-year age range. Below figure shows the resulting baseline age distribution for one sample.



Web Table 1. Correlation matrix of repeated cognitive measures from one simulated data set for Scenario A with 25% mortality.

	C1	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**
C1	1.00						
C2	0.55	1.00					
C3	0.41	0.58	1.00				
C4	0.32	0.50	0.62	1.00			
C5	0.29	0.46	0.53	0.67	1.00		
C6	0.30	0.46	0.54	0.61	0.72	1.00	
C7	0.27	0.44	0.53	0.60	0.67	0.75	1.00

Web Appendix 1. Generation of survival time.

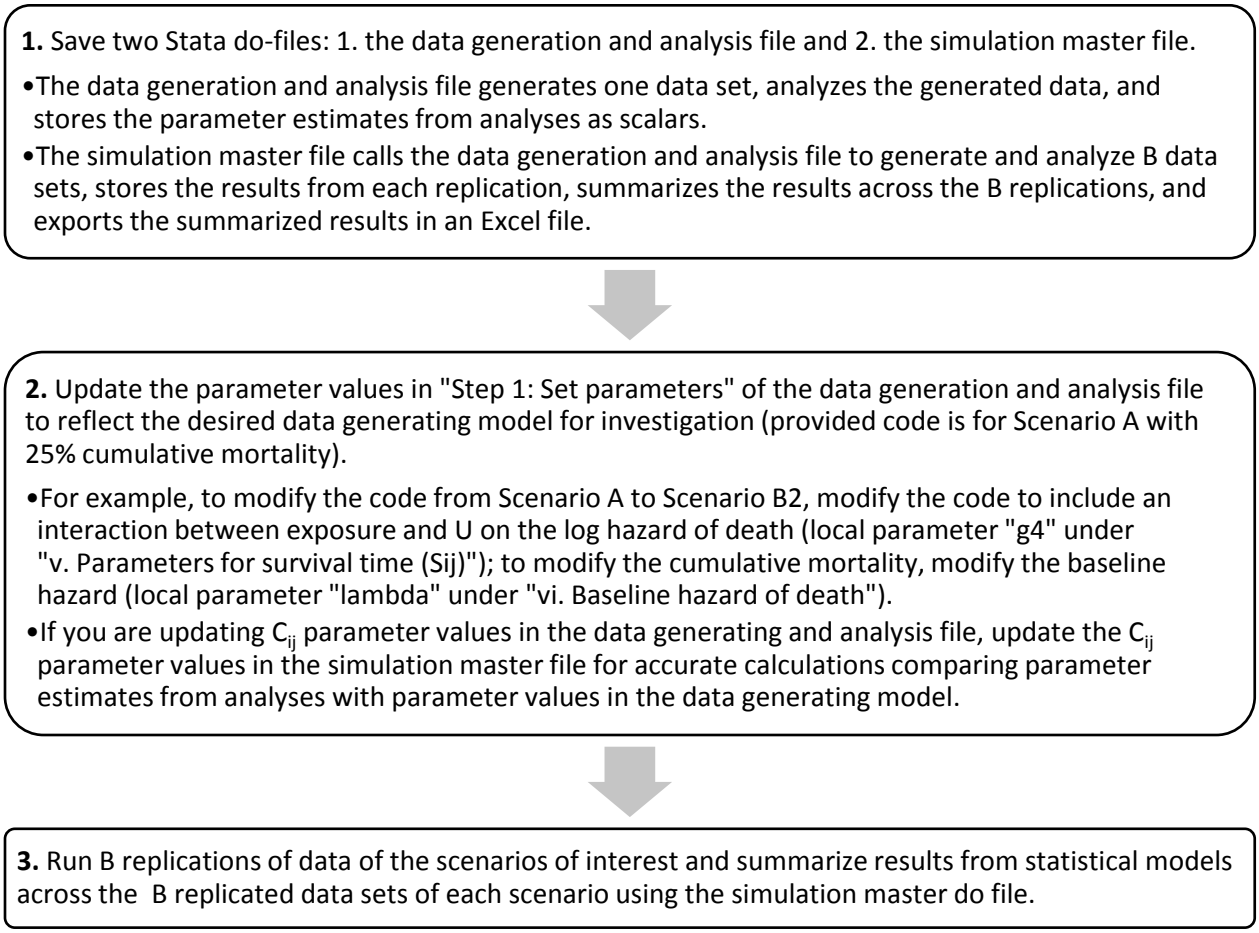
A person's survival time for a given time interval at risk (T_{ij}) is generated using the inverse cumulative hazard function transformation formula described by Bender *et al.* (1):

$$T_{ij} = \frac{-\log(U_{ij})}{\lambda \exp(\gamma_1 exposure_i + \gamma_2 age_{ij} + \gamma_3 U_i + \gamma_4 exposure_i * U_i + \gamma_5 slope_{ij} + \gamma_6 C_{ij})},$$

where U_{ij} is a uniform random variable ($U \sim U[0,1]$) and λ represents the baseline hazard.

1. Bender R, Augustin T, Blettner M. Generating survival times to simulate Cox proportional hazards models. *Statistics in medicine* 2005;24(11):1713-23.

Web Figure 2. Outline of steps for running the simulation code.



Web Table 2. Data generation and analysis file code book.

Variable	Definition
int_time	Time (in years) between cognitive assessments
pexp	Prevalence of exposure
Variances and correlations	
s2z0	Variance of random cognitive intercept ($\sigma^2_{\zeta_0}$)
s2z1	Variance of random cognitive slope ($\sigma^2_{\zeta_1}$)
sz01	Covariance of random intercept and random slope ($\sigma_{\zeta_{01}}$)
s2e	Variance of noise for C_{ij} ($\sigma_\varepsilon{}^2$)
rho	Correlation between noise terms for C_{ij} (ρ)
s2d	Variance of measurement error of C_{ij} (σ^2_δ)
Parameters for C_{ij}	
b00	Group mean cognitive intercept (baseline level of cognitive function) for the unexposed (β_{00})
b01	Effect of exposure on level of cognitive function at baseline (β_{01})
b02	Effect of one year change in baseline age (in years, centered at age 60) on level of cognitive function at baseline (β_{02})
b03	Effect of one unit change in U on level of cognitive function at baseline (β_{03})
b10	Group mean cognitive slope (annual rate of cognitive change) for the unexposed (β_{10})
b11	Effect of exposure on annual rate of cognitive change (β_{11})
b12	Effect of one year change in baseline age on annual rate of cognitive change (β_{12})
b13	Effect of one unit change in U on annual rate of cognitive change (β_{13})
Parameters for S_{ij}	
g1	Effect of the exposure on the log hazard rate of mortality (γ_1)
g2	Effect of one year change in age on the log hazard rate of mortality (γ_2)
g3	Effect of one unit change in U on the log hazard rate of mortality (γ_3)
g4	Additional effect of one unit change in U on the log hazard rate of mortality among people who are exposed (γ_4)
g5	Effect of one unit change in rate of cognitive change on the log hazard rate of mortality (γ_5)
g6	Effect of one unit change in level of cognitive function on the log hazard rate of mortality (γ_6)
lambda	Baseline hazard of death (λ). Increasing the value of this parameter will increase the number of deaths.
time0-time6	Time (in years) from baseline of cognitive assessment 0-6 ($time_{ij}$)
age1-age6	Age (in years) at cognitive assessment 0-6 (age_{ij})
age1_c60-age6_c60	Age centered at age 60 (in years) at cognitive assessment 0-6 (age_c60_{ij})
z0i	Person *i*'s deviation from the group mean intercept (ζ_{0i})
z1i	Person *i*'s deviation from the group mean slope (ζ_{1i})
epsilon0-6	Person *i*'s deviation from her predicted cognitive function at time *j*, where the prediction incorporates both covariates and random coefficients; this deviation reflects both sources correlated over time and time-specific deviations

alpha1-6	Deviation term for person *i*'s at time *j* that is independent of previous deviations.
sd_alpha	Standard deviation of distribution of alphas
true_cogfxn0-6	Person *i*'s true cognitive function at assessments 0-6 (C_{ij})
error0-6	Error in measurement of person *i*'s cognitive function at assessments 0-6
measured_cogfxn0-6	Person *i*'s measured cognitive function at assessments 0-6 (C_{ij}^{*})
avg_slope	Person *i*'s average rate of cognitive throughout the study
slope01-slope56	Person *i*'s rate of cognitive decline between assessments 0- 1, 1-2, etc.
U_01-U_56	Uniform random variable for survival time estimates for between assessments 0- 1, 1-2, etc.
survtime01-survtime56	Person *i*'s survival time (in years) for assessments 0- 1, 1-2, etc.
death01-death56	Indicator variable for person *i*'s death in interval 0-1, 1-2, etc.
study_death	Indicator variable for person *i*'s death during the study
survtime	Person *i*'s time (in years) from baseline to death; if person *i* does not die during the study, person *i*'s survtime is censored at the end of the study
agedeath	Person *i*'s age (in years) at death; if person *i* does not die during the study, person *i*'s ageatdeath is censored at the end of the study
Results from analyses stored as scalars	
p_study_death	Proportion of study participants who died
survtime	Mean follow-up time for study participants
Parameter estimates from conventional linear mixed effects models (linear mixed effects models with random intercepts and slopes allowing for possible correlation of the random intercept and slope and no additional within-person covariance structure)	
mixed_b00_int	Parameter estimate of group mean cognitive intercept (baseline level of cognitive function) for the unexposed ($\hat{\beta}_{00}$)
mixed_b00_int_se	Standard error of parameter estimate of group mean cognitive intercept (baseline level of cognitive function) for the unexposed (SE($\hat{\beta}_{00}$))
mixed_b01_exposure	Parameter estimate of effect of exposure on level of cognitive function at baseline ($\hat{\beta}_{01}$)
mixed_b01_exposure_se	Standard error of parameter estimate of effect of exposure on level of cognitive function at baseline (SE($\hat{\beta}_{01}$))
mixed_b02_age	Parameter estimate of effect of one year change in baseline age on level of cognitive function at baseline ($\hat{\beta}_{02}$)
mixed_b02_age_se	Standard error of parameter estimate of effect of one year change in baseline age on level of cognitive function at baseline (SE($\hat{\beta}_{02}$))
mixed_b10_time	Parameter estimate of group mean cognitive slope (annual rate of cognitive change) for the unexposed ($\hat{\beta}_{10}$)
mixed_b10_time_se	Standard error of parameter estimate of group mean cognitive slope (annual rate of cognitive change) for the unexposed (SE($\hat{\beta}_{10}$))
mixed_b11_exptime	Parameter estimate of effect of exposure on annual rate of cognitive change ($\hat{\beta}_{11}$)
mixed_b11_exptime_se	Standard error of parameter estimate of effect of exposure on annual rate of cognitive change (SE($\hat{\beta}_{11}$))
mixed_b12_agec60time	Parameter estimate of effect of one year change in baseline age on annual rate of cognitive change ($\hat{\beta}_{12}$)
mixed_b12_agec60time_se	Standard error of parameter estimate of one year change in baseline age on annual rate of cognitive change (SE($\hat{\beta}_{12}$)

mixed_z0i_est	Parameter estimate for random intercept
mixed_z0i_se	Standard error of parameter of random intercept
mixed_z1i_est	Parameter estimate of random slope
mixed_z1i_se	Standard error of parameter estimate of random slope
mixed_rho_est	Estimated covariance of random intercepts and random slopes
mixed_rho_se	Standard error of estimated covariance of random intercepts and slopes
Parameter estimates from GEE approach with an autoregressive within-person correlation structure	
gee_b00_int	Parameter estimate of group mean cognitive intercept (baseline level of cognitive function) for the unexposed ($\hat{\beta}_{00}$)
gee_b00_int_se	Standard error of parameter estimate of group mean cognitive intercept (baseline level of cognitive function) for the unexposed (SE($\hat{\beta}_{00}$))
gee_b01_exposure	Parameter estimate of effect of exposure on level of cognitive function at baseline ($\hat{\beta}_{01}$)
gee_b01_exposure_se	Standard error of parameter estimate of effect of exposure on level of cognitive function at baseline (SE($\hat{\beta}_{01}$))
gee_b02_age	Parameter estimate of effect of one year change in baseline age on level of cognitive function at baseline ($\hat{\beta}_{02}$)
gee_b02_age_se	Standard error of parameter estimate of effect of one year change in baseline age on level of cognitive function at baseline (SE($\hat{\beta}_{02}$))
gee_b10_time	Parameter estimate of group mean cognitive slope (annual rate of cognitive change) for the unexposed ($\hat{\beta}_{10}$)
gee_b10_time_se	Standard error of parameter estimate of group mean cognitive slope (annual rate of cognitive change) for the unexposed (SE($\hat{\beta}_{10}$))
gee_b11_exptime	Parameter estimate of effect of exposure on annual rate of cognitive change ($\hat{\beta}_{11}$)
gee_b11_exptime_se	Standard error of parameter estimate of effect of exposure on annual rate of cognitive change (SE($\hat{\beta}_{11}$))
gee_b12_agec60time	Parameter estimate of effect of one year change in baseline age on annual rate of cognitive change ($\hat{\beta}_{12}$)
gee_b12_agec60time_se	Standard error of parameter estimate of one year change in baseline age on annual rate of cognitive change (SE($\hat{\beta}_{12}$))
Parameter estimates from autoregressive linear mixed effects models	
ARmixed_b00_int	Parameter estimate of group mean cognitive intercept (baseline level of cognitive function) for the unexposed ($\hat{\beta}_{00}$)
ARmixed_b00_int_se	Standard error of parameter estimate of group mean cognitive intercept (baseline level of cognitive function) for the unexposed (SE($\hat{\beta}_{00}$))
ARmixed_b01_exposure	Parameter estimate of effect of exposure on level of cognitive function at baseline ($\hat{\beta}_{01}$)
ARmixed_b01_exposure_se	Standard error of parameter estimate of effect of exposure on level of cognitive function at baseline (SE($\hat{\beta}_{01}$))
ARmixed_b02_age	Parameter estimate of effect of one year change in baseline age on level of cognitive function at baseline ($\hat{\beta}_{02}$)
ARmixed_b02_age_se	Standard error of parameter estimate of effect of one year change in baseline age on level of cognitive function at baseline (SE($\hat{\beta}_{02}$))
ARmixed_b10_time	Parameter estimate of group mean cognitive slope (annual rate of cognitive change) for the unexposed ($\hat{\beta}_{10}$)
ARmixed_b10_time_se	Standard error of parameter estimate of group mean cognitive slope (annual rate of cognitive change) for the unexposed (SE($\hat{\beta}_{10}$))

ARmixed_b11_exptime	Parameter estimate of effect of exposure on annual rate of cognitive change ($\hat{\beta}_{11}$)
ARmixed_b11_exptime_se	Standard error of parameter estimate of effect of exposure on annual rate of cognitive change (SE($\hat{\beta}_{11}$))
ARmixed_b12_agec60time	Parameter estimate of effect of one year change in baseline age on annual rate of cognitive change ($\hat{\beta}_{12}$)
ARmixed_b12_agec60time_se	Standard error of parameter estimate of one year change in baseline age on annual rate of cognitive change (SE($\hat{\beta}_{12}$))

Web Table 3. Simulation master file codebook.

Variable	Definition
Variables read in from each replication	
N	Number of observations in each data set
int_time	Number of years between cognitive assessments in each data set
p_study_death	Proportion of study participants who died in each data set
survtime	Mean follow-up time for study participants in each data set
mixed_b00_int, mixed_b00_int_se, mixed_b10_time, mixed_b10_time_se, etc.	Parameter estimates and standard errors of parameter estimates from conventional linear mixed effects models from each data set
gee_b00_int, gee_b00_int_se, gee_b10_time, gee_b10_time_se, etc.	Parameter estimates and standard errors of parameter estimates from the GEE approach from each data set
ARmixed_b00_int, ARmixed_b00_int_se, ARmixed_b01_time, ARmixed_b01_time_se, etc.	Parameter estimates and standard errors of parameter estimates from autoregressive linear mixed effects models from each data set
Summary variables created in the simulation master file to summarize results across all replications	
mean_p_study_death	Mean proportion of study participants who died across data sets
mean_survtime	Mean follow-up time across data sets
Mean of each parameter estimate across replications	
mean_mixed_b00_int	Mean parameter estimate of group mean cognitive intercept (baseline level of cognitive function) for the unexposed ($\overline{\hat{\beta}}_{00}$)
mean_mixed_b10_time	Mean parameter estimate of group mean cognitive slope (annual rate of cognitive change) for the unexposed ($\overline{\hat{\beta}}_{10}$)
mean_mixed_b01_exposure	Mean parameter estimate of effect of exposure on level of cognitive function at baseline ($\overline{\hat{\beta}}_{01}$)
mean_mixed_b11_exptime	Mean parameter estimate of effect of exposure on annual rate of cognitive change ($\overline{\hat{\beta}}_{11}$)
mean_mixed_b02_agec60	Mean parameter estimate of effect of one year change in baseline age (in years, centered at age 60) on level of cognitive function at baseline ($\overline{\hat{\beta}}_{02}$)
mean_mixed_b12_agec60time	Mean parameter estimate of effect of one year change in baseline age on annual rate of cognitive change ($\overline{\hat{\beta}}_{12}$)
mean_gee_b00_int	Mean parameter estimate of group mean cognitive intercept (baseline level of cognitive function) for the unexposed ($\overline{\hat{\beta}}_{00}$)
mean_gee_b10_time	Mean parameter estimate of group mean cognitive slope (annual rate of cognitive change) for the unexposed ($\overline{\hat{\beta}}_{10}$)
mean_gee_b01_exposure	Mean parameter estimate of effect of exposure on level of cognitive function at baseline ($\overline{\hat{\beta}}_{01}$)
mean_gee_b11_exptime	Mean parameter estimate of effect of exposure on annual rate of cognitive change ($\overline{\hat{\beta}}_{11}$)
mean_gee_b02_agec60	Mean parameter estimate of effect of one year change in baseline age (in years, centered at age 60) on level of cognitive function at baseline ($\overline{\hat{\beta}}_{02}$)

mean_gee_b12_agec60time	Mean parameter estimate of effect of one year change in baseline age on annual rate of cognitive change ($\overline{\hat{\beta}}_{12}$)
mean_ARmixed_b00_int	Mean parameter estimate of group mean cognitive intercept (baseline level of cognitive function) for the unexposed ($\overline{\hat{\beta}}_{00}$)
mean_ARmixed_b10_time	Mean parameter estimate of group mean cognitive slope (annual rate of cognitive change) for the unexposed ($\overline{\hat{\beta}}_{10}$)
mean_ARmixed_b01_exposure	Mean parameter estimate of effect of exposure on level of cognitive function at baseline ($\overline{\hat{\beta}}_{01}$)
mean_ARmixed_b11_exptime	Mean parameter estimate of effect of exposure on annual rate of cognitive change ($\overline{\hat{\beta}}_{11}$)
mean_ARmixed_b02_agec60	Mean parameter estimate of effect of one year change in baseline age (in years, centered at age 60) on level of cognitive function at baseline ($\overline{\hat{\beta}}_{02}$)
mean_ARmixed_b12_agec60time	Mean parameter estimate of effect of one year change in baseline age on annual rate of cognitive change ($\overline{\hat{\beta}}_{12}$)
Empirical standard error of each parameter estimate across replications	
empSE_mixed_b00_int	Empirical standard error of parameter estimate of group mean cognitive intercept (baseline level of cognitive function) for the unexposed
empSE_mixed_b01_exposure	Empirical standard error of parameter estimate of group mean cognitive slope (annual rate of cognitive change) for the unexposed
empSE_mixed_b11_exptime	Empirical standard error of parameter estimate of effect of exposure on level of cognitive function at baseline
empSE_mixed_b02_agec60	Empirical standard error of parameter estimate of effect of exposure on annual rate of cognitive change
empSE_mixed_b12_agec60time	Empirical standard error of parameter estimate of effect of one year change in baseline age (in years, centered at age 60) on level of cognitive function at baseline
empSE_gee_b00_int	Empirical standard error of parameter estimate of effect of one year change in baseline age on annual rate of cognitive change
empSE_gee_b10_time	Empirical standard error of parameter estimate of group mean cognitive intercept (baseline level of cognitive function) for the unexposed
empSE_gee_b01_exposure	Empirical standard error of parameter estimate of group mean cognitive slope (annual rate of cognitive change) for the unexposed
empSE_gee_b11_exptime	Empirical standard error of parameter estimate of effect of exposure on level of cognitive function at baseline
empSE_gee_b02_agec60	Empirical standard error of parameter estimate of effect of exposure on annual rate of cognitive change
empSE_gee_b12_agec60time	Empirical standard error of parameter estimate of effect of one year change in baseline age (in years, centered at age 60) on level of cognitive function at baseline
empSE_ARmixed_b00_int	Empirical standard error of parameter estimate of effect of one year change in baseline age on annual rate of cognitive change
empSE_ARmixed_b10_time	Empirical standard error of parameter estimate of group mean cognitive intercept (baseline level of cognitive function) for the unexposed

empSE_ARmixed_b01_exposure	Empirical standard error of parameter estimate of group mean cognitive slope (annual rate of cognitive change) for the unexposed
empSE_ARmixed_b11_exptime	Empirical standard error of parameter estimate of effect of exposure on level of cognitive function at baseline
empSE_ARmixed_b02_agec60	Empirical standard error of parameter estimate of effect of exposure on annual rate of cognitive change
empSE_ARmixed_b12_agec60time	Empirical standard error of parameter estimate of effect of one year change in baseline age (in years, centered at age 60) on level of cognitive function at baseline
Root-mean-square error (RMSE) of each parameter estimate across replications	
RMSE_mixed_b00_int	RMSE of parameter estimate of group mean cognitive intercept (baseline level of cognitive function) for the unexposed
RMSE_mixed_b10_time	RMSE of parameter estimate of group mean cognitive slope (annual rate of cognitive change) for the unexposed
RMSE_mixed_b01_exposure	RMSE of parameter estimate of effect of exposure on level of cognitive function at baseline
RMSE_mixed_b11_exptime	RMSE of parameter estimate of effect of exposure on annual rate of cognitive change
RMSE_mixed_b02_agec60	Empirical standard error of parameter estimate of effect of one year change in baseline age (in years, centered at age 60) on level of cognitive function at baseline
RMSE_mixed_b12_agec60time	Empirical standard error of parameter estimate of effect of one year change in baseline age on annual rate of cognitive change
RMSE_gee_b00_int	RMSE of parameter estimate of group mean cognitive intercept (baseline level of cognitive function) for the unexposed
RMSE_gee_b10_time	RMSE of parameter estimate of group mean cognitive slope (annual rate of cognitive change) for the unexposed
RMSE_gee_b01_exposure	RMSE of parameter estimate of effect of exposure on level of cognitive function at baseline
RMSE_gee_b11_exptime	RMSE of parameter estimate of effect of exposure on annual rate of cognitive change
RMSE_gee_b02_agec60	Empirical standard error of parameter estimate of effect of one year change in baseline age (in years, centered at age 60) on level of cognitive function at baseline
RMSE_gee_b12_agec60time	Empirical standard error of parameter estimate of effect of one year change in baseline age on annual rate of cognitive change
RMSE_ARmixed_b00_int	RMSE of parameter estimate of group mean cognitive intercept (baseline level of cognitive function) for the unexposed
RMSE_ARmixed_b10_time	RMSE of parameter estimate of group mean cognitive slope (annual rate of cognitive change) for the unexposed
RMSE_ARmixed_b01_exposure	RMSE of parameter estimate of effect of exposure on level of cognitive function at baseline
RMSE_ARmixed_b11_exptime	RMSE of parameter estimate of effect of exposure on annual rate of cognitive change
RMSE_ARmixed_b02_agec60	Empirical standard error of parameter estimate of effect of one year change in baseline age (in years, centered at age 60) on level of cognitive function at baseline
RMSE_ARmixed_b12_agec60time	Empirical standard error of parameter estimate of effect of one year change in baseline age on annual rate of cognitive change
Lower bound and upper bound of 95% confidence intervals (CI) for parameter estimates and indicator variable for 95% confidence interval coverage	

lb_mixed_b00_int	Lower bound of 95% CI for parameter estimate of group mean cognitive intercept (baseline level of cognitive function) for the unexposed
ub_mixed_b00_int	Upper bound of 95% CI for parameter estimate of group mean cognitive intercept (baseline level of cognitive function) for the unexposed
coverage_mixed_b00_int	Indicator variable for 95% CI coverage for parameter estimate of group mean cognitive intercept (baseline level of cognitive function) for the unexposed
lb_mixed_b10_time	Lower bound of 95% CI for parameter estimate of group mean cognitive slope (annual rate of cognitive change) for the unexposed
ub_mixed_b10_time	Upper bound of 95% CI for parameter estimate of group mean cognitive slope (annual rate of cognitive change) for the unexposed
coverage_mixed_b10_time	Indicator variable for 95% CI coverage for parameter estimate of group mean cognitive slope (annual rate of cognitive change) for the unexposed
lb_mixed_b01_exposure	Lower bound of 95% CI for parameter estimate of effect of exposure on level of cognitive function at baseline
ub_mixed_b01_exposure	Upper bound of 95% CI for parameter estimate of effect of exposure on level of cognitive function at baseline
coverage_mixed_b01_exposure	Indicator variable for 95% CI coverage for parameter estimate of effect of exposure on level of cognitive function at baseline
lb_mixed_b11_exptime	Lower bound of 95% CI for parameter estimate of effect of exposure on annual rate of cognitive change
ub_mixed_b11_exptime	Upper bound of 95% CI for parameter estimate of effect of exposure on annual rate of cognitive change
coverage_mixed_b11_exptime	Indicator variable for 95% CI coverage for parameter estimate of effect of exposure on annual rate of cognitive change
lb_mixed_b02_agec60	Lower bound of 95% CI for parameter estimate of effect of one year change in baseline age (in years, centered at age 60) on level of cognitive function at baseline
ub_mixed_b02_agec60	Upper bound of 95% CI for parameter estimate of effect of one year change in baseline age (in years, centered at age 60) on level of cognitive function at baseline
coverage_mixed_b02_agec60	Indicator variable for 95% CI coverage for parameter estimate of effect of one year change in baseline age (in years, centered at age 60) on level of cognitive function at baseline
lb_mixed_b12_agec60time	Lower bound of 95% CI for parameter estimate of effect of one year change in baseline age on annual rate of cognitive change
ub_mixed_b12_agec60time	Upper bound of 95% CI for parameter estimate of effect of one year change in baseline age on annual rate of cognitive change
coverage_mixed_b12_agec60time	Indicator variable for 95% CI coverage for parameter estimate of effect of one year change in baseline age on annual rate of cognitive change
lb_gee_b00_int	Lower bound of 95% CI for parameter estimate of group mean cognitive intercept (baseline level of cognitive function) for the unexposed
ub_gee_b00_int	Upper bound of 95% CI for parameter estimate of group mean cognitive intercept (baseline level of cognitive function) for the unexposed
coverage_gee_b00_int	Indicator variable for 95% CI coverage for parameter estimate of group mean cognitive intercept (baseline level of cognitive function) for the unexposed

lb_gee_b10_time	Lower bound of 95% CI for parameter estimate of group mean cognitive slope (annual rate of cognitive change) for the unexposed
ub_gee_b10_time	Upper bound of 95% CI for parameter estimate of group mean cognitive slope (annual rate of cognitive change) for the unexposed
coverage_gee_b10_time	Indicator variable for 95% CI coverage for parameter estimate of group mean cognitive slope (annual rate of cognitive change) for the unexposed
lb_gee_b01_exposure	Lower bound of 95% CI for parameter estimate of effect of exposure on level of cognitive function at baseline
ub_gee_b01_exposure	Upper bound of 95% CI for parameter estimate of effect of exposure on level of cognitive function at baseline
coverage_gee_b01_exposure	Indicator variable for 95% CI coverage for parameter estimate of effect of exposure on level of cognitive function at baseline
lb_gee_b11_exptime	Lower bound of 95% CI for parameter estimate of effect of exposure on annual rate of cognitive change
ub_gee_b11_exptime	Upper bound of 95% CI for parameter estimate of effect of exposure on annual rate of cognitive change
coverage_gee_b11_exptime	Indicator variable for 95% CI coverage for parameter estimate of effect of exposure on annual rate of cognitive change
lb_gee_b02_agec60	Lower bound of 95% CI for parameter estimate of effect of one year change in baseline age (in years, centered at age 60) on level of cognitive function at baseline
ub_gee_b02_agec60	Upper bound of 95% CI for parameter estimate of effect of one year change in baseline age (in years, centered at age 60) on level of cognitive function at baseline
coverage_gee_b02_agec60	Indicator variable for 95% CI coverage for parameter estimate of effect of one year change in baseline age (in years, centered at age 60) on level of cognitive function at baseline
lb_gee_b12_agec60time	Lower bound of 95% CI for parameter estimate of effect of one year change in baseline age on annual rate of cognitive change
ub_gee_b12_agec60time	Upper bound of 95% CI for parameter estimate of effect of one year change in baseline age on annual rate of cognitive change
coverage_gee_b12_agec60time	Indicator variable for 95% CI coverage for parameter estimate of effect of one year change in baseline age on annual rate of cognitive change
lb_ARmixed_b00_int	Lower bound of 95% CI for parameter estimate of group mean cognitive intercept (baseline level of cognitive function) for the unexposed
ub_ARmixed_b00_int	Upper bound of 95% CI for parameter estimate of group mean cognitive intercept (baseline level of cognitive function) for the unexposed
coverage_ARmixed_b00_int	Indicator variable for 95% CI coverage for parameter estimate of group mean cognitive intercept (baseline level of cognitive function) for the unexposed
lb_ARmixed_b10_time	Lower bound of 95% CI for parameter estimate of group mean cognitive slope (annual rate of cognitive change) for the unexposed
ub_ARmixed_b10_time	Upper bound of 95% CI for parameter estimate of group mean cognitive slope (annual rate of cognitive change) for the unexposed
coverage_ARmixed_b10_time	Indicator variable for 95% CI coverage for parameter estimate of group mean cognitive slope (annual rate of cognitive change) for the unexposed
lb_ARmixed_b01_exposure	Lower bound of 95% CI for parameter estimate of effect of exposure on level of cognitive function at baseline

ub_ARmixed_b01_exposure	Upper bound of 95% CI for parameter estimate of effect of exposure on level of cognitive function at baseline
coverage_ARmixed_b01_exposure	Indicator variable for 95% CI coverage for parameter estimate of effect of exposure on level of cognitive function at baseline
lb_ARmixed_b11_exptime	Lower bound of 95% CI for parameter estimate of effect of exposure on annual rate of cognitive change
ub_ARmixed_b11_exptime	Upper bound of 95% CI for parameter estimate of effect of exposure on annual rate of cognitive change
coverage_ARmixed_b11_exptime	Indicator variable for 95% CI coverage for parameter estimate of effect of exposure on annual rate of cognitive change
lb_ARmixed_b02_agec60	Lower bound of 95% CI for parameter estimate of effect of one year change in baseline age (in years, centered at age 60) on level of cognitive function at baseline
ub_ARmixed_b02_agec60	Upper bound of 95% CI for parameter estimate of effect of one year change in baseline age (in years, centered at age 60) on level of cognitive function at baseline
coverage_ARmixed_b02_agec60	Indicator variable for 95% CI coverage for parameter estimate of effect of one year change in baseline age (in years, centered at age 60) on level of cognitive function at baseline
lb_ARmixed_b12_agec60time	Lower bound of 95% CI for parameter estimate of effect of one year change in baseline age on annual rate of cognitive change
ub_ARmixed_b12_agec60time	Upper bound of 95% CI for parameter estimate of effect of one year change in baseline age on annual rate of cognitive change
coverage_ARmixed_b12_agec60time	Indicator variable for 95% CI coverage for parameter estimate of effect of one year change in baseline age on annual rate of cognitive change

Web Appendix 2. Data generation and analysis file.

```
/*************************************************************************/
/***   Simulation Scenario A: No anticipated bias                      ***/
/***   25% cumulative incidence of mortality                           ***/
/***   Modify the local parameters to modify the scenario              ***/
/*************************************************************************/

set more off
clear

/*Create blank data set*/
set obs 1500 //creates blank dataset with 1,500 observations
gen id = _n  //creates id numbers


/*Step 1: Set parameters*/

*i. Specify time (in years) between cognitive assessments
local int_time = 1.5

*ii. Specify prevalence of binary exposure (exp)
local pexp = 0.5

*iii. Variances and correlations
local s2z0 = 0.2    //variance of random cognitive intercept
local s2z1 = 0.005  //variance of random cognitive slope
local sz01 = 0.01   //covariance of random intercept and random slope
local s2e = 0.70    //variance of unexplained variation in Cij
local rho = 0.40    //correlation between noise terms for Cij
local s2d = 0.19    //variance of measurement error of Cij
/*Notes for modifications:
   1. To eliminate random intercept and random slope terms, set
      sigma_square_zeta0(s2z0)=0, sigma_square_zeta1(s2z1)=0,
      and sigma_zeta_01(sz01)=0.
   2. To eliminate autoregressive covariance structure, set rho=0.
   3. To eliminate measurement error, set sigma_square_delta(s2d)=0.*/

*iv. Parameters for cognitive function (Cij)
local b00 = 0       //cognitive intercept for unexposed
local b01 = 0       //effect of exposure on cognitive intercept
local b02 = -0.05   //effect of age on cognitive intercept
local b03 = 0       //effect of U on cognitive intercept
local b10 = -0.05   //cognitive slope for unexposed
local b11 = 0       //effect of exposure on cognitive slope
local b12 = -0.005  //effect of age on cognitive slope
local b13 = -0.05   //effect of U on cognitive slope

*v. Parameters for survival time (Sij)
local g1 = 0.69     //effect of exposure on log hazard of death
local g2 = 0.086    //effect of TD age on log hazard of death
local g3 = 0        //effect of U on log hazard of death
local g4 = 0        //interaction effect of exposure & U on log hazard of death
local g5 = 0        //effect of TD cognitive slope on log hazard of death
local g6 = 0        //effect of TD cognitive function on log hazard of death

*vi. Baseline hazard of death
local lambda = 0.004315 //25% die over study period
*local lambda = 0.01173 //50% die over study period
*local lambda = 0.02763 //75% die over study period
```

```
/*Step 2: Generate variable for baseline age in years.
Have min age 60 and max age ~100. Similar to most cohort studies that enroll
participants age 60+, we want the age distribution to be right-skewed
(less older people). We achieved this by using a beta distribution*/
gen age0 = 60 + rbeta(2,4)*40
gen age0_c60 = age0 - 60 /*age centered at age 60 for modeling*/


/*Step 3: Generate exposure variable with prevalence pexp by generating a U[0,1)
distribution and setting exposure=0 if the random number<pexp, else exposure=1*/
gen exposure = runiform()<`pexp'


/*Step 4: Generate U, a continuous time-constant variable, U~N(0,1)*/
gen U = rnormal()


/*Step 5: Create visit times and ages for each cognitive assessment. The study
will include 7 cognitive assessment waves "int_time" years apart*/
*Generate visit times
gen time0 = 0
gen time1 = `int_time'
gen time2 = time1 + `int_time'
gen time3 = time2 + `int_time'
gen time4 = time3 + `int_time'
gen time5 = time4 + `int_time'
gen time6 = time5 + `int_time'

*Generate age at each visit
*age0 already generated
gen age1 = age0 + time1
gen age2 = age0 + time2
gen age3 = age0 + time3
gen age4 = age0 + time4
gen age5 = age0 + time5
gen age6 = age0 + time6
*Generate age (centered at 60 years) at each visit
*age0_c60 already generated
gen age1_c60 = age0_c60 + time1
gen age2_c60 = age0_c60 + time2
gen age3_c60 = age0_c60 + time3
gen age4_c60 = age0_c60 + time4
gen age5_c60 = age0_c60 + time5
gen age6_c60 = age0_c60 + time6


/*Step 6: Generate "true" and "measured" cognitive function at each wave,
where measured cognitive function = true cognitive function + error*/
/*Cij = b00 + b01*exposurei + b02*age0i + b03*Ui
      + (b10 + b11*exposurei + b12*age0i + b13*Ui)*timeij
      + z0i +z1i*timeij + epsilonij*/

/*Generate random terms for slope and intercept, zeta_0i (z0i) and zeta_1i (z1i),
where zeta_0i and zeta_1i covary*/
matrix c = (`s2z0', `sz01'\ `sz01', `s2z1')
drawnorm z0i z1i, cov(c)

/*Generate autoregressive noise term (unexplained variance in Cij) for each visit*/
gen sd_alpha = sqrt((1-`rho'*`rho'))*`s2e'

gen alpha1 = sd_alpha*rnormal()
gen alpha2 = sd_alpha*rnormal()
gen alpha3 = sd_alpha*rnormal()
```

```
gen alpha4 = sd_alpha*rnormal()
gen alpha5 = sd_alpha*rnormal()
gen alpha6 = sd_alpha*rnormal()

gen epsilon0 = sqrt(`s2e')*rnormal()
gen epsilon1 = `rho'*epsilon0 + alpha1
gen epsilon2 = `rho'*epsilon1 + alpha2
gen epsilon3 = `rho'*epsilon2 + alpha3
gen epsilon4 = `rho'*epsilon3 + alpha4
gen epsilon5 = `rho'*epsilon4 + alpha5
gen epsilon6 = `rho'*epsilon5 + alpha6

/*Generate true cognitive function*/
gen true_cogfxn0 = `b00' +`b01'*exposure +`b02'*age0_c60 +`b03'*U ///
                   +(`b10' +`b11'*exposure +`b12'*age0_c60 +`b13'*U)*0 ///
                   +z0i +z1i*0 +epsilon0
gen true_cogfxn1 = `b00' +`b01'*exposure +`b02'*age0_c60 +`b03'*U ///
                   +(`b10' +`b11'*exposure +`b12'*age0_c60 +`b13'*U)*time1 ///
                   +z0i +z1i*time1 +epsilon1
gen true_cogfxn2 = `b00' +`b01'*exposure +`b02'*age0_c60 +`b03'*U ///
                   +(`b10' +`b11'*exposure+`b12'*age0_c60 +`b13'*U)*time2 ///
                   +z0i +z1i*time2 +epsilon2
gen true_cogfxn3 = `b00' +`b01'*exposure +`b02'*age0_c60 +`b03'*U ///
                   +(`b10' +`b11'*exposure +`b12'*age0_c60 +`b13'*U)*time3 ///
                   +z0i +z1i*time3 +epsilon3
gen true_cogfxn4 = `b00' +`b01'*exposure +`b02'*age0_c60 +`b03'*U ///
                   +(`b10' +`b11'*exposure +`b12'*age0_c60 +`b13'*U)*time4 ///
                   +z0i +z1i*time4 +epsilon4
gen true_cogfxn5 = `b00' +`b01'*exposure +`b02'*age0_c60 +`b03'*U ///
                   +(`b10' +`b11'*exposure +`b12'*age0_c60 +`b13'*U)*time5 ///
                   +z0i +z1i*time5 +epsilon5
gen true_cogfxn6 = `b00' +`b01'*exposure +`b02'*age0_c60 +`b03'*U ///
                   +(`b10' +`b11'*exposure +`b12'*age0_c60 +`b13'*U)*time6 ///
                   +z0i +z1i*time6 +epsilon6


/*Generate measured cognitive function (true cognitive function measured with error)*/
*C_ij = C_ij + error_ij
*First, generate delta term (delta_ij = measurement error for Cij) for each visit
gen delta0 = sqrt(`s2d')*rnormal()
gen delta1 = sqrt(`s2d')*rnormal()
gen delta2 = sqrt(`s2d')*rnormal()
gen delta3 = sqrt(`s2d')*rnormal()
gen delta4 = sqrt(`s2d')*rnormal()
gen delta5 = sqrt(`s2d')*rnormal()
gen delta6 = sqrt(`s2d')*rnormal()

*Next, generate measured cognitive function as C_ij* = C_ij + delta_ij
gen measured_cogfxn0 = true_cogfxn0 +delta0
gen measured_cogfxn1 = true_cogfxn1 +delta1
gen measured_cogfxn2 = true_cogfxn2 +delta2
gen measured_cogfxn3 = true_cogfxn3 +delta3
gen measured_cogfxn4 = true_cogfxn4 +delta4
gen measured_cogfxn5 = true_cogfxn5 +delta5
gen measured_cogfxn6 = true_cogfxn6 +delta6


/*Step 7: Calculate true rate of change in cognitive function (true slope) for
each interval*/
gen avg_slope = (`b10' +`b11'*exposure +`b12'*age0_c60 +`b13'*U +z1i)
gen slope01 = avg_slope +(epsilon1-epsilon0)/`int_time'
gen slope12 = avg_slope +(epsilon2-epsilon1)/`int_time'
gen slope23 = avg_slope +(epsilon3-epsilon2)/`int_time'
```

```
gen slope34 = avg_slope +(epsilon4-epsilon3)/`int_time'
gen slope45 = avg_slope +(epsilon5-epsilon4)/`int_time'
gen slope56 = avg_slope +(epsilon6-epsilon5)/`int_time'


/*Step 8: Generate survival time for each person. At each cognitive assessment
wave j, a "time to death" value is generated for each person who has remained
alive up to that wave. The time to death variable is generated based on the
person's covariate history up to that point as a random variable drawn from an
exponential survival distribution (based on the hazard function specified below).
If the randomly generated time to death exceeds the length of the interval
between waves j and j+1, the person is considered alive at wave j+1 and a new
survival time is generated for the next interval conditional on history up to
the start of the interval. This process is repeated until the person's survival
time falls within a given interval or the end of the study (7 waves), whichever
comes first.
h(tij|x) = lambda*exp(g1*exposurei + g2*ageij + g3*Ui + g4*exposurei*Ui +
                    g5*slopeij + g6Cij)
A person's survival time for a given time interval at risk is generated using
the inverse cumulative hazard function transformation formula described by
Bender et al. (Stat Med 2011)*/

*i. Generate uniform random variable for generating survival times
gen U_01 = runiform()
gen U_12 = runiform()
gen U_23 = runiform()
gen U_34 = runiform()
gen U_45 = runiform()
gen U_56 = runiform()


*ii. Generate survival time for each interval
/*Interval 0-1*/
*Generate survival time from time 0
gen survtime01 = -ln(U_01)/(`lambda'*exp(`g1'*exposure +`g2'*age0_c60 +`g3'*U ///
                    +`g4'*exposure*U +`g5'*slope01 +`g6'*true_cogfxn0))
*Generate death indicator for interval 0-1
gen death01 = 0
replace death01 = 1 if (survtime01<`int_time')

/*Interval 1-2*/
*Generate survival time from time 1
gen survtime12 = -ln(U_12)/(`lambda'*exp(`g1'*exposure +`g2'*age1_c60 +`g3'*U ///
                          +`g4'*exposure*U +`g5'*slope12 +`g6'*true_cogfxn1)) ///
                          if (death01==0)
*Generate death indicator for interval 1-2
gen death12 = 0 if (death01==0)
replace death12 = 1 if (survtime12<`int_time')

/*Interval 2-3*/
*Generate survival time from time 2
gen survtime23 = -ln(U_23)/(`lambda'*exp(`g1'*exposure +`g2'*age2_c60 +`g3'*U ///
                          +`g4'*exposure*U +`g5'*slope23 +`g6'*true_cogfxn2)) ///
                          if (death01==0 & death12==0)
*Generate death indicator for interval 2-3
gen death23 = 0 if (death01==0 & death12==0)
replace death23 = 1 if (survtime23<`int_time')

/*Interval 3-4*/
*Generate survival time from time 3
gen survtime34 = -ln(U_34)/(`lambda'*exp(`g1'*exposure +`g2'*age3_c60 +`g3'*U ///
                          +`g4'*exposure*U +`g5'*slope34 +`g6'*true_cogfxn3)) ///
                          if (death01==0 & death12==0 & death23==0)
```

```
*Generate death indicator for interval 3-4
gen death34 = 0 if (death01==0 & death12==0 & death23==0)
replace death34 = 1 if (survtime34<`int_time')

/*Interval 4-5*/
*Generate survival time from time 4
gen survtime45 = -ln(U_45)/(`lambda'*exp(`g1'*exposure +`g2'*age4_c60 +`g3'*U ///
                          +`g4'*exposure*U +`g5'*slope45 +`g6'*true_cogfxn4)) ///
                          if (death01==0 & death12==0 & death23==0 & death34==0)
*Generate death indicator for interval 4-5
gen death45 = 0 if (death01==0 & death12==0 & death23==0 & death34==0)
replace death45 = 1 if (survtime45<`int_time')

/*Interval 5-6*/
*Generate survival time from time 5
gen survtime56 = -ln(U_56)/(`lambda'*exp(`g1'*exposure +`g2'*age5_c60 +`g3'*U ///
                          +`g4'*exposure*U +`g5'*slope56 +`g6'*true_cogfxn5)) ///
                          if (death01==0 & death12==0 & death23==0 & death34==0 &
death45==0)
*Generate death indicator for interval 5-6
gen death56 = 0 if (death01==0 & death12==0 & death23==0 & death34==0 & death45==0)
replace death56 = 1 if (survtime56<`int_time')


*iv. Generate variable for death during study:
gen study_death = 0
replace study_death = 1 if (death01==1 | death12==1 | death23==1 | death34==1 ///
             | death45==1 | death56==1)


*v. Generate variable for overall survival time for study:
gen survtime = .
*survtime for people who died during the study:
replace survtime = survtime01 if (death01==1)
replace survtime = time1 + survtime12 if (death12==1)
replace survtime = time2 + survtime23 if (death23==1)
replace survtime = time3 + survtime34 if (death34==1)
replace survtime = time4 + survtime45 if (death45==1)
replace survtime = time5 + survtime56 if (death56==1)
*survtime for people who didn't die during study--censor at final study visit:
replace survtime = time6 if (study_death==0)

*vi. Generate variable for age at death:
gen agedeath = age0+survtime


/*Step 9: Censor cognitive function values after death*/
replace time1 = . if (death01==1)
replace time2 = . if (death01==1 | death12==1)
replace time3 = . if (death01==1 | death12==1 | death23==1)
replace time4 = . if (death01==1 | death12==1 | death23==1 | death34==1)
replace time5 = . if (death01==1 | death12==1 | death23==1 | death34==1 | death45==1)
replace time6 = . if (death01==1 | death12==1 | death23==1 | death34==1 | death45==1 |
death56==1)

replace true_cogfxn1 = . if (death01==1)
replace true_cogfxn2 = . if (death01==1 | death12==1)
replace true_cogfxn3 = . if (death01==1 | death12==1 | death23==1)
replace true_cogfxn4 = . if (death01==1 | death12==1 | death23==1 | death34==1)
replace true_cogfxn5 = . if (death01==1 | death12==1 | death23==1 | death34==1 |
death45==1)
replace true_cogfxn6 = . if (death01==1 | death12==1 | death23==1 | death34==1 |
death45==1 | death56==1)
```

```
replace measured_cogfxn1 = . if (death01==1)
replace measured_cogfxn2 = . if (death01==1 | death12==1)
replace measured_cogfxn3 = . if (death01==1 | death12==1 | death23==1)
replace measured_cogfxn4 = . if (death01==1 | death12==1 | death23==1 | death34==1)
replace measured_cogfxn5 = . if (death01==1 | death12==1 | death23==1 | death34==1 |
death45==1)
replace measured_cogfxn6 = . if (death01==1 | death12==1 | death23==1 | death34==1 |
death45==1 | death56==1)

/*****************************************************************************/
/***   End data generation                                                 ***/
/*****************************************************************************/


/*****************************************************************************/
/***   MODELS                                                              ***/
/*****************************************************************************/

/*****************************************/
*pull N
scalar N= _N
/*****************************************/

/*****************************************/
*pull int time
scalar int_time = `int_time'
/*****************************************/

/*****************************************/
*pull percentage of deaths
summarize study_death, meanonly
scalar p_study_death = r(mean)
summarize survtime, meanonly
scalar survtime = r(mean)
/*****************************************/


/*Data generating model for cognitive function:
Cij = b00 + b01*exposurei + b02*age0i + b03*Ui + (b10 + b11*exposurei +
      + b12*age0i + b13*Ui)*timeij + z0i + z1i*timeij + epsilonij*/


/*****************************************/
/*Conventional linear mixed effects model (linear mixed effects models with
random intercepts and slopes allowing for possible correlation of the random
intercept and slope and no additional within-person covariance structure)*/
reshape long measured_cogfxn time, i(id) j(year)
xtmixed measured_cogfxn i.exposure##c.time c.age0_c60##c.time|| id: time, mle cov(uns)
*store fixed effect estimates as scalars;
matrix var=e(V)
*matrix list var
scalar mixed_b00_int = _b[_cons]
scalar mixed_b00_int_se = _se[_cons]
scalar mixed_b01_exposure = _b[1.exposure]
scalar mixed_b01_exposure_se = _se[1.exposure]
scalar mixed_b02_agec60 = _b[age0_c60]
scalar mixed_b02_agec60_se = _se[age0_c60]
scalar mixed_b10_time = _b[time]
scalar mixed_b10_time_se = _se[time]
scalar mixed_b11_exptime = _b[1.exposure#c.time]
scalar mixed_b11_exptime_se = _se[1.exposure#c.time]
scalar mixed_b12_agec60time = _b[age0_c60#c.time]
```

```
scalar mixed_b12_agec60time_se = _se[age0_c60#c.time]
/*****************************************/
*store random effect estimates as scalars;
matrix var=e(b)
_diparm lns1_1_2, f(exp(@)) d(exp(@))
scalar mixed_z0i_est = r(est)
scalar mixed_z0i_se = r(se)
_diparm lns1_1_1, f(exp(@)) d(exp(@))
scalar mixed_z1i_est = r(est)
scalar mixed_z1i_se = r(se)
_diparm atr1_1_1_2, f(tanh(@)) d(1-tanh(@)^2)
scalar mixed_rho_est = r(est)
scalar mixed_rho_se = r(se)
/*****************************************/


/*****************************************/
/*GEE approach*/
*reshape long measured_cogfxn time, i(id) j(year)
xtset id year
xtgee measured_cogfxn i.exposure##c.time c.age0_c60##c.time, i(id) corr(ar1) //|| id:
time, mle cov(uns)
*store estimates as scalars
matrix var=e(V)
*matrix list var
scalar gee_b00_int = _b[_cons]

scalar gee_b00_int_se = _se[_cons]
scalar gee_b01_exposure = _b[1.exposure]
scalar gee_b01_exposure_se = _se[1.exposure]
scalar gee_b10_time = _b[time]
scalar gee_b10_time_se = _se[time]
scalar gee_b11_exptime = _b[1.exposure#c.time]
scalar gee_b11_exptime_se = _se[1.exposure#c.time]
scalar gee_b02_agec60 = _b[age0_c60]
scalar gee_b02_agec60_se = _se[age0_c60]
scalar gee_b12_agec60time = _b[age0_c60#c.time]
scalar gee_b12_agec60time_se = _se[age0_c60#c.time]
/*****************************************/


/*****************************************/
/*Autoregressive linear mixed effects model
(autoregressive within-person covariance structure)*/
*reshape long measured_cogfxn time, i(id) j(year) //reshape with censoring at death
xtmixed measured_cogfxn i.exposure##c.time c.age0_c60##c.time|| id: time, mle cov(uns)
residuals(ar1, t(year))
*store fixed effect estimates as scalars;
matrix var=e(V)
*matrix list var
scalar ARmixed_b00_int = _b[_cons]
scalar ARmixed_b00_int_se = _se[_cons]
scalar ARmixed_b01_exposure = _b[1.exposure]
scalar ARmixed_b01_exposure_se = _se[1.exposure]
scalar ARmixed_b02_agec60 = _b[age0_c60]
scalar ARmixed_b02_agec60_se = _se[age0_c60]
scalar ARmixed_b10_time = _b[time]
scalar ARmixed_b10_time_se = _se[time]
scalar ARmixed_b11_exptime = _b[1.exposure#c.time]
scalar ARmixed_b11_exptime_se = _se[1.exposure#c.time]
scalar ARmixed_b12_agec60time = _b[age0_c60#c.time]
scalar ARmixed_b12_agec60time_se = _se[age0_c60#c.time]
/*****************************************/
```

Web Appendix 3. Simulation master file.

```
/*******************************************************************************/
/***   Simulation Master File                                                ***/

/*"simulate" performs Monte Carlo-type simulations.

The basic form of the code is:
"simulate <list_of_scalars_collected_from_each_repetition>, ///
reps(B) seed(#): do <data_generation_and_analysis_do_file>"

This will replicate B data sets using the data generating code and perform the
analyses in the do file. Then it will pull the results from each replication we
stored as scalars so we can summarize results across all B repetitions.
-Note that the first lines of code call in and assign the value from each
 scalar to a variable with the same name.
-"reps(B)" specifies the number of replications.
-"seed(#)" specifies the random number seed.

Simulate documentation: http://www.stata.com/manuals13/rsimulate.pdf

The rest of the code in this do file summarizes the results from the B
replications and stores the summarized results in an Excel file.*/
/*******************************************************************************/

cd "C:\Users\location_of_data_generation_and_analysis_do_file"

set more off
clear all

local B = 1000 //number of replications

simulate ///
N=N int_time=int_time p_study_death=p_study_death survtime=survtime ///
mixed_b00_int=mixed_b00_int mixed_b00_int_se=mixed_b00_int_se ///
mixed_b10_time=mixed_b10_time mixed_b10_time_se=mixed_b10_time_se ///
mixed_b01_exposure=mixed_b01_exposure mixed_b01_exposure_se=mixed_b01_exposure_se ///
mixed_b11_exptime=mixed_b11_exptime mixed_b11_exptime_se=mixed_b11_exptime_se ///
mixed_b02_agec60=mixed_b02_agec60 mixed_b02_agec60_se=mixed_b02_agec60_se ///
mixed_b12_agec60time=mixed_b12_agec60time
mixed_b12_agec60time_se=mixed_b12_agec60time_se ///
gee_b00_int=gee_b00_int gee_b00_int_se=gee_b00_int_se  ///
gee_b10_time=gee_b10_time gee_b10_time_se=gee_b10_time_se ///
gee_b01_exposure=gee_b01_exposure gee_b01_exposure_se=gee_b01_exposure_se  ///
gee_b11_exptime=gee_b11_exptime gee_b11_exptime_se=gee_b11_exptime_se ///
gee_b02_agec60=gee_b02_agec60 gee_b02_agec60_se=gee_b02_agec60_se ///
gee_b12_agec60time=gee_b12_agec60time gee_b12_agec60time_se=gee_b12_agec60time_se ///
ARmixed_b00_int=ARmixed_b00_int ARmixed_b00_int_se=ARmixed_b00_int_se ///
ARmixed_b10_time=ARmixed_b10_time ARmixed_b10_time_se=ARmixed_b10_time_se ///
ARmixed_b01_exposure=ARmixed_b01_exposure
ARmixed_b01_exposure_se=ARmixed_b01_exposure_se ///
ARmixed_b11_exptime=ARmixed_b11_exptime ARmixed_b11_exptime_se=ARmixed_b11_exptime_se
///
ARmixed_b02_agec60=ARmixed_b02_agec60 ARmixed_b02_agec60_se=ARmixed_b02_agec60_se ///
ARmixed_b12_agec60time=ARmixed_b12_agec60time
ARmixed_b12_agec60time_se=ARmixed_b12_agec60time_se, ///
reps(`B') seed(67208113): do ScenarioA

/*Parameters for Cij for comparing parameter estimates with true effects. These
values must match with values in data generation file for accurate calculations*/
local b00 = 0       //cognitive intercept for unexposed
local b01 = 0       //effect of exposure on cognitive intercept
```

```
local b02 = -0.05   //effect of age on cognitive intercept
local b03 = 0       //effect of U on cognitive intercept
local b10 = -0.05   //cognitive slope for unexposed
local b11 = 0       //effect of exposure on cognitive slope
local b12 = -0.005  //effect of age on cognitive slope
local b13 = -0.05   //effect of U on cognitive slope

/*Number of people in cohort*/
local N = N
local int_time = int_time

/*Proportion died and average follow up time*/
summarize p_study_death, meanonly
scalar mean_p_study_death = r(mean)
scalar r_mean_p_study_death = round(mean_p_study_death,0.01)
local pctdied = r_mean_p_study_death*100
summarize survtime, meanonly
scalar mean_survtime  = r(mean)

/*Parameter estimates: Across all B repetitions, calculate mean value of each
parameter estimate from models of repeated measures of cognitive function*/
foreach b in mixed_b00_int mixed_b10_time mixed_b01_exposure mixed_b11_exptime
mixed_b02_agec60 mixed_b12_agec60time ///
gee_b00_int gee_b10_time gee_b01_exposure gee_b11_exptime gee_b02_agec60
gee_b12_agec60time ///
ARmixed_b00_int ARmixed_b10_time ARmixed_b01_exposure ARmixed_b11_exptime
ARmixed_b02_agec60 ARmixed_b12_agec60time {
summarize `b', meanonly
scalar mean_`b' = r(mean)
}

/*Standard errors (SE): Across all B repetitions, calculate the mean value of each SE
estimate from models for repeated measures of cognitive function*/
foreach b in mixed_b00_int_se mixed_b10_time_se mixed_b01_exposure_se
mixed_b11_exptime_se mixed_b02_agec60_se mixed_b12_agec60time_se ///
gee_b00_int_se gee_b10_time_se gee_b01_exposure_se gee_b11_exptime_se
gee_b02_agec60_se gee_b12_agec60time_se ///
ARmixed_b00_int_se ARmixed_b10_time_se ARmixed_b01_exposure_se ARmixed_b11_exptime_se
ARmixed_b02_agec60_se ARmixed_b12_agec60time_se {
qui summarize `b', meanonly
scalar mean_`b' = r(mean)
}

/*Root-mean-square error (RMSE): Across all B repetitions, calculate the mean value of
RMSE from models for repeated measures of cognitive function. This involves two
steps*/
/*Step 1: Calculate empirical SE (SD of estimates of interest from all repetitions)*/
foreach b in mixed_b00_int mixed_b10_time mixed_b01_exposure mixed_b11_exptime
mixed_b02_agec60 mixed_b12_agec60time ///
gee_b00_int gee_b10_time gee_b01_exposure gee_b11_exptime gee_b02_agec60
gee_b12_agec60time ///
ARmixed_b00_int ARmixed_b10_time ARmixed_b01_exposure ARmixed_b11_exptime
ARmixed_b02_agec60 ARmixed_b12_agec60time {
summarize `b'
scalar empSE_`b' = r(sd)
}

/*Step 2: Calculate root-mean-square error (RMSE), which is the square root of the
mean squared deviation of the estimated effect of exposure on rate of cognitive
decline (beta_hat) from the true value (beta=0):
RMSE = sqrt[(mean_beta_hat - beta)^2 +(SE(beta_hat)^2]*/
scalar RMSE_mixed_b00_int = sqrt((mean_mixed_b00_int-`b00')*(mean_mixed_b00_int-
`b00')+empSE_mixed_b00_int*empSE_mixed_b00_int)
```

```
scalar RMSE_mixed_b10_time = sqrt((mean_mixed_b10_time-`b10')*(mean_mixed_b10_time-
`b10')+empSE_mixed_b10_time*empSE_mixed_b10_time)
scalar RMSE_mixed_b01_exposure = sqrt((mean_mixed_b01_exposure-
`b01')*(mean_mixed_b01_exposure-
`b01')+empSE_mixed_b01_exposure*empSE_mixed_b01_exposure)
scalar RMSE_mixed_b11_exptime = sqrt((mean_mixed_b11_exptime-
`b11')*(mean_mixed_b11_exptime-`b11')+empSE_mixed_b11_exptime*empSE_mixed_b11_exptime)
scalar RMSE_mixed_b02_agec60 = sqrt((mean_mixed_b02_agec60-
`b02')*(mean_mixed_b02_agec60-`b02')+empSE_mixed_b02_agec60*empSE_mixed_b02_agec60)
scalar RMSE_mixed_b12_agec60time = sqrt((mean_mixed_b12_agec60time-
`b12')*(mean_mixed_b12_agec60time-
`b12')+empSE_mixed_b12_agec60time*empSE_mixed_b12_agec60time)
scalar RMSE_gee_b00_int = sqrt((mean_gee_b00_int-`b00')*(mean_gee_b00_int-
`b00')+empSE_gee_b00_int*empSE_gee_b00_int)
scalar RMSE_gee_b10_time = sqrt((mean_gee_b10_time-`b10')*(mean_gee_b10_time-
`b10')+empSE_gee_b10_time*empSE_gee_b10_time)
scalar RMSE_gee_b01_exposure = sqrt((mean_gee_b01_exposure-
`b01')*(mean_gee_b01_exposure-`b01')+empSE_gee_b01_exposure*empSE_gee_b01_exposure)
scalar RMSE_gee_b11_exptime = sqrt((mean_gee_b11_exptime-`b11')*(mean_gee_b11_exptime-
`b11')+empSE_gee_b11_exptime*empSE_gee_b11_exptime)
scalar RMSE_gee_b02_agec60 = sqrt((mean_gee_b02_agec60-`b02')*(mean_gee_b02_agec60-
`b02')+empSE_gee_b02_agec60*empSE_gee_b02_agec60)
scalar RMSE_gee_b12_agec60time = sqrt((mean_gee_b12_agec60time-
`b12')*(mean_gee_b12_agec60time-
`b12')+empSE_gee_b12_agec60time*empSE_gee_b12_agec60time)
scalar RMSE_ARmixed_b00_int = sqrt((mean_ARmixed_b00_int-`b00')*(mean_ARmixed_b00_int-
`b00')+empSE_ARmixed_b00_int*empSE_ARmixed_b00_int)
scalar RMSE_ARmixed_b10_time = sqrt((mean_ARmixed_b10_time-
`b10')*(mean_ARmixed_b10_time-`b10')+empSE_ARmixed_b10_time*empSE_ARmixed_b10_time)
scalar RMSE_ARmixed_b01_exposure = sqrt((mean_ARmixed_b01_exposure-
`b01')*(mean_ARmixed_b01_exposure-
`b01')+empSE_ARmixed_b01_exposure*empSE_ARmixed_b01_exposure)
scalar RMSE_ARmixed_b11_exptime = sqrt((mean_ARmixed_b11_exptime-
`b11')*(mean_ARmixed_b11_exptime-
`b11')+empSE_ARmixed_b11_exptime*empSE_ARmixed_b11_exptime)
scalar RMSE_ARmixed_b02_agec60 = sqrt((mean_ARmixed_b02_agec60-
`b02')*(mean_ARmixed_b02_agec60-
`b02')+empSE_ARmixed_b02_agec60*empSE_ARmixed_b02_agec60)
scalar RMSE_ARmixed_b12_agec60time = sqrt((mean_ARmixed_b12_agec60time-
`b12')*(mean_ARmixed_b12_agec60time-
`b12')+empSE_ARmixed_b12_agec60time*empSE_ARmixed_b12_agec60time)

/*Generate indicator for 95% confidence interval (CI) coverage for each repetition:
Step 1 is to calculate the lower bound (lb) and upper bound (ub) of the 95% CI; Step
2: Generate an indicator variable for whether the 95% CI for a parameter estimate
(beta_hat) includes the true parameter value (beta).*/
*intercept (mixed_b00)
      gen lb_mixed_b00_int = mixed_b00_int - 1.96*mixed_b00_int_se
      gen ub_mixed_b00_int = mixed_b00_int + 1.96*mixed_b00_int_se
      gen coverage_mixed_b00_int = (ub_mixed_b00_int > `b00' & lb_mixed_b00_int <
`b00')
*time (mixed_b10)
      gen lb_mixed_b10_time = mixed_b10_time - 1.96*mixed_b10_time_se
      gen ub_mixed_b10_time = mixed_b10_time + 1.96*mixed_b10_time_se
      gen coverage_mixed_b10_time = (ub_mixed_b10_time > `b10' & lb_mixed_b10_time <
`b10')
*exposure(mixed_b01)
      gen lb_mixed_b01_exposure = mixed_b01_exposure - 1.96*mixed_b01_exposure_se
      gen ub_mixed_b01_exposure = mixed_b01_exposure + 1.96*mixed_b01_exposure_se
      gen coverage_mixed_b01_exposure = (ub_mixed_b01_exposure > `b01' &
lb_mixed_b01_exposure < `b01')
*exposure*time (mixed_b11)
      gen lb_mixed_b11_exptime = mixed_b11_exptime - 1.96*mixed_b11_exptime_se
```

```
      gen ub_mixed_b11_exptime = mixed_b11_exptime + 1.96*mixed_b11_exptime_se
      gen coverage_mixed_b11_exptime = (ub_mixed_b11_exptime > `b11' &
lb_mixed_b11_exptime < `b11')
*agec60 (mixed_b02)
      gen lb_mixed_b02_agec60 = mixed_b02_agec60 - 1.96*mixed_b02_agec60_se
      gen ub_mixed_b02_agec60 = mixed_b02_agec60 + 1.96*mixed_b02_agec60_se
      gen coverage_mixed_b02_agec60 = (ub_mixed_b02_agec60 > `b02' &
lb_mixed_b02_agec60 < `b02')
*agec60*time (mixed_b12)
      gen lb_mixed_b12_agec60time = mixed_b12_agec60time -
1.96*mixed_b12_agec60time_se
      gen ub_mixed_b12_agec60time = mixed_b12_agec60time +
1.96*mixed_b12_agec60time_se
      gen coverage_mixed_b12_agec60time = (ub_mixed_b12_agec60time > `b12' &
lb_mixed_b12_agec60time< `b12')

*intercept (gee_b00)
      gen lb_gee_b00_int = gee_b00_int - 1.96*gee_b00_int_se
      gen ub_gee_b00_int = gee_b00_int + 1.96*gee_b00_int_se
      gen coverage_gee_b00_int = (ub_gee_b00_int > `b00' & lb_gee_b00_int < `b00')
*time (gee_b10)
      gen lb_gee_b10_time = gee_b10_time - 1.96*gee_b10_time_se
      gen ub_gee_b10_time = gee_b10_time + 1.96*gee_b10_time_se
      gen coverage_gee_b10_time = (ub_gee_b10_time > `b10' & lb_gee_b10_time < `b10')
*exposure(gee_b01)
      gen lb_gee_b01_exposure = gee_b01_exposure - 1.96*gee_b01_exposure_se
      gen ub_gee_b01_exposure = gee_b01_exposure + 1.96*gee_b01_exposure_se
      gen coverage_gee_b01_exposure = (ub_gee_b01_exposure > `b01' &
lb_gee_b01_exposure < `b01')
*exposure*time (gee_b11)
      gen lb_gee_b11_exptime = gee_b11_exptime - 1.96*gee_b11_exptime_se
      gen ub_gee_b11_exptime = gee_b11_exptime + 1.96*gee_b11_exptime_se
      gen coverage_gee_b11_exptime = (ub_gee_b11_exptime > `b11' & lb_gee_b11_exptime
< `b11')
*agec60 (gee_b02)
      gen lb_gee_b02_agec60 = gee_b02_agec60 - 1.96*gee_b02_agec60_se
      gen ub_gee_b02_agec60 = gee_b02_agec60 + 1.96*gee_b02_agec60_se
      gen coverage_gee_b02_agec60 = (ub_gee_b02_agec60 > `b02' & lb_gee_b02_agec60 <
`b02')
*agec60*time (gee_b12)
      gen lb_gee_b12_agec60time = gee_b12_agec60time - 1.96*gee_b12_agec60time_se
      gen ub_gee_b12_agec60time = gee_b12_agec60time + 1.96*gee_b12_agec60time_se
      gen coverage_gee_b12_agec60time = (ub_gee_b12_agec60time > `b12' &
lb_gee_b12_agec60time< `b12')

*intercept (ARmixed_b00)
      gen lb_ARmixed_b00_int = ARmixed_b00_int - 1.96*ARmixed_b00_int_se
      gen ub_ARmixed_b00_int = ARmixed_b00_int + 1.96*ARmixed_b00_int_se
      gen coverage_ARmixed_b00_int = (ub_ARmixed_b00_int > `b00' & lb_ARmixed_b00_int
< `b00')
*time (ARmixed_b10)
      gen lb_ARmixed_b10_time = ARmixed_b10_time - 1.96*ARmixed_b10_time_se
      gen ub_ARmixed_b10_time = ARmixed_b10_time + 1.96*ARmixed_b10_time_se
      gen coverage_ARmixed_b10_time = (ub_ARmixed_b10_time > `b10' &
lb_ARmixed_b10_time < `b10')
*exposure(ARmixed_b01)
      gen lb_ARmixed_b01_exposure = ARmixed_b01_exposure -
1.96*ARmixed_b01_exposure_se
      gen ub_ARmixed_b01_exposure = ARmixed_b01_exposure +
1.96*ARmixed_b01_exposure_se
      gen coverage_ARmixed_b01_exposure = (ub_ARmixed_b01_exposure > `b01' &
lb_ARmixed_b01_exposure < `b01')
*exposure*time (ARmixed_b11)
```

```
      gen lb_ARmixed_b11_exptime = ARmixed_b11_exptime - 1.96*ARmixed_b11_exptime_se
      gen ub_ARmixed_b11_exptime = ARmixed_b11_exptime + 1.96*ARmixed_b11_exptime_se
      gen coverage_ARmixed_b11_exptime = (ub_ARmixed_b11_exptime > `b11' &
lb_ARmixed_b11_exptime < `b11')
*agec60 (ARmixed_b02)
      gen lb_ARmixed_b02_agec60 = ARmixed_b02_agec60 - 1.96*ARmixed_b02_agec60_se
      gen ub_ARmixed_b02_agec60 = ARmixed_b02_agec60 + 1.96*ARmixed_b02_agec60_se
      gen coverage_ARmixed_b02_agec60 = (ub_ARmixed_b02_agec60 > `b02' &
lb_ARmixed_b02_agec60 < `b02')
*agec60*time (ARmixed_b12)
      gen lb_ARmixed_b12_agec60time = ARmixed_b12_agec60time -
1.96*ARmixed_b12_agec60time_se
      gen ub_ARmixed_b12_agec60time = ARmixed_b12_agec60time +
1.96*ARmixed_b12_agec60time_se
      gen coverage_ARmixed_b12_agec60time = (ub_ARmixed_b12_agec60time > `b12' &
lb_ARmixed_b12_agec60time< `b12')

/*Create summary variable for 95% CI coverage for each parameter estimate*/
foreach b in coverage_mixed_b00_int coverage_mixed_b10_time
coverage_mixed_b01_exposure coverage_mixed_b11_exptime coverage_mixed_b02_agec60
coverage_mixed_b12_agec60time ///
coverage_gee_b00_int coverage_gee_b10_time coverage_gee_b01_exposure
coverage_gee_b11_exptime coverage_gee_b02_agec60 coverage_gee_b12_agec60time ///
coverage_ARmixed_b00_int coverage_ARmixed_b10_time coverage_ARmixed_b01_exposure
coverage_ARmixed_b11_exptime coverage_ARmixed_b02_agec60
coverage_ARmixed_b12_agec60time {
summarize `b', meanonly
scalar P`b' = r(mean)
}

/*Z test statistic and p-value for rejecting true value of each parameter*/
foreach b in mixed_b00_int gee_b00_int ARmixed_b00_int {
scalar Z_`b' = ((mean_`b' - `b00')/(empSE_`b'/sqrt(`B')))
scalar pvalue_`b'=2*(1-normal(abs(Z_`b')))
}
foreach b in mixed_b10_time gee_b10_time ARmixed_b10_time {
scalar Z_`b' = ((mean_`b' - `b10')/(empSE_`b'/sqrt(`B')))
scalar pvalue_`b'=2*(1-normal(abs(Z_`b')))
}
foreach b in mixed_b01_exposure gee_b01_exposure ARmixed_b01_exposure {
scalar Z_`b' = ((mean_`b' - `b01')/(empSE_`b'/sqrt(`B')))
scalar pvalue_`b'=2*(1-normal(abs(Z_`b')))
}
foreach b in mixed_b11_exptime gee_b11_exptime ARmixed_b11_exptime {
scalar Z_`b' = ((mean_`b' - `b11')/(empSE_`b'/sqrt(`B')))
scalar pvalue_`b'=2*(1-normal(abs(Z_`b')))
}
foreach b in mixed_b02_agec60 gee_b02_agec60 ARmixed_b02_agec60 {
scalar Z_`b' = ((mean_`b' - `b02')/(empSE_`b'/sqrt(`B')))
scalar pvalue_`b'=2*(1-normal(abs(Z_`b')))
}
foreach b in mixed_b12_agec60time gee_b12_agec60time ARmixed_b12_agec60time {
scalar Z_`b' = ((mean_`b' - `b12')/(empSE_`b'/sqrt(`B')))
scalar pvalue_`b'=2*(1-normal(abs(Z_`b')))
}

/*******************************************************************************/
/***   End of data steps   ***/
/*******************************************************************************/
```

```
/*************************************************************************/
/***   Final step is to store results in a Excel file       ***/
/*Note: "putexcel" was a new command in Stata 13. If you are running the
simulations in an earlier version of Stata, you can look at the results by
asking Stata to list the scalars with the summarized results. For example,
"scalar list mean_mixed_b11_exptime RMSE_mixed_b11_exptime Pcoverage_
mixed_b11_exptime survtime p_study_death" will output the mean beta estimate
for the effect of exposure on annual rate of cognitive decline and the
corresponding RMSE and 95% CI coverage estimates from the conventional linear
mixed effects model and the mean survival time and mean proportion of study
participants who died.*/
/*************************************************************************/

/***Conventional linear mixed effects model results; Store in sheet "MixedResults"***/
putexcel B1=("avg est beta") C1=("avg est SE") D1=("empirical SE") E1=("RMSE")
F1=("95% CI coverage") G1=("TS") H1=("p-value")  ///
A2=("Scenario A") A3=("Intercept") A4=("Time (years)") A5=("Exposure")
A6=("Exposure*Time") A7=("Age (years, c60)") A8=("Age*Time") ///
B3=(mean_mixed_b00_int) B4=(mean_mixed_b10_time) B5=(mean_mixed_b01_exposure)
B6=(mean_mixed_b11_exptime) B7=(mean_mixed_b02_agec60) B8=(mean_mixed_b12_agec60time)
///
C3=(mean_mixed_b00_int_se) C4=(mean_mixed_b10_time_se) C5=(mean_mixed_b01_exposure_se)
C6=(mean_mixed_b11_exptime_se) C7=(mean_mixed_b02_agec60_se)
C8=(mean_mixed_b12_agec60time_se) ///
D3=(empSE_mixed_b00_int) D4=(empSE_mixed_b10_time) D5=(empSE_mixed_b01_exposure)
D6=(empSE_mixed_b11_exptime) D7=(empSE_mixed_b02_agec60)
D8=(empSE_mixed_b12_agec60time) ///
E3=(RMSE_mixed_b00_int) E4=(RMSE_mixed_b10_time) E5=(RMSE_mixed_b01_exposure)
E6=(RMSE_mixed_b11_exptime) E7=(RMSE_mixed_b02_agec60) E8=(RMSE_mixed_b12_agec60time)
///
F3=(Pcoverage_mixed_b00_int) F4=(Pcoverage_mixed_b10_time)
F5=(Pcoverage_mixed_b01_exposure) F6=(Pcoverage_mixed_b11_exptime)
F7=(Pcoverage_mixed_b02_agec60) F8=(Pcoverage_mixed_b12_agec60time) ///
G3=(Z_mixed_b00_int) G4=(Z_mixed_b10_time) G5=(Z_mixed_b01_exposure)
G6=(Z_mixed_b11_exptime) G7=(Z_mixed_b02_agec60) G8=(Z_mixed_b12_agec60time) ///
H3=(pvalue_mixed_b00_int) H4=(pvalue_mixed_b10_time) H5=(pvalue_mixed_b01_exposure)
H6=(pvalue_mixed_b11_exptime) H7=(pvalue_mixed_b02_agec60)
H8=(pvalue_mixed_b12_agec60time) ///
using SimulationResults_N`N'_B`B'_pctdied`pctdied', sheet("MixedResults") modify


/***GEE approach results; Store in sheet "GEEResults"***/
putexcel B1=("avg est beta") C1=("avg est SE") D1=("empirical SE") E1=("RMSE")
F1=("95% CI coverage") G1=("TS") H1=("p-value")  ///
A2=("Scenario A") A3=("Intercept") A4=("Time (years)") A5=("Exposure")
A6=("Exposure*Time") A7=("Age (years, c60)") A8=("Age*Time") ///
B3=(mean_gee_b00_int) B4=(mean_gee_b10_time) B5=(mean_gee_b01_exposure)
B6=(mean_gee_b11_exptime) B7=(mean_gee_b02_agec60) B8=(mean_gee_b12_agec60time) ///
C3=(mean_gee_b00_int_se) C4=(mean_gee_b10_time_se) C5=(mean_gee_b01_exposure_se)
C6=(mean_gee_b11_exptime_se) C7=(mean_gee_b02_agec60_se)
C8=(mean_gee_b12_agec60time_se) ///
D3=(empSE_gee_b00_int) D4=(empSE_gee_b10_time) D5=(empSE_gee_b01_exposure)
D6=(empSE_gee_b11_exptime) D7=(empSE_gee_b02_agec60) D8=(empSE_gee_b12_agec60time) ///
E3=(RMSE_gee_b00_int) E4=(RMSE_gee_b10_time) E5=(RMSE_gee_b01_exposure)
E6=(RMSE_gee_b11_exptime) E7=(RMSE_gee_b02_agec60) E8=(RMSE_gee_b12_agec60time) ///
F3=(Pcoverage_gee_b00_int) F4=(Pcoverage_gee_b10_time) F5=(Pcoverage_gee_b01_exposure)
F6=(Pcoverage_gee_b11_exptime) F7=(Pcoverage_gee_b02_agec60)
F8=(Pcoverage_gee_b12_agec60time) ///
G3=(Z_gee_b00_int) G4=(Z_gee_b10_time) G5=(Z_gee_b01_exposure) G6=(Z_gee_b11_exptime)
G7=(Z_gee_b02_agec60) G8=(Z_gee_b12_agec60time) ///
H3=(pvalue_gee_b00_int) H4=(pvalue_gee_b10_time) H5=(pvalue_gee_b01_exposure)
H6=(pvalue_gee_b11_exptime) H7=(pvalue_gee_b02_agec60) H8=(pvalue_gee_b12_agec60time)
///
```

```
using SimulationResults_N`N'_B`B'_pctdied`pctdied', sheet("GEEResults") modify

/***Autoregressive linear mixed effects model; Store in sheet named
"ARmixedResults"***/
putexcel B1=("avg est beta") C1=("avg est SE") D1=("empirical SE") E1=("RMSE")
F1=("95% CI coverage") G1=("TS") H1=("p-value")  ///
A2=("Scenario A") A3=("Intercept") A4=("Time (years)") A5=("Exposure")
A6=("Exposure*Time") A7=("Age (years, c60)") A8=("Age*Time") ///
B3=(mean_ARmixed_b00_int) B4=(mean_ARmixed_b10_time) B5=(mean_ARmixed_b01_exposure)
B6=(mean_ARmixed_b11_exptime) B7=(mean_ARmixed_b02_agec60)
B8=(mean_ARmixed_b12_agec60time) ///
C3=(mean_ARmixed_b00_int_se) C4=(mean_ARmixed_b10_time_se)
C5=(mean_ARmixed_b01_exposure_se) C6=(mean_ARmixed_b11_exptime_se)
C7=(mean_ARmixed_b02_agec60_se) C8=(mean_ARmixed_b12_agec60time_se) ///
D3=(empSE_ARmixed_b00_int) D4=(empSE_ARmixed_b10_time) D5=(empSE_ARmixed_b01_exposure)
D6=(empSE_ARmixed_b11_exptime) D7=(empSE_ARmixed_b02_agec60)
D8=(empSE_ARmixed_b12_agec60time) ///
E3=(RMSE_ARmixed_b00_int) E4=(RMSE_ARmixed_b10_time) E5=(RMSE_ARmixed_b01_exposure)
E6=(RMSE_ARmixed_b11_exptime) E7=(RMSE_ARmixed_b02_agec60)
E8=(RMSE_ARmixed_b12_agec60time) ///
F3=(Pcoverage_ARmixed_b00_int) F4=(Pcoverage_ARmixed_b10_time)
F5=(Pcoverage_ARmixed_b01_exposure) F6=(Pcoverage_ARmixed_b11_exptime)
F7=(Pcoverage_ARmixed_b02_agec60) F8=(Pcoverage_ARmixed_b12_agec60time) ///
G3=(Z_ARmixed_b00_int) G4=(Z_ARmixed_b10_time) G5=(Z_ARmixed_b01_exposure)
G6=(Z_ARmixed_b11_exptime) G7=(Z_ARmixed_b02_agec60) G8=(Z_ARmixed_b12_agec60time) ///
H3=(pvalue_ARmixed_b00_int) H4=(pvalue_ARmixed_b10_time)
H5=(pvalue_ARmixed_b01_exposure) H6=(pvalue_ARmixed_b11_exptime)
H7=(pvalue_ARmixed_b02_agec60) H8=(pvalue_ARmixed_b12_agec60time) ///
using SimulationResults_N`N'_B`B'_pctdied`pctdied', sheet("ARmixedResults") modify


/**Proportion of sample who died and survival times; Store in sheet named
"DeathInfo"**/
putexcel A1=("Scenario") B1=("Proportion died") C1=("Avg. f/u time") ///
A2=("ScenarioA") B2=(mean_p_study_death) C2=(mean_survtime) ///
using SimulationResults_N`N'_B`B'_pctdied`pctdied', sheet("DeathInfo") modify


/**# individuals (N), # simulations (B), and time between visits; Store in sheet named
"NandBinfo"**/
putexcel A1=("Scenario") B1=("N (# individuals)") C1=("B (# simulations)") D1=("time
btwn visits") ///
A2=("ScenarioA") B2=(`N') C2=(`B') D2=(`int_time') ///
using SimulationResults_N`N'_B`B'_pctdied`pctdied', sheet("NandBinfo") modify
/*************************************************************************/
```

Web Table 4. Scenario A-C simulation results for the estimated effect of exposure on rate of cognitive change (β_{11}) per 10 years from naïve linear mixed effects models with random effects and autoregressive within-person covariance structure: archetype input values (Table 1).

	Low mortality (25%)			Intermediate mortality (50%)			High mortality (75%)		
	$\overline{\hat{\beta}}_{11}$	RMSE	Coverage[a]	$\overline{\hat{\beta}}_{11}$	RMSE	Coverage[a]	$\overline{\hat{\beta}}_{11}$	RMSE	Coverage[a]
Scenario A	0.002	0.085	0.956	0.000	0.094	0.959	-0.002	0.120	0.957
Scenario B1	0.049	0.099	0.912	0.076	0.125	0.887	0.095	0.155	0.881
Scenario B2	0.169	0.190	0.501	0.295	0.311	0.146	0.421	0.437	0.064
Scenario C1	0.093	0.125	0.808	0.183	0.206	0.527	0.311	0.331	0.234
Scenario C2	0.090	0.122	0.818	0.171	0.194	0.546	0.287	0.307	0.260

Abbreviations: RMSE, root-mean-square error. [a]Coverage = proportion of times the 95% confidence interval for $\hat{\beta}_{11}$ includes β_{11}.

Web Table 5. Scenario B1 (exposure and an unmeasured U affect mortality) simulation results for the estimated effect of exposure on rate of cognitive change (β_{11}) per 10 years from naïve linear mixed effects models with random effects and autoregressive within-person covariance structure: alternative input values (more moderate effect sizes).

	Low mortality (25%)			Intermediate mortality (50%)			High mortality (75%)		
	$\overline{\hat{\beta}}_{11}$	RMSE	Coverage[a]	$\overline{\hat{\beta}}_{11}$	RMSE	Coverage[a]	$\overline{\hat{\beta}}_{11}$	RMSE	Coverage[a]
Under archetypical conditions[b]	0.049	0.099	0.912	0.076	0.125	0.887	0.095	0.155	0.881
Moderate effect of exposure on mortality[c]	0.030	0.090	0.933	0.045	0.108	0.919	0.056	0.135	0.922
Moderate effect of U on mortality[d]	0.036	0.093	0.936	0.064	0.115	0.908	0.089	0.150	0.902
Moderate effect of U on cognitive decline[e]	0.026	0.087	0.938	0.039	0.104	0.919	0.048	0.130	0.933

Abbreviations: RMSE, root-mean-square error. [a]Coverage = proportion of times the 95% confidence interval for $\hat{\beta}_{11}$ includes β_{11}; [b]Results in this row replicate results for Scenario B1 shown in Web Table 4 for convenience of comparisons; [c]Input value for γ_1=0.40 (HR=1.5) instead of 0.69 (HR=2.0); [d]Input value for γ_3=0.69 (HR=2.0) instead of 1.61 (HR=5.0); [e]Input value for β_{13}=-0.025 instead of -0.05.

Web Table 6. Scenario B2 (exposure and an unmeasured U have more than multiplicative effects on mortality hazard) simulation results for the estimated effect of exposure on rate of cognitive change (β_{11}) per 10 years from naïve linear mixed effects models with random effects and autoregressive within-person covariance structure: alternative input values (more moderate effect sizes).

	Low mortality (25%)			Intermediate mortality (50%)			High mortality (75%)		
	$\overline{\hat{\beta}}_{11}$	RMSE	Coverage[a]	$\overline{\hat{\beta}}_{11}$	RMSE	Coverage[a]	$\overline{\hat{\beta}}_{11}$	RMSE	Coverage[a]
Under archetypical conditions[b]	0.169	0.190	0.501	0.295	0.311	0.146	0.421	0.437	0.064
Moderate effect of exposure on mortality[c]	0.158	0.180	0.550	0.275	0.292	0.196	0.392	0.410	0.094
Moderate effect of exposure and U on mortality[d]	0.083	0.119	0.846	0.165	0.191	0.592	0.266	0.292	0.405
Moderate effect of U on cognitive decline[e]	0.087	0.120	0.826	0.150	0.177	0.648	0.213	0.242	0.555

Abbreviations: RMSE, root-mean-square error. [a]Coverage = proportion of times the 95% confidence interval for $\hat{\beta}_{11}$ includes β_{11}; [b]Results in this row replicate results for Scenario B2 shown in Web Table 4 for convenience of comparisons; [c]Input value for γ_1=0.40 (HR=1.5) instead of 0.69 (HR=2.0); [d]Input value for γ_4=0.69 HR=2.0) instead of 1.61 (HR=5.0); [e]Input value for β_{13}=-0.025 instead of -0.05.

Web Table 7. Scenario C1 (exposure and rate of cognitive change affect mortality) simulation results for the estimated effect of exposure on rate of cognitive change (β_{11}) per 10 years from naïve linear mixed effects models with random effects and autoregressive within-person covariance structure: alternative input values (more moderate effect sizes).

	Low mortality (25%)			Intermediate mortality (50%)			High mortality (75%)		
	$\overline{\hat{\beta}}_{11}$	RMSE	Coverage[a]	$\overline{\hat{\beta}}_{11}$	RMSE	Coverage[a]	$\overline{\hat{\beta}}_{11}$	RMSE	Coverage[a]
Under archetypical conditions[b]	0.093	0.125	0.808	0.183	0.206	0.527	0.311	0.331	0.234
Moderate effect of exposure on mortality[c]	0.054	0.099	0.903	0.105	0.141	0.817	0.177	0.210	0.677
Moderate effect of rate of cognitive change on mortality[d]	0.042	0.094	0.928	0.096	0.136	0.832	0.187	0.223	0.674

Abbreviations: RMSE, root-mean-square error. [a]Coverage = proportion of times the 95% confidence interval for $\hat{\beta}_{11}$ includes β_{11}; [b]Results in this row replicate results for Scenario C1 shown in Web Table 4 for convenience of comparisons; [c]Input value for γ_1=0.40 (HR=1.5) instead of 0.69 (HR=2.0); [d]Input value for γ_5=-1.0 (HR=0.4) instead of -2.9 (HR=0.1).