Published in final edited form as:

Eur J Epidemiol. 2017 January ; 32(1): 55–66. doi:10.1007/s10654-016-0197-8.

Multiple Imputation of Cognitive Performance as a Repeatedly Measured Outcome

Andreea M. Rawlings¹, Yingying Sang¹, A. Richey Sharrett¹, Josef Coresh¹, Michael Griswold², Anna M. Kucharska-Newton³, Priya Palta³, Lisa M. Wruck⁴, Alden L. Gross¹, Jennifer A. Deal¹, Melinda C. Power^{1,5}, and Karen Bandeen-Roche⁶

¹Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland ²Center of Biostatistics and Bioinformatics, University of Mississippi Medical Center, Jackson, Mississippi ³Department of Epidemiology, University of North Carolina, Chapel Hill, North Carolina ⁴Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina ⁵Department of Epidemiology and Biostatistics, George Washington University Milken Institute of Public Health, Washington, District of Columbia ⁶Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland

Abstract

Background—Longitudinal studies of cognitive performance are sensitive to dropout, as participants experiencing cognitive deficits are less likely to attend study visits, which may bias estimated associations between exposures of interest and cognitive decline. Multiple imputation is a powerful tool for handling missing data, however its use for missing cognitive outcome measures in longitudinal analyses remains limited.

Methods—We use multiple imputation by chained equations (MICE) to impute cognitive performance scores of participants who did not attend the 2011–2013 exam of the Atherosclerosis Risk in Communities Study. We examined the validity of imputed scores using observed and simulated data under varying assumptions. We examined differences in the estimated association between diabetes at baseline and 20-year cognitive decline with and without imputed values.

Andreea M. Rawlings, PhD MS, Welch Center for Prevention, Epidemiology and Clinical Research and the Johns Hopkins Bloomberg School of Public Health, 2024 E. Monument Street, Suite 2-600, Baltimore MD 21287, Tel: 443-287-4169; arawlin2@jhu.edu.

ARIC-NCS Steering Committee: Thomas H. Mosley, PhD (Chair); Josef Coresh (Co-Chair), MD, PhD; Marilyn Albert, PhD; Alvaro Alonso, MD, PhD; Christie M. Ballantyne, MD; Eric Boerwinkle, PhD; David Couper, PhD; Gerardo Heiss, MD, PhD; Clifford Jack, MD; Barbara Klein, MD, MPH; Ron Klein, MD, MPH; David Knopman, MD; Natalie Kurinij, PhD (National Eye Institute Project Office); Claudio Moy, PhD (National Institute of Neurological Disorders and Stroke Project Officer); and Jacqueline Wright, PhD (NHLBI Project Officer). Ex Officio Members: Laura Coker, PhD, Aaron Folsom, MD, MPH, Rebecca F. Gottesman, MD, PhD, A. Richey Sharrett, MD, DrPH, Lynne E. Wagenknecht, DrPH, and Lisa Miller Wruck, PhD.

ARIC-NCS Data Analysis Committee (drafted and critically revised all analysis plans): A. Richey Sharrett, MD, DrPH (Chair), Karen Bandeen-Roche, PhD (Senior Statistician), Andrea L.C. Schneider, MD, PhD, Josef Coresh, MD, PhD, Jennifer A. Deal, PhD, Rebecca F. Gottesman, MD, PhD, Michael Griswold, PhD, Alden L. Gross, PhD, Thomas H. Mosley, PhD, Melinda C. Power PhD, Andreea M. Rawlings, MS, Lisa Miller Wruck, PhD, and Shoshana Ballew, PhD (Epidemiologist coordinator)

ARIC-NCS Neurocognitive Committee: Thomas Mosley, PhD (Chair), Rebecca F. Gottesman, MD, PhD (Co-Chair), Alvaro Alonso, MD, PhD, Laura Coker, PhD, David Couper, PhD, David Knopman, MD, Guy McKhann, MD, Ola Selnes, PhD, and A. Richey Sharrett, MD, DrPH.

Disclosures: none.

Lastly, we discuss how different analytic methods (mixed models and models fit using generalized estimate equations) and choice of for whom to impute result in different estimands.

Results—Validation using observed data showed MICE produced unbiased imputations. Simulations showed a substantial reduction in the bias of the 20-year association between diabetes and cognitive decline comparing MICE (3-4% bias) to analyses of available data only (16-23% bias) in a construct where missingness was strongly informative but realistic. Associations between diabetes and 20-year cognitive decline were substantially stronger with MICE than in available-case analyses.

Conclusions—Our study suggests when informative data are available for non-examined participants, MICE can be an effective tool for imputing cognitive performance and improving assessment of cognitive decline, though careful thought should be given to target imputation population and analytic model chosen, as they may yield different estimands.

Keywords

bias; cognitive function; epidemiologic methods; missing data; multiple imputation; prospective study

Introduction

Missing data is common in epidemiologic studies. In longitudinal studies, the focus is often on how a baseline exposure is associated with changes in an outcome. Since participants who do not attend subsequent study visits are likely informatively different from those who do attend, associations may be biased if missing data are not handled appropriately.

Multiple imputation is a powerful tool for dealing with missing data¹⁻⁴. However, use of imputation for cognitive outcome studies remains limited^{5,6}, perhaps because other methods are effective for correcting potential biases. For example, maximum likelihood methods, routinely used in fitting mixed models, account for biases when missingness is random with respect to variables included in primary analyses⁷. Inverse probability of attrition weighting methods and shared parameter models⁸⁻¹¹ are also used to address biases associated with dropout and death and may incorporate additional variables not included in primary analyses.

Multiple imputation is particularly useful when data are available for at least a subset of participants who did not attend all study visits, such as participants with low cognitive performance who are typically less likely to attend follow-up examinations¹²⁻¹⁵. Large epidemiologic studies with repeat examinations often collect such data through morbidity and mortality surveillance or follow-up telephone calls, and may identify participants believed to have cognitive impairment. When such data are also collected for individuals who attend study visits, multiple imputation may be particularly useful in addressing potential biases compared to other analytical methods, and may be a useful tool for translating this auxiliary information into a cognitive battery score. Substantially informative but where auxiliary variables, which are strongly correlated to the dropout and longitudinal processes, can be utilized to make model assumptions about missingness more plausible.

In this study we aimed to document the utility of multiple imputation to address data incompleteness in settings where dropout may be substantially informative but where auxiliary data were collected to bridge information obtained from fully and partially followed individuals. These auxiliary data are not confounders (rather, alternative measures of cognitive function), and cannot readily be included in analytic models. However these data lend themselves well to inclusion in multiple imputation procedures, allowing for unbiased estimates of associations of interest if all variables relevant to dropout are included in the imputation, making assumptions about missingness more plausible. To illustrate the utility of multiple imputation using chained equations (MICE)^{2,16} to address missing data, we present a case study implementing MICE in a complex situation: estimating the association between diabetes at baseline (exposure) and cognitive performance (the outcome, measured three times over 20 years of follow-up). Specifically, we show analytic findings, and evaluate their robustness, based on imputing cognitive performance scores of participants from the Atherosclerosis Risk in Communities (ARIC) Study who did not attend the 2011-2013 exam. We also evaluate the utility of MICE in a simulation study generating data under different assumptions regarding dropout mechanisms, and show that the accuracy of imputation improves with the use of auxiliary information. Finally, using multiple imputation in longitudinal analyses compared to cross-sectional analyses presents unique issues, such as for whom to impute and the choice of an analytic model, which can affect the inferences and target estimands. There is considerable debate in the literature regarding accounting for attrition due to death¹⁷, and our study documents how choice of analytic method and for whom to impute can impact estimates of interest.

Methods

Study population

ARIC is a community-based, prospective cohort of 15,792 middle-aged adults from four communities in Maryland, Minnesota, Mississippi, and North Carolina¹⁸. Participants were examined at four triennial visits, beginning in 1987-1989. A fifth examination occurred in 2011-2013. Participants in North Carolina and Mississippi also had cognitive assessment at Brain and Carotid MRI visits (2004-2006, N=2790). Baseline for the present study was visit 2 in 1990-1992 (where cognitive assessment began); we excluded participants who did not attend baseline (N=1444) or who were neither black nor white (N=91). Institutional review boards from each center approved the study, and all participants provided informed consent.

Diabetes assessment

Diabetes at visit 2 was defined as self-reported physician diagnosis, diabetes medication use, or a hemoglobin A1c level $\geq 6.5\%$.

Cognitive assessment at study visits

Cognitive function was assessed at visits 2, 4, and 5 using 3 tests: Delayed Word Recall¹⁹, Digit Symbol Substitution²⁰, and Word Fluency²¹. We standardized each test score to visit 2 by subtracting the test mean (at visit 2) from each participant's score and dividing by the test standard deviation (SD, at visit 2). A global Z score, calculated by averaging the Z score of

the three tests, was likewise standardized to visit 2. The global Z score was the outcome of interest and the focus of the imputation.

Auxiliary measures of cognitive function

Information about cognitive function for participants who did not attend visit 5 was available through the modified Telephone Interview for Cognitive Status (TICS-m) questionnaire, suspect dementia status, and the Clinical Dementia Rating (CDR) scale.

The TICS-m, a test of cognitive function given over the telephone^{22–24}, was offered to all participants who did not attend visit 5 (completed for N=1327), and to a random subsample of participants who attended visit 5 (N=255).

Participants were classified as having suspect dementia based on information obtained by telephone with the participant or their proxy, or an ICD-9 code of dementia appearing in any position in hospital discharge records²⁵ (N=1462). If participants with suspect dementia did not complete visit 5 or the TICS-M, their proxies were sought to complete a CDR. Suspect dementia status was available for all participants in ARIC.

For participants with suspect dementia, interviews were sought with proxy informants. The CDR was completed by telephone with informants familiar with the participant's current cognitive status (for living participants) or cognitive status 12 months prior to death. It covers six domains (memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care). For deceased participants, interviewers were carefully instructed to focus on change in cognitive status occurring 12 or more months prior to death, and to avoid reports of pre-terminal cognitive decline. Considering difficulties in attempting to reach proxies of participants who died more than 10 years prior to visit 5 and that few participants would be expected to have dementia prior to this date (mean age was 70), a CDR was sought only for participants who died after 2004.

Interviewers scored each of the six domains using a scale of 0 (no impairment), 0.5 (questionable), 1 (mild), 2 (moderate), and 3 (severe impairment). The CDR sum of boxes (total score) ranged from 0 to 18. The CDR was collected on 885 participants who did not attend visit 5 (N=575 with suspect dementia) and from 2856 who attended visit 5 (N=176 with suspect dementia).

Diabetes association with cognitive decline over 20 years

To examine the association between diabetes and cognitive change over 20 years, we considered two modeling strategies: mixed-effects models and models fit using generalized estimating equations (GEE). Time since baseline (visit 2) was modeled using a linear spline with a knot at six years (median time to visit 4). Our mixed models included one random intercept and a random slope for each time spline term, with random effects assumed to be independent. The coefficients of interest were the interaction terms between diabetes and each time spline term, which indicate differential decline over time among persons with diabetes at baseline compared to those without. A limitation of the mixed model approach is that there is implicit imputation of cognitive scores beyond death (via the random effects), resulting in inferences to an immortal cohort²⁶. While such an estimand may be of interest,

we also applied models fit using GEE and independent working correlation with robust variance estimation. Using GEE with independent working correlation avoids the potentially undesired implicit imputation effect when participants' data are missing.

The different analytic strategies (mixed and GEE), combined with the persons for whom we impute (no imputation, imputation for living, or imputation for living and deceased participants), give rise to different estimands, which should be carefully considered when choosing an imputation and analysis strategy. For this study we focus on three analyses: 1) using MICE to impute missing participants' scores (living and deceased) and estimating associations of interest using a mixed model with random intercept and slopes, 2) using MICE to impute missing participants' scores (living and deceased) and estimating associations of interest using GEE with independent working correlation, and 3) restricting the population to 20-year survivors, using MICE to impute scores for participants living at visit 5, and conducting analyses using both mixed models and GEE. The first analysis infers the findings to be expected if those dying had remained alive subsequently, continuing on their trajectory while living. The second targets the population average association between diabetes and cognitive decline while participants were alive. The last targets the population average association conditional on participants surviving 20 years. Lastly, we compare results from these analyses to those obtained without the use of imputation and for 1) and 2) above, to those using imputation only for participants living to visit 5, where the deceased contribute data only while living.

All models were adjusted for demographic, behavioral, and cardiovascular risk factors as have been previously used²⁷ (See Figure 2 legend).

Multiple imputation

Missing data can be classified as follows^{28,29}: missing completely at random (MCAR) when missingness does not depend on either observed or unobserved data; missing at random (MAR) when, after conditioning on observed data, missingness does not depend on unobserved data; or missing not at random (MNAR), when missingness depends on unobserved data (such as unmeasured dementia status).

Multiple imputation replaces missing data with plausible values, and has been demonstrated to produce asymptotically unbiased estimates when missing data are MAR or MCAR^{2,30}. To account for the uncertainty of the imputation and ensure correct standard error estimation, multiple imputations are performed²⁸. Multiple imputation by chained equations (MICE) involves a series of imputation models, where each variable containing missing data is regressed on all other variables, including previously imputed missing variables^{2,16,30,31}. The flexibility of MICE to impute different data types (categorical, continuous, binary, etc) makes it an attractive tool for use in practice. We used 25 sets of imputations, although we observed stability in estimates after 6-7 imputations (eFigure1). For participants alive at visit 5, scores were imputed at the median visit date. For participants who were deceased by visit 5, scores were imputed 6 months prior to death. In analyses using imputation for the living and deceased, all records with imputed values were retained. In analyses conditional on survival, all records after a participant was deceased were dropped from the dataset. In analyses conditional on survival to visit 5, all records of persons who died during follow-up

were dropped from the dataset. After using MICE to impute cognitive performance scores, we estimated the association of diabetes with 20-year cognitive change, using mixed models and models fit using GEE as described above, by conducting analyses separately on each imputed dataset, and combining the estimated coefficients and standard errors from each analysis using Rubin's rules⁴. In both the mixed model and GEE analyses, outcome means were modeled with a linear link and were related to time by a spline model as described above.

The imputation model for global Z score at visit 5 included the same variables as the mixed-effects longitudinal model (described above) as well as variables collected from annual telephone calls, TICS-M, suspect dementia status, CDRs, and global Z scores from visits 2 and 4. From the annual telephone call most nearly preceding visit 5 we included the following variables (all coded yes/no): coronary heart disease, diabetes status, hypertension status, history of stroke, self-reported poor health, and an indicator of whether a proxy report was needed. We selected these variables *a priori* based on knowledge of their association with probability of dropout and cognitive function. Interaction terms between suspect dementia and education, race-field center, prior visit Z scores, CDR, diabetes, and hypertension were also included. Interaction terms were needed because suspect dementia modified outcome relationships with prior cognitive performance and other covariates. For example, if a person with suspect dementia was found by CDR to have severe impairments, cognitive performance at an earlier exam may be less informative relative to current performance. We included variables for time since baseline as a basis for timing imputation of visit 4 and visit 5 scores. Additionally, we examined whether an indicator for death modified the relationship between covariates and cognitive function by including interaction terms between the death indicator and other covariates or interest (CDR, prior Z scores, suspect dementia, and diabetes). These interactions were not significant and were not included in the final imputation model. We examined including other visit-based variables from clinical chemistries, medical/health history, anthropometry, and medication survey. Since these additional variables did not improve the imputation or change the results of longitudinal analyses using these imputations, they were not used in the analyses reported here.

Lastly, we compared imputed values with and without the use of auxiliary information to gauge the improvement in imputation with the use of this data. While few participants with suspect dementia came to visit 5, auxiliary information in this group would be most informative for imputing scores.

Validation and simulation

We used two validation approaches. First, we set to missing cognitive scores of a random sample of participants who attended visit 5, and then compared imputed with observed values. To validate imputations under an MCAR missingness assumption among participants alive at visit 5, we randomly selected 20% of participants and set their Z score to missing; to validate under a MAR missingness assumption, we used a logit model to allow the probability of missingness to differ by the following baseline variables: age, race-center, education, diabetes, global Z from visit 2, and diabetes*global Z from visit 2. Because none

of the persons who died attended visit 5, we could not use the same approach to validate cognitive Z scores for the deceased. As an alternative, we used Z scores obtained from participants who attended the Brain or Carotid MRI visits (2004-2006). Scores for participants who died are imputed across a wider time range (2004-2013). To obtain observed Z scores proximal to when Z scores for the deceased were imputed, we conducted our validation analysis for deceased participants by comparing the observed Z scores from the 2004-2006 visits to imputed scores in those who died within two years of the Brain or Carotid MRI visit. The 2004-2006 data were used only for validation, and were not used in the imputation model.

As a second validation approach, we evaluated the performance of MICE using a simulation study addressing patterns of missingness corresponding to MCAR, MAR, and MNAR. We retained the observed values of all covariates for individuals from the ARIC population and simulated Z scores using a mixed-effects model that included age, race-field center, sex, body mass index, suspect dementia, diabetes, hypertension, and interaction terms suspect dementia*time, hypertension*time, current cigarette smoking*time, and diabetes*time, modeling time using spline terms (described above). Including suspect dementia was necessary to allow us to retain the correlations between risk factors and cognitive decline. Additionally, CDR data were specifically sought for persons with suspect dementia, so including it was necessary in generating “believable” Z scores. Using this model, persons with hypertension, diabetes, or smokers, by design, had accelerated cognitive decline. Coefficients of each variable and the random-effects parameters were chosen to be similar to values estimated from the cohort (these were estimated by fitting this same model using observed ARIC data). Simulation specifications are detailed in the Appendix.

To model probabilities of dropout and death, we created four scenarios reflecting different dropout mechanisms. Scenario 1 (MCAR) assumed death and dropout occurred completely at random. Scenario 2 (MAR) assumed the probabilities of death and dropout (separately) depended on prior visit global Z score, diabetes, hypertension, and smoking status modeled using a multinomial logistic regression. These variables are included in our analytic model and hence result in a MAR scenario. Scenario 3 (MAR for the variables included in the imputation, MNAR for variables included in primary analyses) assumed the probabilities of death and dropout depended on prior visit global Z score (visit 2 and 4), diabetes, hypertension, smoking status, suspect dementia, and a diabetes*suspect dementia interaction (suspect dementia collected around visit 5). In this scenario, the variables suspect dementia and its interaction with diabetes are included in the MICE to impute scores, but are not included in our analytic model, resulting in MAR and MNAR for MICE and our analytic model, respectively. Scenario 3 is most consistent with what we believe the true missingness pattern in ARIC to be. Scenario 4 (MNAR) assumed that dropout depended only on simulated visit 5 global Z scores (i.e. unobserved scores), and that death among dropouts was random with a probability of 0.4. For each scenario, we analyzed data using available-case analysis, MICE restricted to participants living at the time of visit 5, and MICE including both living and dead participants. Estimates for all scenarios were obtained using both mixed-effects modeling and using GEE with independent correlation structure. We also calculated standard errors, bias, and confidence interval (CI) coverage (the percentage of simulations where the true association was contained in the 95% CI). Bias was calculated

both relative to a trajectory-among-those-living-for-visits and to a trajectory-up-to-death. The “truth” trajectories were estimated using the same models described above (mixed/GEE) in these simulated data prior to implementing the dropout or death scenarios (i.e. prior to creating the missing data).

For both validation approaches, MICE was used to perform imputation, and the longitudinal analyses were conducted using linear-link mixed models or linear-link models fit with GEE. Analyses were completed using Stata/SE Version 13.1 (StataCorp, College Station, TX).

Results

Of the original visit 2 cohort, 55% did not attend visit 5, with approximately equal percentages due to death (29%) and dropout but living prior to visit 5 (26%); these values also represent the percentages of the original visit 2 cohort for which imputation was performed. Comparatively, 16% of the visit 2 cohort did not attend visit 4 (but were living) and 3.8% were deceased by visit 4 (eTable 1). Compared to participants who attended visit 5, participants who died by visit 5 tended to be older at baseline (age 60 vs 55 years), were more likely to have diabetes (24% vs 8%) and a history of stroke (4% vs 1%), and had worse baseline cognitive performance (Table 1). Additionally, 15% of the deceased were suspected of having dementia, compared to only 4% among participants who attended visit 5.

Validation results based on observed data are shown in Figure 1. MICE produced unbiased imputed values regardless of whether an MCAR or a MAR approach was used to select the validation sample (Figure 1, **Panels A and B**). Additionally, imputed values were unbiased in subgroups defined by race, education, diabetes, cognitive performance at visit 4, and suspect dementia (not shown). Among both MCAR and MAR validation samples, and by these subgroups, mean differences between imputed and observed global Z ranged from -0.03 to +0.02 Z scores, and the r-squared from a linear fit model between observed and average imputed scores ranged from 0.65 to 0.68. As shown in Figure 1, **Panel C**, among 74 participants who died less than 2 years after attending the Brain or Carotid MRI visits, agreement between the imputed and observed global Z scores was excellent. The mean difference was -0.02 Z scores, and the r-squared was 0.70 from the linear fit model where observations were weighted relative to time since the brain or carotid visit (calculated as 1/time, such that deaths closer to the visit received higher weights). Finally, Figure 1, **Panel D** shows the distribution of imputed scores at visit 5, by CDR availability, among persons with suspect dementia. The characteristics of participants without a CDR were similar to those with a CDR (eTable 2). However, because the informant could not be located (and CDRs were not obtained), the average imputed scores were higher by 0.55 Z scores than the average imputed score for participants whose informant was interviewed. This result implies that when a CDR could not be obtained, we had insufficient information with which to impute a plausibly low enough cognitive score. Using the auxiliary information on cognitive function in MICE resulted in a higher r-squared, lower root mean squared error, and lower imputed scores among participants with suspect dementia, compared to MICE without the auxiliary information (eFigures 2 and 3).

Simulation results are in Table 2. When data were MCAR (scenario 1) or MAR (scenario 2), all methods yielded approximately unbiased estimates, as expected. In scenario 3, where dropout depended on suspect dementia, available-case analysis (i.e. no imputation) using mixed models yielded a 16% bias, which was reduced to 4% with imputation for living participants, and a reduction from 23% to 3% with imputation for living and deceased. We observed similar trends for scenario 3 using GEE, with a 33% bias reduced to 2% with imputation in the living, and a 55% bias reduced to 10% with imputation in living and deceased. We also saw that targeting a trajectory among the living (using GEE) can result in an anticonservative bias if the imputation includes living and deceased participants (22% and 34% anticonservative bias in scenarios 2 and 3). In scenario 4, where participants were missing based on their unobserved cognitive function, no method yielded unbiased results (bias $\approx$ 18-32%).

Estimates of 20-year additional cognitive decline in persons with diabetes compared to those without are shown in Figure 2. For both mixed model and model fit with GEE, imputation yielded larger estimates of additional decline due to diabetes, compared to available case analysis, although confidence intervals overlapped. Mirroring the simulation study, estimates of additional decline were consistently smaller for GEE (-0.14 – -0.17 Z) compared to the mixed model (-0.21 Z). Estimates across models were nearly identical when we conditioned on surviving through the end of follow-up (i.e. restricted analyses to participants who survived 20-years).

Discussion

In this community-based cohort study, we used multiple imputation by chained equations to impute cognitive performance as the outcome for subsequent epidemiologic questions. Validation analyses showed that MICE yielded unbiased imputations of cognitive performance for both living and deceased participants, with the exception that the procedure may not specify scores plausibly low enough for persons with suspect dementia whose informants could not be interviewed. In those believed to have dementia, the use of auxiliary information yielded improved imputation of cognitive scores compared to imputation without this information. We showed that estimates of the associations of diabetes with 20-year cognitive decline were substantially further from the null with the use of MICE, compared to an available-case analysis. Simulations showed that when data are informatively missing and related additional data are available, MICE may produce less biased estimates of associations of interest compared to available-case analysis. Lastly, we demonstrated changes in estimates depending on for whom we impute (living or living and deceased) and with the analytic approach (mixed models or independence GEE).

We note several limitations to our simulation study. First, suspect dementia was built into the data-generating model in our simulations but not included as a covariate in subsequent mixed models. The need to do so highlights implications of unobserved covariates even in a MAR scenario: analyses including only participants living at visits and those also incorporating the deceased may target different estimands as a result of the groups' differentiation by the unobserved covariate. Second, while we chose parameters for simulation models that we believe are realistic (coefficients were obtained from models

using observed ARIC data), simulation results depend on assumptions made about the generating, death, and dropout models.

We imputed scores for the dead 6 months before death, with attempts made to ignore pre-terminal changes. However, certain analysis methods, such as mixed-models, may implicitly impute cognitive trajectories beyond death, giving inferences for an immortal cohort^{17,26}, and we observed an indication of this occurring in our analyses. Such analyses (mixed model with imputation for all participants) target an estimand that can be interpreted as the 20-year population average association had participants remained alive and under observation (via study visits or ancillary information). Immortal cohort inferences may be of interest as interventions to prolong life continue to improve, and people with cognitive impairment live longer^{32,33}, and may also be of particular interest in causal inference. While this approach has merits, it also has limitations. Imputations for dead participants are placed before death, which can occur along a wide time interval, while imputations for living participants are anchored to visits. Thus, the former imputation gives different statistical leverage to those who died. One remedy is potentially the use of linear models fit with GEE using an independent working correlation matrix. While standard errors need to be carefully estimated in this scenario, this approach may be useful when using imputed scores, as it would avoid the potentially undesired effect of implicit imputation and inferences for an immortal cohort. Such analysis targets an estimand that represents the whole population's cognitive natural history up to, but not beyond death, or the population average association while living.

Estimating trajectories of cognitive function using data only at clinic visits of living participants has the advantage of being directly informed by observations timed independently of adverse outcomes. Resulting estimands of cognitive decline among persons remaining alive is also relevant as persons with diabetes may be interested in knowing the cognitive trajectory they might expect assuming they survive 20 years. This estimand in a mortal cohort resulted from restricting the population to survivors, and estimates across models (GEE vs mixed) were nearly identical. However, ignoring the stronger association of diabetes with cognitive decline in persons who die or drop out due to dementia is likely to inadequately represent the natural history of the entire baseline population.

In our mixed model analyses we debated which imputation approach should be prioritized: one imputing outcomes for participants lost to follow up due to death (outcomes timed six months prior to death), or one imputing outcomes only for living participants. We concluded that the former was more consistent with the immortal inference provided by the mixed model approach. However we continue to see value for mixed models retaining only information while living (whether imputed or not), to address a mortal inference. The latter avoids the analytic limitations noted two paragraphs above, as well possibly as conceptual ones. Such an analysis may more validly estimate the mean rate of change at a given time than the GEE alternative, because it explicitly addresses this estimand (as opposed to “population average” differences in means across time for GEE).

The choice of when and for whom we impute the outcome deserves careful thought. While our study saw relatively similar results under two imputation scenarios (imputing only for

living participants and imputing for living and deceased) and analytic methods (mixed model and GEE), others may not. While our study does not settle the more controversial question of what approach is preferable for dealing with the potential bias induced when attrition is due to death^{17,34}, it adds to the literature in this area. Though guidance regarding multiple imputation is available^{2,16,31,35}, less is known about its utilization in epidemiologic studies for imputing cognitive outcomes in longitudinal analyses and the analytic issues that arise in this setting as we have documented. While methods such as inverse probability weighting or likelihood-based approaches are more common³⁶, multiple imputation may be ideal for handling missing data when valuable information is available only in a subset of participants, as is the case in our and other community-based cohort studies. More research is needed to determine if a combined approach using both imputation and inverse probability weighting in epidemiologic studies would yield improved estimates^{37,38}.

Advantages of MICE include its flexibility in imputing different data types (e.g. categorical, continuous, etc.), and relative ease of implementation using standard statistical packages. A disadvantage of MICE may be its atheoretical nature. Specifically, the series of conditional models may lead to situations where the joint distributions are incompatible. However studies have shown that MICE appears to be generally robust against such incompatibility^{1,30,39}. MICE does not necessarily produce unbiased estimates when data are missing based on unobserved information; in such scenarios, analyses to explicate sensitivity of findings to the strength of non-ignorable associations are optimal^{40,41}. A notable disadvantage of MICE is program run time, which required hours to run. Finally, careful thought should be given to collection of alternative data to supplement the data collected at regular study visits, whether through proxies, phone calls, or other surveillance. We showed that this information improved imputation in informative subgroups, and such supplemental data are invaluable to characterize and minimize informative missingness, provided that one avoids differential information bias.

In summary, our results suggest that when informative data are available for participants who do not attend study visits, MICE is an effective tool for imputing cognitive performance as the outcome, and may improve assessment of cognitive decline.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C, with the ARIC carotid MRI examination funded by U01HL075572-01). Neurocognitive data is collected by U01 HL096812, HL096814, HL096899, HL096902, HL096917 from the NHLBI and the National Institute of Neurological Disorders and Stroke, and with previous brain MRI examinations funded by R01-HL70825 from the NHLBI. The authors thank the staff and participants of the ARIC study for their important contributions.

AMR was supported by NIH/NHLBI grant T32 HL007024. MCP is supported by the NIA grant T32 AG027668. PP was supported by NIH/NHLBI grant T32 HL007055. LMW was supported by 1U01HL096899-01 and HHSN268201100005C.

AMR had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. AMR, ARS, JC, MCP, JAD, and KBR were involved in the design and implementation of the analytic methods, and AMKN and PP provided additional insight on validation analyses. AMR and YS performed statistical analyses. All authors contributed to the interpretation of the results and provided critical revisions to the manuscript.

References

1. Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychol Methods*. 2002; 7(2): 147–177. [Accessed March 22, 2015] <http://www.ncbi.nlm.nih.gov/pubmed/12090408>. [PubMed: 12090408]
2. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med*. 2011; 30(4):377–399. DOI: 10.1002/sim.4067 [PubMed: 21225900]
3. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol*. 2006; 59(10):1087–1091. DOI: 10.1016/j.jclinepi.2006.01.014 [PubMed: 16980149]
4. Rubin, DB. *Multiple Imputation for Nonresponse Surveys*. New York, NY: Wiley; 1987.
5. Seaman SR, Bartlett JW, White IR. Multiple imputation of missing covariates with nonlinear effects and interactions: an evaluation of statistical methods. *BMC Med Res Methodol*. 2012; 12(1):46.doi: 10.1186/1471-2288-12-46 [PubMed: 22489953]
6. Klebanoff MA, Cole SR. Use of multiple imputation in the epidemiologic literature. *Am J Epidemiol*. 2008; 168(4):355–357. DOI: 10.1093/aje/kwn071 [PubMed: 18591202]
7. Laird NM. Missing data in longitudinal studies. *Stat Med*. 1988; 7(1-2):305–315. DOI: 10.1002/sim.4780070131 [PubMed: 3353609]
8. Gad AM, Darwish NMM. A Shared Parameter Model for Longitudinal Data with Missing Values. *Am J Appl Math Stat*. 2013; 1(2):30–35. DOI: 10.12691/ajams-1-2-3
9. Creemers, a, Hens, N., Aerts, M., Molenberghs, G., Verbeke, G., Kenward, MG. Generalized shared-parameter models and missingness at random. *Stat Modelling*. 2011; 11(4):279–310. DOI: 10.1177/1471082X1001100401
10. 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. DOI: 10.1097/EDE.0b013e318230e861 [PubMed: 21989136]
11. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000; 11(5):550–560. [Accessed August 25, 2015] <http://www.ncbi.nlm.nih.gov/pubmed/10955408>. [PubMed: 10955408]
12. Rabbitt P, Diggle P, Holland F, McInnes L. Practice and drop-out effects during a 17-year longitudinal study of cognitive aging. *J Gerontol B Psychol Sci Soc Sci*. 2004; 59(2):P84–P97. [Accessed August 13, 2015] <http://www.ncbi.nlm.nih.gov/pubmed/15014091>. [PubMed: 15014091]
13. Shen C, Gao S. A mixed-effects model for cognitive decline with non-monotone non-response from a two-phase longitudinal study of dementia. *Stat Med*. 2007; 26(2):409–425. DOI: 10.1002/sim.2454 [PubMed: 16345034]
14. Gard T, Hölzel BK, Lazar SW. The potential effects of meditation on age-related cognitive decline: a systematic review. *Ann N Y Acad Sci*. 2014; 1307:89–103. DOI: 10.1111/nyas.12348 [PubMed: 24571182]
15. Gerstorf D, Herlitz A, Smith J. Stability of sex differences in cognition in advanced old age: the role of education and attrition. *J Gerontol B Psychol Sci Soc Sci*. 2006; 61(4):P245–P249. [Accessed August 13, 2015] <http://www.ncbi.nlm.nih.gov/pubmed/16855037>. [PubMed: 16855037]
16. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res*. 2011; 20(1):40–49. DOI: 10.1002/mpr.329 [PubMed: 21499542]
17. Jones M, Mishra GD, Dobson A. Analytical results in longitudinal studies depended on target of inference and assumed mechanism of attrition. *J Clin Epidemiol*. 2015; 68(10):1165–1175. DOI: 10.1016/j.jclinepi.2015.03.011 [PubMed: 25920943]

18. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol.* 1989; 129(4):687–702. <http://www.ncbi.nlm.nih.gov/pubmed/2646917>. [PubMed: 2646917]
19. Knopman DS, Ryberg S. A verbal memory test with high predictive accuracy for dementia of the Alzheimer type. *Arch Neurol.* 1989; 46(2):141–145. <http://www.ncbi.nlm.nih.gov/pubmed/2916953>. [PubMed: 2916953]
20. Wechsler D. *Manual for the Wechsler Adult Intelligence Scale, Revised.* 1981
21. Benton, A., Hamsher, K. *Multilingual Aphasia Examination.* 2nd. Oowa City, IA: AJA Associates; 1989.
22. Brandt J, Spencer M, Folstein M. The Telephone Interview for Cognitive Status. *Neuropsychiatry, Neuropsychol Behav Neurol.* 1988; 1(2):111–118.
23. Knopman DS, Roberts RO, Geda YE, et al. Validation of the Telephone Interview for Cognitive Status-modified in Subjects with Normal Cognition, Mild Cognitive Impairment, or Dementia. *Neuroepidemiology.* 2010; 34(1):34–42. DOI: 10.1159/000255464 [PubMed: 19893327]
24. Plassman BLPD, Newman TTBS, Welsh KAPD, Helms MBS, Breitner, John CS MD, MPH. Properties of the Telephone Interview for Cognitive Status: application in epidemiological and longitudinal studies. *Neuropsychiatry, Neuropsychol Behav Neurol.* 1994; 7(3):235–241.
25. *Manual 19 Surveillance of Dementia in the ARIC Cohort.* 2015
26. Kurland BF, Johnson LL, Egleston BL, Diehr PH. Longitudinal Data with Follow-up Truncated by Death: Match the Analysis Method to Research Aims. *Stat Sci.* 2009; 24(2):211.doi: 10.1214/09-STS293 [PubMed: 20119502]
27. Rawlings AM, Sharrett AR, Schneider ALC, et al. Diabetes in Midlife and Cognitive Change Over 20 Years: A Cohort Study. *Ann Intern Med.* 2014; 161(11):785–793. DOI: 10.7326/M14-0737 [PubMed: 25437406]
28. Little, RJA., Rubin, DB. *Statistical Analysis with Missing Data.* 2nd. Hoboken, NJ: Wiley; 2002.
29. Little RJA. Modeling the Drop-Out Mechanism in Repeated-Measures Studies. *J Am Stat Assoc.* 1995; 90(431):1112.doi: 10.2307/2291350
30. Van Buuren S, Brand JPL, Groothuis-Oudshoorn CGM, Rubin DB. Fully conditional specification in multivariate imputation. *J Stat Comput Simul.* 2006; 76(12):1049–1064. DOI: 10.1080/10629360600810434
31. Bartlett JW, Seaman SR, White IR, Carpenter JR. Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model. *Stat Methods Med Res.* 2014; doi: 10.1177/0962280214521348
32. Manton KG, Gu X, Lamb VL. Long-Term Trends in Life Expectancy and Active Life Expectancy in the United States. *Popul Dev Rev.* 2006; 32(1):81–105. DOI: 10.1111/j.1728-4457.2006.00106.x
33. Sauvaget C, Tsuji I, Haan MN, Hisamichi S. Trends in dementia-free life expectancy among elderly members of a large health maintenance organization. *Int J Epidemiol.* 1999; 28(6):1110–1118. [Accessed June 27, 2016] <http://www.ncbi.nlm.nih.gov/pubmed/10661655>. [PubMed: 10661655]
34. Chaix B, Evans D, Merlo J, Suzuki E. 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. discussion 132–137. DOI: 10.1097/EDE.0b013e3182319159 [PubMed: 22157307]
35. Stuart, Ea, Azur, M., Frangakis, C., Leaf, P. Multiple imputation with large data sets: a case study of the Children's Mental Health Initiative. *Am J Epidemiol.* 2009; 169(9):1133–1139. DOI: 10.1093/aje/kwp026 [PubMed: 19318618]
36. Power MC, Tchetgen EJ, Sparrow D, Schwartz J, Weisskopf MG. Blood pressure and cognition: factors that may account for their inconsistent association. *Epidemiology.* 2013; 24(6):886–893. DOI: 10.1097/EDE.0b013e3182a7121c [PubMed: 24030502]
37. Seaman SR, White IR, Copas AJ, Li L. Combining multiple imputation and inverse-probability weighting. *Biometrics.* 2012; 68(1):129–137. DOI: 10.1111/j.1541-0420.2011.01666.x [PubMed: 22050039]

38. Han P. Combining Inverse Probability Weighting and Multiple Imputation to Improve Robustness of Estimation. *Scand J Stat.* 2015; :n/a–n/a. DOI: 10.1111/sjos.12177
39. Brand JPL. Development, implementation and evaluation of multiple imputation strategies for the statistical analysis of incomplete data sets. 1999
40. Scharfstein DO, Irizarry RA. Generalized additive selection models for the analysis of studies with potentially nonignorable missing outcome data. *Biometrics.* 2003; 59(3):601–613. [Accessed August 23, 2015] <http://www.ncbi.nlm.nih.gov/pubmed/14601761>. [PubMed: 14601761]
41. Greenland S. Basic methods for sensitivity analysis of biases. *Int J Epidemiol.* 1996; 25(6):1107–1116. [Accessed August 23, 2015] <http://www.ncbi.nlm.nih.gov/pubmed/9027513>. [PubMed: 9027513]

Figure 1.

Validation of multiply imputed global Z score using existing data. Multiple imputation was done using chained equations, and 25 imputations were obtained and averaged for display in each plot. **Panel A:** 20% validation sample to simulate missing completely at random (MCAR) data. All participants had a 0.2 probability of being selected. If selected, participants' Z scores at visit 5 were set to missing and imputed. **Panel B:** 20% validation sample to simulate missing at random (MAR) data. Participants had varying probabilities of selection into the validation sample, with probabilities varying by education (less than high school, high school, greater than high school), race (black, white), diabetes (yes, no), or global Z score at visit 2 in the bottom 25th percentile (yes, no), and all interactions. If selected, participants' Z scores at visit 5 were set to missing and imputed. **Panel C:** Validation of imputed scores for participants who were deceased within 2 years of either the brain or carotid MRI visits, which took place 2004-2006 (N=74). Square size is inversely related to difference between visit date and death date (i.e. weight=1/time between visits), such that larger squares indicate death closer to the time of visit. Linear fit line is weighted using the inverse of difference between visit date and death date, and yielded an r-squared of 0.70. **Panel D:** Distribution of imputed scores among people with suspected dementia (N=1,462) by CDR status, adjusted for vital status (living/deceased).

Figure 2. Estimated 20-year additional decline in cognitive performance for persons with diabetes compared to persons without, by model type and use of imputation

Mixed: estimates and 95% CIs are from mixed-effects models using time since baseline as the time axis, modeled using a spline term with a knot at 6 years, the median time between visits 2 and 4. Random effects were random intercept and two random slopes, one for each time spline term.

GEE: estimates and 95% CIs are from models fit using generalized estimating equations with independent working correlation and robust variance estimation.

All models were adjusted for age, age squared, race-field center (Maryland (white race); Minnesota (white race); North Carolina (white race); North Carolina (black race); Mississippi (black race)), sex, education (less than high school; high school, high school equivalent, or vocational school; or college, graduate, or professional school), cigarette smoking status (current; former; never), alcohol consumption status (current; former; never), body mass index (kg/m^2), hypertension (yes or no), history of stroke (yes or no), apolipoprotein E $\epsilon 4$ genotype (0, 1, or 2 alleles). Interaction terms between the time spline terms and age, sex, race-field center, education, history of stroke, and apolipoprotein E $\epsilon 4$ genotype were also included in the model. All covariates were assessed at visit 2 (baseline) except education, race, and sex (visit 1). 25 imputations were generated by chained equations. Sample sizes for mixed/GEE were as follows: No imputation: participants=13482, observations=29616; Imputation of outcome for living participants: participants= 14151, observations=37854; imputation of outcome for living and deceased participants: participants = 14151, observations = 40019. For analyses conditioned on 20-year survival they were: No imputation: participants=10039, observations=23445; Imputation of outcome for living participants: participants= 10039, observations=30117.

Abbreviations: GEE, generalized estimating equations.

Table 1
Participant baseline characteristics by vital status at visit 5

	Total	Attended visit 5	Alive, did not attend visit 5	Deceased by visit 5
N (%)	14,229	6,340 (45)	3,713 (26)	4,176 (29)
Age	57.0 (5.7)	55.1 (5.2)	57.4 (5.7)	59.8 (5.4)
Female, %	55.4	58.9	61.3	45.0
Black, %	24.7	22.6	22.9	29.5
HbA1c, %	5.8 (1.2)	5.6 (0.9)	5.7 (1.0)	6.2 (1.7)
Diabetes, %	13.6	7.5	12.2	24.2
Body mass index, kg/m ²	28.0 (5.4)	27.6 (5.1)	28.1 (5.5)	28.3 (5.8)
History of CHD, %	5.8	2.6	3.9	12.3
History of stroke, %	1.9	0.8	1.2	4.3
Hypertension, %	36.1	27.4	36.3	49.2
APOE e4 alleles, %				
0	69.1	71.1	68.9	66.3
1	28.2	26.6	28.5	30.5
2	2.6	2.3	2.6	3.2
Education, %				
Less than high school	21.9	14.7	23.5	31.4
High school	41.5	41.9	43.7	38.9
College/vocational	36.6	43.4	32.9	29.7
Smoking, %				
Current	22.4	16.2	21.0	33.1
Former	37.9	38.5	36.6	38.2
Never	39.7	45.3	42.4	28.7
Drinking, %				
Current	56.3	60.9	53.8	51.5
Former	21.2	17.1	21.0	27.6
Never	22.5	22.0	25.2	20.8
Measures of cognitive function				
Global Z	0.00 (1.00)	0.24 (0.93)	-0.01 (0.94)	-0.37 (1.04)
DWRT, words recalled	6.6 (1.5)	6.9 (1.5)	6.6 (1.5)	6.2 (1.6)
DSST, number completed	44.6 (14.2)	48.1 (13.5)	44.6 (13.4)	39.1 (14.4)
WFT, words generated	33.2 (12.5)	35.0 (12.2)	32.7 (12.1)	30.8 (12.8)
Suspected Dementia by Visit 5, %	10.3	4.2	15.6	14.7
CDR sum of boxes [*]	2.8 (4.7)	1.3 (2.4)	7.3 (6.8)	8.0 (6.9)
TICS-M	34.2 (7.4)	34.2 (7.7)	34.2 (7.3)	-

Values shown as % or mean (SD). All variables measured at visit 2 (1990-1992) except CDR, which was collect around the time of visit 5 (2011-2013) via contact with participants or a proxy. Suspected dementia was ascertained prior to visit 5 from hospitalization records with an ICD-9 code for dementia or from contact with participants or their proxy where cognitive impairment was indicated.

^{*} Available for a subset of participants, N=3741

Abbreviations: CHD, coronary heart disease; DWRT, delayed word recall test; DSST, digit symbol substitution test; WFT, word fluency test; CDR, clinical dementia rating; TICS-M, modified telephone interview for cognitive status.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2
Simulation results of estimated 20-year additional decline for persons with diabetes compared to those without, examining two imputation scenarios (imputing for participants living at visit 5 and living and deceased by visit 5) under two different analytic approaches (mixed and independence GEE) and four missingness mechanisms

Analytic approach and Scenario	“Truth” - Analyses in simulated cohort before implementing dropout/death				Imputation	
	Living Participants	Living and deceased participants	Available case	Living Participants	Living Participants	Living and deceased participants
Mixed						
1 MCAR	Mean effect (SE)	-0.236 (0.028)	-0.237 (0.024)	-0.235 (0.035)	-0.231 (0.034)	-0.229 (0.033)
	Empirical SE	0.0261	0.0202	0.0341	0.0322	0.0313
	Bias (%)	(reference)	-	-0.001 (0%)	-0.005 (2%)	-0.007 (3%)
	Bias (%)	-	(reference)	-0.001 (1%)	-0.005 (2%)	-0.007 (3%)
	CI coverage	(reference)	-	100%	100%	98%
	CI coverage	-	(reference)	100%	100%	98%
2 MAR						
2 MAR	Mean effect (SE)	-0.229 (0.031)	-0.237 (0.024)	-0.225 (0.042)	-0.221 (0.043)	-0.231 (0.044)
	Empirical SE	0.0309	0.0202	0.0402	0.0389	0.0403
	Bias (%)	(reference)	-	-0.004 (2%)	-0.008 (4%)	0.002 (-1%)
	Bias (%)	-	(reference)	-0.012 (5%)	-0.016 (7%)	-0.006 (2%)
	CI coverage	(reference)	-	99%	99%	100%
	CI coverage	-	(reference)	96%	93%	96%
3 MAR for MICE, MNAR for available case						
3 MAR for MICE, MNAR for available case	Mean effect (SE)	-0.215 (0.031)	-0.237 (0.024)	-0.182 (0.042)	-0.208 (0.044)	-0.229 (0.047)
	Empirical SE	0.0298	0.0202	0.0436	0.0448	0.0452
	Bias (%)	(reference)	-	-0.033 (16%)	-0.008 (4%)	0.014 (-7%)
	Bias (%)	-	(reference)	-0.055 (23%)	-0.029 (12%)	-0.007 (3%)
	CI coverage	(reference)	-	94%	98%	100%
	CI coverage	-	(reference)	73%	94%	94%
4 MNAR						
4 MNAR	Mean effect (SE)	-0.222 (0.028)	-0.237 (0.024)	-0.168 (0.035)	-0.174 (0.037)	-0.175 (0.037)
	Empirical SE	0.0249	0.0202	0.0341	0.0308	0.0307

Analytic approach and Scenario	“Truth” - Analyses in simulated cohort before implementing dropout/death					Imputation	
	Living Participants		Living and deceased participants		Available case	Living Participants	Living and deceased participants
GEE	Bias (%)	(reference)	-		-0.053 (24%)	-0.047 (21%)	-0.046 (21%)
	Bias (%)	-	(reference)		-0.068 (29%)	-0.062 (26%)	-0.061 (26%)
	CI coverage	(reference)	-		72%	85%	80%
	CI coverage	-	(reference)		51%	65%	66%
1 MCAR	Mean effect (SE)	-0.248 (0.034)	-0.249 (0.028)		-0.248 (0.045)	-0.231 (0.039)	-0.232 (0.036)
	Empirical SE	0.0386	0.0270		0.045	0.039	0.036
	Bias (%)	(reference)	-		0.000 (0%)	-0.017 (7%)	-0.015 (6%)
	Bias (%)	-	(reference)		-0.001 (0%)	-0.018 (7%)	-0.016 (7%)
2 MAR	CI coverage	(reference)	-		100%	100%	97%
	CI coverage	-	(reference)		99%	98%	99%
	Mean effect (SE)	-0.189 (0.039)	-0.249 (0.028)		-0.168 (0.055)	-0.196 (0.048)	-0.230 (0.043)
	Empirical SE	0.0385	0.0270		0.0523	0.0462	0.0404
3 MAR for MICE, MNAR for available case	Bias (%)	(reference)	-		-0.020 (11%)	0.008 (-4%)	0.042 (-22%)
	Bias (%)	-	(reference)		-0.080 (32%)	-0.052 (21%)	-0.019 (8%)
	CI coverage	(reference)	-		99%	100%	88%
	CI coverage	-	(reference)		72%	72%	93%
4 MNAR	Mean effect (SE)	-0.168 (0.039)	-0.249 (0.028)		-0.113 (0.053)	-0.164 (0.049)	-0.225 (0.046)
	Empirical SE	0.0386	0.0270		0.0516	0.0513	0.0473
	Bias (%)	(reference)	-		-0.055 (33%)	-0.004 (2%)	0.057 (-34%)
	Bias (%)	-	(reference)		-0.136 (55%)	-0.085 (34%)	-0.024 (10%)
5 MNAR	CI coverage	(reference)	-		89%	97%	75%
	CI coverage	-	(reference)		27%	46%	91%
	Mean effect (SE)	-0.225 (0.035)	-0.249 (0.028)		-0.147 (0.045)	-0.168 (0.040)	-0.185 (0.037)
	Empirical SE	0.0366	0.0270		0.0494	0.0435	0.0395
6 MNAR	Bias (%)	(reference)	-		-0.078 (35%)	-0.057 (25%)	-0.040 (18%)
	Bias (%)	-	(reference)		-0.078 (35%)	-0.057 (25%)	-0.040 (18%)

Analytic approach and Scenario	“Truth” - Analyses in simulated cohort before implementing dropout/death				Imputation	
	Living Participants	Living and deceased participants	Available case	Living Participants	Living and deceased participants	
Bias (%)	-	(reference)	-0.102 (41%)	-0.080 (32%)	-0.063 (26%)	
CI coverage	(reference)	-	62%	89%	82%	
CI coverage	-	(reference)	40%	61%	61%	

Values in the “truth” columns are calculated using the entire simulated cohort before implementing the death and dropout censoring of cognitive scores. The “living participants” column excludes the simulated cognitive trajectory of persons who die between visits 4 and 5. The “living and deceased” participants column includes all simulated cognitive trajectories in the estimation. Mean effect is an average of 100 simulations. The standard error (SE) of the mean effect is the square root of mean variances across 100 simulations. Bias is calculated as the mean effect estimate from each method (available case, imputation in living participants, imputation in both living and deceased participants) minus the mean effect estimate from the truth. Negative values indicate underestimation of the true effect and positive values represent overestimation. Bias % is calculated as the estimated bias divided by the true effect (e.g. $0.007/0.237 = 3\%$). The empirical SE is the standard deviation of the mean effect across 100 simulations. CI coverage is the percentage of the simulations where the confidence interval for the estimated effect includes the true effect.

Mixed: estimates are from mixed-effects models using time since baseline as the time axis, modeled using a spline term with a knot at 6 years, the median time between visits 2 and 4. Random effects were random intercept and two random slopes, one for each time spline term, and were assumed to be independent.

GEE: estimates are from models fit using generalized estimating equations with independent working correlation and robust variance estimation. **Scenario 1:** Death and dropout simulated to be missing completely at random, with probabilities of 0.29 and 0.37, respectively, chosen to match proportions observed in ARIC. **Scenario 2:** Death and dropout simulated to depend on prior visit global Z score, diabetes, hypertension, and smoking status. **Scenario 3:** Death and dropout simulated to depend on prior visit global Z score, diabetes, hypertension, current smoking status, and suspected dementia. As a result, the “complete case”, which uses a mixed model, is MNAR (suspect dementia not included in the mixed model), but MICE is MAR (suspect dementia is included for imputation). This scenario is more consistent with what we believe the true missingness pattern in ARIC to be. **Scenario 4:** Dropout simulated to depend on visit 5 global Z scores (i.e. unobserved scores), and death simulated to be missing completely at random.