International Journal of Epidemiology, 2018, 1507–1517
doi: 10.1093/ije/dyy124
Advance Access Publication Date: 13 July 2018
Original article

Cognitive and Behavioural Outcomes

Does selective survival before study enrolment attenuate estimated effects of education on rate of cognitive decline in older adults? A simulation approach for quantifying survival bias in life course epidemiology

Elizabeth Rose Mayeda,[1,2]* Teresa J Filshtein,[2,3] Yorghos Tripodis,[4] M Maria Glymour[2] and Alden L Gross[5]

[1]Department of Epidemiology, University of California, Los Angeles Fielding School of Public Health, Los Angeles, CA, USA, [2]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, USA, [3]Department of Statistics, University of California, Davis, Davis, CA, USA, [4]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA and [5]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

*Corresponding author. Department of Epidemiology, University of California, Los Angeles Fielding School of Public Health, Box 951772, 46-070B CHS Los Angeles, CA 90095-1772, USA. E-mail: ermayeda@ph.ucla.edu

Editorial decision 21 May 2018; Accepted 29 May 2018

Abstract

Background: The relationship between education and late-life cognitive decline is controversial. Selective survival between early life, when education is typically completed, and late life, when cognitive ageing studies take place, could attenuate effect estimates.

Methods: We quantified potential survival bias (collider-stratification bias) in estimation of the effect of education on late-life cognitive decline by simulating hypothetical cohorts of 20-year-olds and applying cumulative mortality from US life tables. For each of four causal scenarios (2000 replications each), we compared the estimated versus causal effect of education on cognitive decline over 9 years, starting at age 60, 75 or 90 in random samples of $n = 2000$ people who survived to each age.

Results: Effects of education on cognitive decline were underestimated when both education and *U*, another determinant of cognitive decline, influenced mortality (collider-stratification bias). The magnitude of bias was sensitive to the magnitude of the effect of *U* on cognitive decline and whether there was a multiplicative interaction between education and *U* on mortality. For example, when there was a multiplicative interaction between education and *U* on mortality, 95% confidence interval coverage of the causal effect ranged from 83.4% to 50.4% at age 60 and 25.8% to 0.2% at age 90.

Conclusions: Selective survival could lead to underestimation of effects of education on late-life cognitive decline. Our simulations map survival bias to testable assumptions about underlying causal structures.

Key words: Life course, education, cognitive decline, survival bias, selection bias, collider-stratification bias, simulation

Key Messages

- Selective survival before study enrolment could lead to underestimation of effects of education on late-life cognitive decline due to mortality between the ages when education is completed and when studies of cognitive ageing typically begin.
- In simulations quantifying potential survival bias, the magnitude of bias was large only when there was a multiplicative interaction between exposure (i.e. education) and *U*, a common cause of mortality, and the outcome (i.e. late-life cognitive decline). The magnitude of bias was more sensitive to the magnitude of the effect of *U* on the outcome than the magnitude of the effect of *U* on mortality.
- Selective survival before study initiation may bias a broad range of life course epidemiology studies. With relevant information on survival processes, simulations can be used to quantify the plausible range of bias.

Introduction

Education is associated with better health across numerous outcomes[1,2] and is thought to be important for cognitive health.[3] Education is associated with lower risk of Alzheimer's disease and related dementias,[4] but longitudinal studies suggest that this association may be driven by the link between education and level of cognitive function, not rate of cognitive decline.[5–10] This distinction is critical because cognitive decline is the hallmark of progression of Alzheimer's disease and related dementias. Although it is possible that education has no effect on cognitive decline, selective survival is an alternative explanation. Selective survival between early life, when education is typically completed, and late life, when studies of cognitive ageing typically begin, could lead to underestimation of possible effects of education on cognitive decline because education is associated with survival.[11,12] Recent studies estimating effects of education on cognitive decline have used methods to address selective survival after study enrolment,[7,8] but few have available data to account for selective survival between early life and study initiation.

We used simulations to systematically quantify potential bias in estimation of the effect of education on late-life rate of cognitive decline arising from selective survival between early and late life under several causal structures consistent with collider-stratification bias.[13,14] Our simulations focused on education and cognitive decline, but survival bias is a potential source of bias for a broad range of exposures and outcomes of importance in life course epidemiology.

Methods

Hypothetical cohort study

We considered a hypothetical cohort study of the effect of completion of high school education on late-life cognitive decline. The hypothetical population cohort comprised $n = 100\,000$ 20-year-olds (when high school education is typically completed). The association between completion of high school education and cognitive decline was estimated among people who survived to late life (ages 60, 75 and 90) under four causal scenarios (described below). For each scenario, we selected random samples of $n = 2000$ people surviving to age 60 (Age 60 Study), 75 (Age 75 Study) or 90 (Age 90 Study) for inclusion in hypothetical cognition studies. In each hypothetical study, cognitive function was measured on seven occasions over 9 years, a typical sample size and assessment schedule for cognitive ageing studies.[9,15] We considered three age cohorts because most studies recruit a range of ages, and the magnitude of survival bias arising from collider stratification may differ by enrolment age.

Education and survival distributions for the hypothetical cohort were based on estimates for people born in the USA around 1930: we simulated the data such that 60% of the $n = 100\,000$ individuals in the 20-year-old cohort completed high school (60% ≥high school; 40% <high school)[16] and 69%, 35% and 3% of the 20-year-old cohort survived to ages 60, 75 and 90, respectively.[17] Lower education was associated with higher cumulative mortality at all ages (odds ratio = 2.0 for <high school versus ≥high

school).[12] We set the average rate of cognitive decline in the ≥high school group to $\beta_{10} = -0.05$ standardized [N(0, 1)] units annually. We set the total effect of education (<high school versus ≥high school) on cognitive decline to −0.05 standardized units annually (i.e. <high school doubled rate of cognitive decline relative to ≥high school). To focus on potential bias arising from selective survival between exposure and outcome assessment, we assumed no other form of selection into the analysis cohort, no attrition after study enrolment, no measurement error and no confounding.

Education influences mortality, so some people who survive with ≥high school education would not survive if they had <high school education; the effect of education on cognitive ageing is considered undefined for these individuals because one of the counterfactuals cannot be defined. We therefore focus on the Survivor Average Causal Effect, i.e. the effect of education on cognitive decline among individuals who would survive (to the age at which cognitive ageing was evaluated) regardless of education level.[18–20]

Data-generating process

The causal structures guiding our data-generating processes are shown in Figure 1. For clarity, we posit specific constructs represented by the unmeasured determinants of cognitive decline (the '*U*' variables), but numerous factors are imaginable. We assigned each person *i*: (i) education level$_i$ (0 if ≥high school, 1 if <high school) from Bernoulli(0.40); (ii) $U1_i$, which we refer to as an Alzheimer's disease polygenic risk score (AD-GRS),[21,22] $U1_i \sim$N(0, 1); and (iii) $U2_i$, which we refer to as complexity of primary lifetime occupation (occupational complexity),[23,24] generated as a consequence of education:

$$U2_i = 0 + \alpha_o \text{education}_i + e_i \quad (1),$$

where α_0 = effect of education on occupational complexity and $e_i \sim$N(0, 1). For consistency in the direction of effects of variables in our simulations, *U2* is coded such that higher values are harmful.

We generated survival to ages 60, 75 and 90 for each person as a function of education, *U1* and *U2*. First, we generated probability of death for each person *i* by age *k* using equation 2:

$$\text{P(death}_{ik}) = \frac{\exp(\gamma_{0k} + \gamma_1 \text{education}_i + \gamma_2 U1_i + \gamma_3 U2_i + \gamma_4 U1_i * \text{education}_i)}{1 + \exp(\gamma_{0k} + \gamma_1 \text{education}_i + \gamma_2 U1_i + \gamma_3 U2_i + \gamma_4 U1_i * \text{education}_i)} \quad (2)$$

Next, we generated probability of survival as P(survival$_{ik}$) = 1 − P(death$_{ik}$) and vital status as a binary variable with mean P(survival$_{ik}$). For each person *i* at cognitive assessment wave *j*, we generated a value of cognitive function, C$_{ij}$, following a linear mixed-effects 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 Table 1). Although we use the term time$_{ij}$, time at wave *j* does not vary between people:

$$C_{ij} = \beta_{00} + \beta_{01}\text{education}_i + \beta_{02}U1_i + \beta_{03}U2_i + (\beta_{10} + \beta_{11}\text{education}_i + \beta_{12}U1_i + \beta_{13}U2_i) * \text{time}_{ij} + \zeta_{0i} + \zeta_{1i}\text{time}_{ij} + \varepsilon_{ij} \quad (3)$$

Note that the data-generating model for cognitive function specifies an indirect effect of education on cognitive change via *U2* (β_{13}). Thus, the total effect of education on cognitive change is $\beta_{11} + \alpha_0 * \beta_{13}$.

Causal scenarios under investigation

We considered four causal structures. In all structures, education influenced survival to late life (age 60, 75 or 90) ($\exp(\gamma_1) = 2.0$). Causal Structure 1 (Figure 1a) was a base scenario with no anticipated bias. In this scenario, no correlate of cognitive decline affected survival. In Causal Structures 2 and 3 (Figure 1b), we added an effect of *U1* (AD-GRS) on survival. In Causal Structure 2 (collider stratification), there was no interaction on the odds ratio scale between education and *U1* in determining mortality ($\exp(\gamma_2) = 2.0$; $\exp(\gamma_4) = 1.0$). In Causal Structure 3 (collider stratification with interaction), education and *U1* interacted on the odds ratio scale in determining mortality, such that *U1* only influenced mortality for people with <high school education ($\exp(\gamma_2) = 1.0$; $\exp(\gamma_4) = 2.0$). In Causal Structure 4 (collider stratification with mediation; Figure 1c), *U1* had no effect on mortality, but *U2* (occupational complexity), which partially mediated the effect of education on cognitive decline, influenced mortality ($\exp(\gamma_3) = 2.0$). Causal Structures 2 and 4 were similar; the distinction between them is that *U1* was not affected by education and *U2* partially mediated the effect of education on cognitive decline.

Input parameter values

In all causal structures, we maintained the magnitude of effects of education on cognitive decline and cumulative mortality described earlier. For each causal structure, we first generated data with the moderate parameter inputs displayed in Table 2, selecting values that represent effect sizes within the bounds of associations observed in published data. Note that all variables are coded such that higher values are harmful for cognitive outcomes and

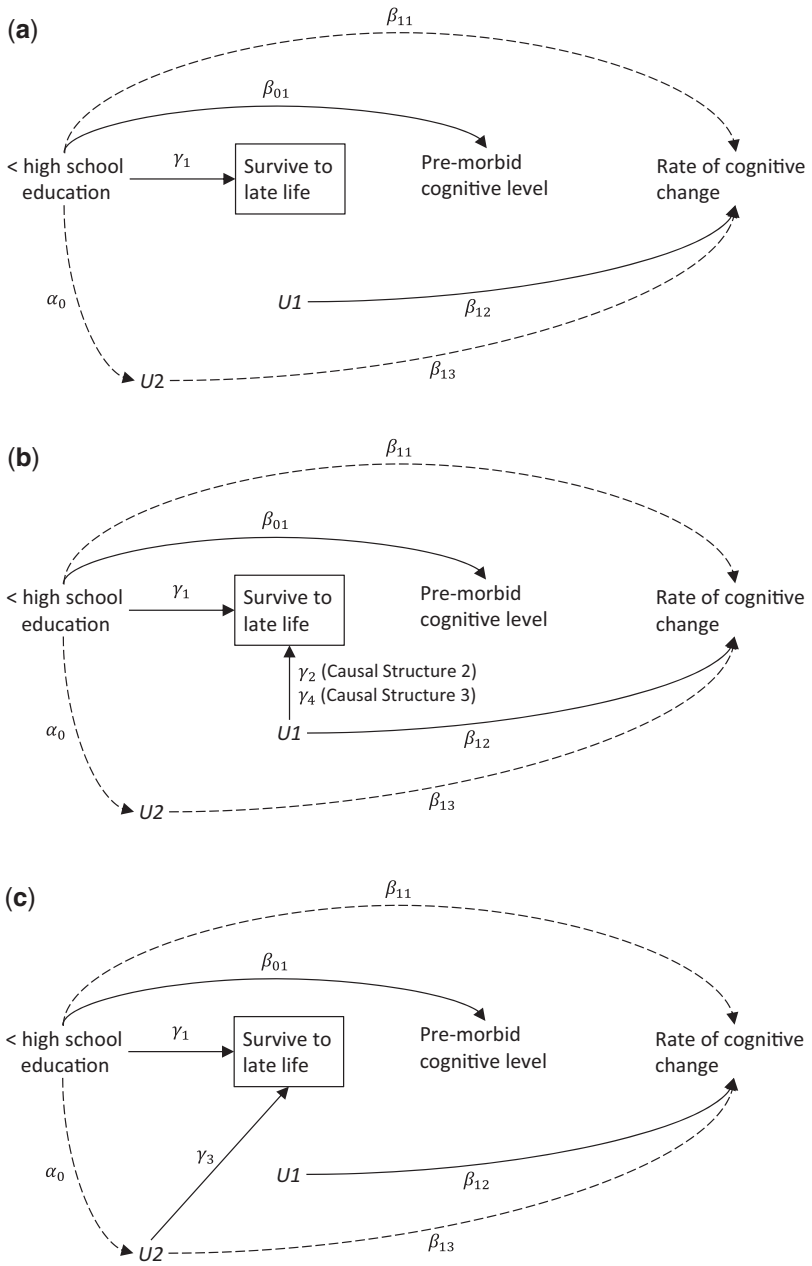

Figure 1. Causal diagrams illustrating causal structures under investigation. Gamma coefficients for effects on mortality and beta coefficients for effects on cognitive outcomes refer to equations 2 and 3. In all scenarios, education influenced survival to late life (age 60, 75 or 90). Causal Structure 1 (no bias anticipated): no determinant of cognitive decline other than education affected mortality. Causal Structures 2 (collider stratification) and 3 (collider stratification with interaction): *U1*, which we conceptualized as Alzheimer's disease polygenetic risk score, influenced mortality; in Causal Structure 2, there was no interaction between education and *U1* on mortality; in Causal Structure 3, there was a multiplicative interaction between education and *U1* on mortality. Causal Structure 4 (collider stratification with mediation): *U2*, which we conceptualized as occupational complexity, partially mediated the effect of education on rate of cognitive decline and mortality. Note that all variables are coded such that higher values are harmful for cognitive outcomes and survival.

Table 1. Input variance and covariance values for generating C_{ij} (cognitive function)

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

Table 2. Moderate input parameters. Shaded columns represent parameters for which alternative input values are considered (β_{12}, β_{13}, γ_2, γ_3, γ_4)

	U2	Cognitive function								Cumulative mortality				
Causal Structure[a]	α_0	β_{00}	β_{01}	β_{02}	β_{03}	β_{10}	β_{11}[b]	β_{12}	β_{13}	γ_{0k}[c]	γ_1	γ_2	γ_3	γ_4
1: no bias anticipated	0.50	0	−0.500	0	0	−0.050	−0.0375	−0.025	−0.025	Varies	ln(2)	0	0	0
2: collider stratification	0.50	0	−0.500	0	0	−0.050	−0.0375	−0.025	−0.025	Varies	ln(2)	ln(2)	0	0
3: collider stratification with interaction	0.50	0	−0.500	0	0	−0.050	−0.0375	−0.025	−0.025	Varies	ln(2)	0	0	ln(2)
4: collider stratification with mediation[d]	0.50	0	−0.500	0	0	−0.050	−0.0375	−0.025	−0.025	Varies	$\ln(\sqrt{2})$	0	ln(2)	0

[a]All causal structures are variations of the general data generation structure, which are:

- Generate $U1_i$ (Alzheimer's disease polygenic risk score) for each person i from $U1_i \sim N(0, 1)$.
- Generate $U2_i$ (occupational complexity, coded such that higher values are harmful) for each person i: $U2_i = 0 + \alpha_o \text{education}_i + e_i$ (equation 1).
- Generate death for each person i by age k: $P(\text{death}_{ik}) = \frac{\exp(\gamma_{0k}+\gamma_1 \text{education}_i+\gamma_2 U1_i+\gamma_3 U2_i+\gamma_4 U1_i*\text{education}_i)}{1+\exp(\gamma_{0k}+\gamma_1 \text{education}_i+\gamma_2 U1_i+\gamma_3 U2_i+\gamma_4 U1_i*\text{education}_i)}$ (equation 2).
- Generate cognitive function for each person i at wave j: $C_{ij} = \beta_{00} + \beta_{01}\text{education}_i + \beta_{02}U1_i + \beta_{03}U2_i + (\beta_{10} + \beta_{11}\ \text{education}_i + \beta_{12}U1_i + \beta_{13}U2_i)*\text{time}_{ij} + \zeta_{0i} + \zeta_{1i}\text{time}_{ij} + \varepsilon_{ij}$ (equation 3).
- Total effect of education on annual rate of cognitive change: $\beta_{11} + \alpha_0 * \beta_{13} = -0.0375 + 0.50 * -0.025 = -0.050$.

[b]When alternative values of β_{13} (effect of *U2* on cognitive change in equation 3) were considered, we varied β_{11} (direct effect of education cognitive change in equation 3) to maintain total effect of education on annual rate of cognitive change −0.05.

[c]Values for γ_{0k} (log odds of mortality for people with reference values for all covariates at age k in equation 2) varied to maintain the same cumulative mortality at ages 60, 75 or 90 across causal structures: 69% survival to age 60, 35% survival to age 75 and 3% survival to age 90.

[d]In Causal Structure 4, effects of *U2* on mortality were introduced. Since education influenced *U2*, we modified γ_1 (direct effect of education on mortality) from ln(2) to $\ln(\sqrt{2})$ in this causal structure to maintain the same total effect of education on mortality as in other causal structures (ln(2)).

survival. Next, we examined more aggressive parameter inputs (varying one input at a time), selecting values we consider on the higher bound of effect sizes likely in real data. We varied the magnitudes of: (i) the effect of *U1* on cognitive change (β_{12}); (ii) the effect of *U2* on cognitive change (β_{13}); (iii) the effect of *U1* on mortality (γ_2); (iv) multiplicative effects of education and *U1* on mortality (γ_4); and (v) the effect of *U2* on mortality (γ_3).

Assessment of survival bias in estimated effect of education on annual rate of cognitive change

We estimated the total effect of education on cognitive change ($\hat{\beta}$) in the Age 60, 75 and 90 Study samples (i.e. random samples of $n = 2000$ survivors to ages 60, 75 and 90) using linear mixed-effects models with random intercepts and slopes allowing for correlation of the random intercepts and random slopes, and no additional within-person covariance structure. Models estimating the education-cognitive decline association do not adjust for *U1* or *U2* because, in practice, such measures may be unavailable.

We generated $S = 2000$ iterations of sample generation for each causal scenario. In our notation, β represents the causal and $\hat{\beta}$ represents the estimated total effect of education on annual rate of cognitive change. We calculated the mean estimate across the S simulated samples, $\bar{\hat{\beta}} = \sum_{b=1}^{S} \frac{\hat{\beta}_b}{S}$, and calculated percentage bias as $\left(\frac{\bar{\hat{\beta}} - \beta}{\beta}\right) * 100$. We assessed variability in estimates with the empirical standard error, the standard deviation of the estimates across the samples, as $\sqrt{\frac{1}{(S-1)}\sum_{b=1}^{S}(\hat{\beta}_b - \bar{\hat{\beta}})^2}$. We calculated 95% confidence interval (CI) coverage as the percentage of samples in which the 95% CI for $\hat{\beta}$ included β. We generated boxplots of $\hat{\beta}$ across samples.

For causal structures consistent with collider-stratification bias (Causal Structures 2–4), potential bias is driven by non-causal associations between education and the '*U*' variables among survivors to late life, which consequently confound the education-cognitive decline association. To illustrate this, we plotted histograms of mean *U1* (Causal Structures 2 and 3) and *U2* (Causal Structure 4) by education level at age 20 and in the Age 60, 75 and 90 Studies.

Simulations were carried out in R version 3.2.4. Simulation code and documentation are available online [https://github.com/ermayeda/education_cognitive_decline_simulation].

Results

In Causal Structure 1 (no bias anticipated), the estimated total effect of education on annual rate of cognitive change (β) was unbiased, as expected, with effect estimates centred around -0.05 standardized units and approximately correct 95% CI coverage (i.e. ~95% of the 95% CIs included the causal effect, $\beta = -0.05$) for all age groups (Table 3).

Table 3. Simulation results from Causal Structures 1–4 for the estimated total effect of education (<high school versus ≥high school) on annual rate of cognitive decline, based on input parameters from Table 1 (moderate input parameters)

Causal Structure	Age 60 Study				Age 75 Study				Age 90 Study			
	$\bar{\hat{\beta}}$[a]	% bias[b]	Empirical SE[c]	95% CI coverage[d]	$\bar{\hat{\beta}}$[a]	% bias[b]	Empirical SE[c]	95% CI coverage[d]	$\bar{\hat{\beta}}$[a]	% bias[b]	Empirical SE[c]	95% CI coverage[d]
1: no bias anticipated	−0.050	0.7	0.006	95.4	−0.050	0.3	0.006	94.7	−0.050	0.6	0.007	95.8
2: collider stratification	−0.048	−4.0	0.006	94.2	−0.048	−4.7	0.006	93.8	−0.050	−1.0	0.007	94.6
3: collider stratification with interaction	−0.044	−12.5	0.006	83.4	−0.038	−23.1	0.006	55.6	−0.033	−33.5	0.006	25.8
4: collider stratification with mediation	−0.048	−4.5	0.006	93.7	−0.048	−4.1	0.007	93.2	−0.050	−0.7	0.007	94.7

[a]$\bar{\hat{\beta}}$ is the mean of the estimated total effect of education on rate of cognitive decline across the S = 2000 simulated samples.
[b]% bias is the bias as a percentage of the total causal effect of education on annual rate of cognitive change in simulations (β = -0.05).
[c]Empirical standard error (SE) is the standard deviation of the estimates across the S = 2000 samples.
[d]95% confidence interval (CI) coverage is the percentage of simulations in which the 95% confidence interval for $\hat{\beta}$ included β.

The estimated total effect of education on cognitive decline was biased toward the null to varying degrees in all other causal structures (Table 3, Figure 2). In Causal Structure 2 (collider stratification: education and *U1* influence mortality), the magnitude of bias with moderate input parameters was relatively small: the total effect of education on cognitive decline was underestimated by 4.0% in the Age 60 Study and 4.7% in the Age 75 Study. The bias was negligible (−1.0%) in the Age 90 Study. The 95% CI coverage was approximately correct in all three age groups. Results were similar for Causal Structure 4 (collider stratification with mediation: education and *U2* influence mortality).

In Causal Structure 3 (collider stratification with interaction: multiplicative interaction between education and *U1* on mortality), the magnitude of bias with moderate input parameters was considerably larger and increased monotonically with age of the study sample (i.e. higher cumulative mortality). The 95% CI coverage ranged from 83% in the Age 60 Study to 26% in the Age 90 Study.

The magnitude of bias increased with more aggressive input parameters in causal structures consistent with collider stratification bias (Causal Structures 2, 3 and 4) (Table 4). The magnitude of bias was particularly sensitive to the magnitude of the effects of the '*U*' variables on cognitive decline (*U1* in Causal Structures 2 and 3; *U2* in Causal Structure 4). However, the bias remained negligible in the Age 90 Study under Causal Structures 2 and 4.

The bias in Causal Structures 2 (collider stratification) and 3 (collider stratification with interaction) was driven by selective survival of people with more protective (lower) values of *U1*, the variable we conceptualized as AD-GRS (Figure 3a). AD-GRS distributions did not differ by education at age 20. In Causal Structure 2 (collider stratification), AD-GRS influenced mortality, so people surviving to late life tended to have more protective AD-GRS. Among survivors to ages 60 and 75, this was especially true for the <high school education group, but among survivors to age 90, there was little difference in the distribution of AD-GRS by education, resulting in minimal bias. In Causal Structure 3 (collider stratification with interaction), AD-GRS only influenced mortality for people with <high school education; this caused the AD-GRS distribution to shift to more protective values with age only for the <high school education group, and the difference in distribution of AD-GRS by education (and magnitude of bias) increased monotonically with age.

The bias in Causal Structure 4 (collider stratification with mediation) was driven by selective survival of people with more protective values of *U2*, the variable we conceptualized as main lifetime occupational complexity (Figure 3b). People with more protective occupations were more likely to survive to late life, although similar to Causal Structure 2, the difference in distribution of occupational complexity by education (and magnitude of bias) was non-monotonic with age.

Discussion

We used simulations to quantify potential bias in estimation of the effect of education on rate of cognitive decline in older adults arising from selective survival between early life, when education is typically completed, and late life, when studies of cognitive ageing typically begin. In our simulations, lower educational attainment had a harmful effect on late-life cognitive decline. We examined several causal structures consistent with collider-stratification bias in hypothetical cohorts established at various ages in late life. In all causal structures, all determinants of cognitive

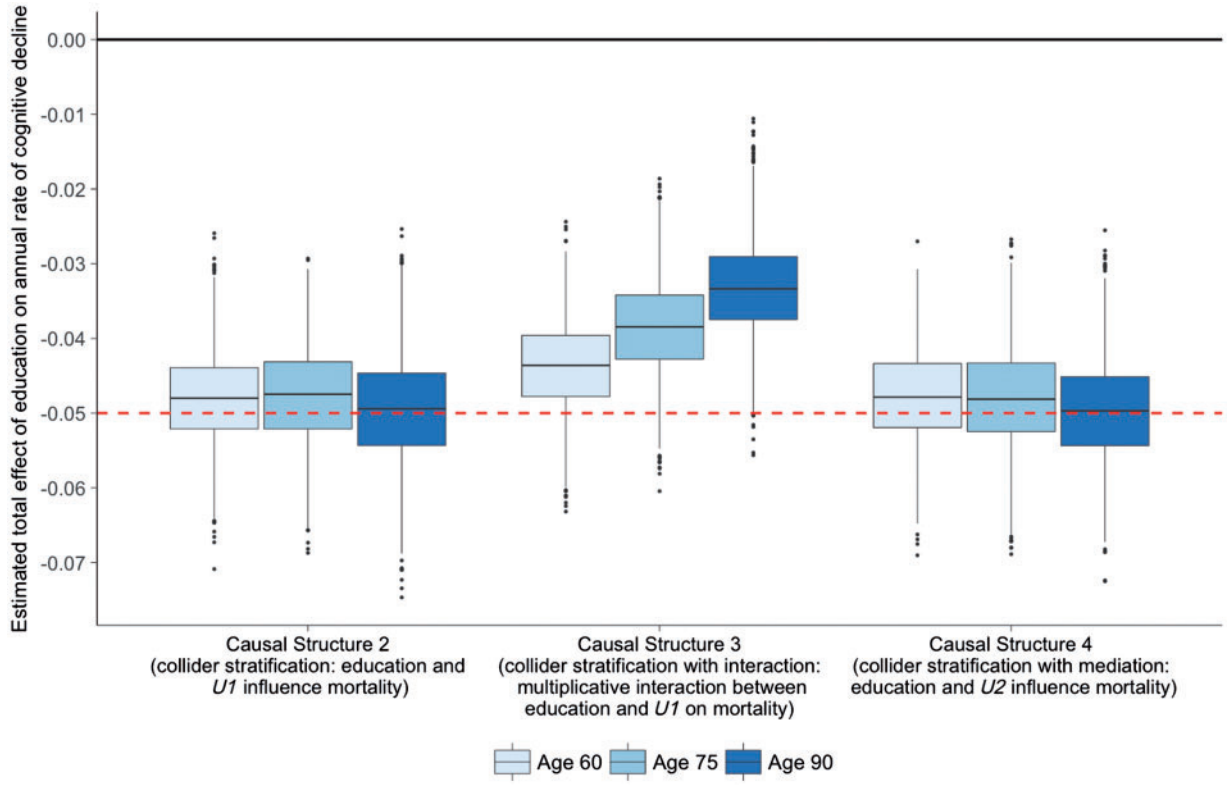

Figure 2. Boxplots of estimated total effect of education on annual rate of cognitive decline ($\widehat{\beta}$) across S = 2000 simulated samples with moderate input parameters. The box indicates the median and first and third quartiles of estimates across simulated samples. The whiskers indicate the lowest and highest estimates within 1.5*(interquartile range) and the dots indicate the outliers. The dotted line indicates the total causal effect of education on annual rate of cognitive decline in our simulations.

decline and survival were harmful and the bias was toward the null. We varied the magnitude of effects in our causal structures, but did not consider effects of opposite signs because harmful effects seemed most plausible to us; however, collider-stratification bias would differ substantially if the signs of effects in our causal structures were reversed, for example if we simulated a '*U*' variable that benefited survival but had deleterious effects on cognitive decline. The magnitude of bias was sensitive to how the causal structure was parameterized, particularly to multiplicative interactions between education and '*U*' variables on mortality and the magnitude of the effects of '*U*' variables on cognitive decline. We expected higher cumulative mortality (older age at follow-up) to increase bias, but this was only the case in Causal Structure 3 (collider stratification with interaction).

The magnitude of bias was greatest in the causal structure with a multiplicative interaction between education and AD-GRS (*U1*) on mortality (Causal Structure 3, Figure 1b). This is consistent with previous research documenting that the magnitude of collider-stratification bias tends to be larger when there is sufficient multiplicative interaction between the exposure and a collider-inducing variable (i.e. a common cause of selection and the outcome) in their influence on selection than in comparable causal structures in the absence of this interaction.[25–30] Although this is expected on average, other features of the causal structure, such as an interaction between the collider-inducing variable and exposure in their influence on the outcome, could offset this in specific situations. The plausibility of Causal Structure 3 can be difficult to determine, since *U1* may be unmeasured. However, the interaction we specified in our simulations was fairly large (odds ratio = 2.0 for effect of 1-unit higher *U1* in <high school group; no effect in ≥high school group). Evaluating the range of plausible interactions to guide honest bias analyses is an important domain for researchers focusing on predictors of survival. In causal structures featuring collider stratification without interaction (Causal Structures 2 and 4), the magnitude of bias was modest and agnostic to whether the '*U*' variable was on the causal pathway from education and cognitive decline. However, options for mitigating bias in these scenarios differ. When the '*U*' variable is not on the causal

Table 4. Simulation results for Causal Structures 2, 3 and 4 (causal structures consistent with collider-stratification bias) for the estimated total effect of education (<high school versus ≥high school) on annual rate of cognitive decline, based on alternative input values (more aggressive effect sizes)

	Age 60 Study				Age 75 Study				Age 90 Study			
	$\bar{\hat{\beta}}$[a]	% bias[b]	Empirical SE[c]	95% CI coverage[d]	$\bar{\hat{\beta}}$[a]	% bias[b]	Empirical SE[c]	95% CI coverage[d]	$\bar{\hat{\beta}}$[a]	% bias[b]	Empirical SE[c]	95% CI coverage[d]
Causal Structure 2 (collider stratification: education and *U1* influence mortality)												
Moderate input parameters (Table 3)	−0.048	−4.0	0.006	94.2	−0.048	−4.7	0.006	93.8	−0.050	−1.0	0.007	94.6
More aggressive β_{12}[e]	−0.046	−8.1	0.007	90.3	−0.045	−9.5	0.007	89.3	−0.049	−1.8	0.007	95.2
More aggressive γ_2[f]	−0.047	−5.4	0.006	92.8	−0.047	−6.8	0.006	91.2	−0.049	−2.2	0.007	94.6
Causal Structure 3 (collider stratification with interaction: multiplicative interaction between education and *U1* on mortality)												
Moderate input parameters (Table 3)	−0.044	−12.5	0.006	83.4	−0.038	−23.1	0.006	55.6	−0.033	−33.5	0.006	25.8
More aggressive β_{12}[e]	−0.037	−25.3	0.007	50.4	−0.027	−46.5	0.007	6.8	−0.017	−67.0	0.007	0.2
More aggressive γ_4[g]	−0.041	−18.2	0.006	70.5	−0.034	−32.1	0.006	28.7	−0.025	−51.0	0.006	1.5
Causal Structure 4 (collider stratification with mediation: education and *U2* influence mortality)												
Moderate input parameters (Table 3)	−0.048	−4.5	0.006	93.7	−0.048	−4.1	0.007	93.2	−0.050	−0.7	0.007	94.7
More aggressive β_{13}[h]	−0.044	−12.6	0.007	85.6	−0.043	−14.0	0.007	84.0	−0.048	−3.4	0.008	93.7
More aggressive γ_3[i]	−0.047	−5.4	0.006	93.8	−0.047	−6.1	0.006	92.5	−0.049	−2.0	0.007	94.4

[a] $\bar{\hat{\beta}}$ is the mean of the estimated total effect of education on rate of cognitive decline across the S = 2000 simulated samples.
[b] % bias is the bias as a percentage of the total causal effect of education on rate of cognitive change in simulations (β = -0.05).
[c] Empirical standard error (SE) is the standard deviation of the estimates across the S = 2000 samples.
[d] 95% confidence interval (CI) coverage is the percentage of simulations in which the 95% confidence interval for $\hat{\beta}$ included β.
[e] Changed β_{12} (effect of *U1* on cognitive change in equation 3) from -0.025 to -0.050; held other parameter inputs constant.
[f] Changed γ_2 (effect of *U1* on mortality in equation 2) from ln(2) to ln(3); held other parameter inputs constant.
[g] Changed γ_4 (multiplicative effects of education and *U1* on mortality in equation 2) from ln(2) to ln(3); held other parameter inputs constant.
[h] Changed β_{13} (effect of *U2* on cognitive change in equation 3) from -0.025 to -0.075; to maintain the same total effect of education on rate of cognitive change as in other scenarios (-0.05), we changed β_{11} (direct effect of education cognitive change in equation 3) from -0.0375 to -0.0125; held other parameter inputs constant.
[i] Changed γ_3 (effect of *U2* on mortality in equation 2) from ln(2) to ln(3); to maintain the same total effect of education on mortality as in other scenarios (ln(2)), we changed γ_1 (direct effect of education on mortality in equation 2) from $\ln(\sqrt{2})$ to $\ln\left(\frac{2}{\sqrt{3}}\right)$; held other parameter inputs constant.

pathway of interest, researchers can control for bias by adjusting for the '*U*' variable in regression models. In contrast, if the '*U*' variable is on the causal pathway of interest, adjusting for the '*U*' variable risks introducing overadjustment bias.[31]

By varying input parameters one by one, we identified parameters with the greatest impact on bias. The magnitude of bias was particularly sensitive to the magnitude of the effects of the '*U*' variables on cognitive decline. Effects of the '*U*' variables on mortality were less influential. Intuitively, this is because collider stratification bias was driven by the education-'*U*' association induced among survivors to old age, resulting in a spurious education-cognitive decline association. The magnitude of bias is constrained by both the education-'*U*' association (induced by collider stratification) and the '*U*'-cognitive decline association (due to a causal effect of '*U*' on cognitive decline). In general, associations induced by collider stratification are not as large as effects of either cause on the collider.[32] Therefore, increasing the effect of '*U*' on mortality does not increase the '*U*'-education association proportionally. In contrast, the magnitude of the '*U*'-cognitive decline association is proportional to the bias induced by the selection process.

We expected higher cumulative mortality (older age at follow-up) to increase bias, but this was not the case in all causal structures. In causal structures with collider-stratification bias without interaction, the magnitude of bias was negligible in the Age 90 Study. As Figure 3 illustrates, this arose because survivors to age 90 (3% of the population alive at age 20) were so highly selected that virtually everyone had extremely protective values of '*U*' variables, regardless of education. Consequently, the '*U*' variables do not confound the education-cognitive decline association in this age group. This extreme selection process may contribute to findings that many typical risk factors are not associated with cognitive and other health outcomes in the oldest-old.[33–36]

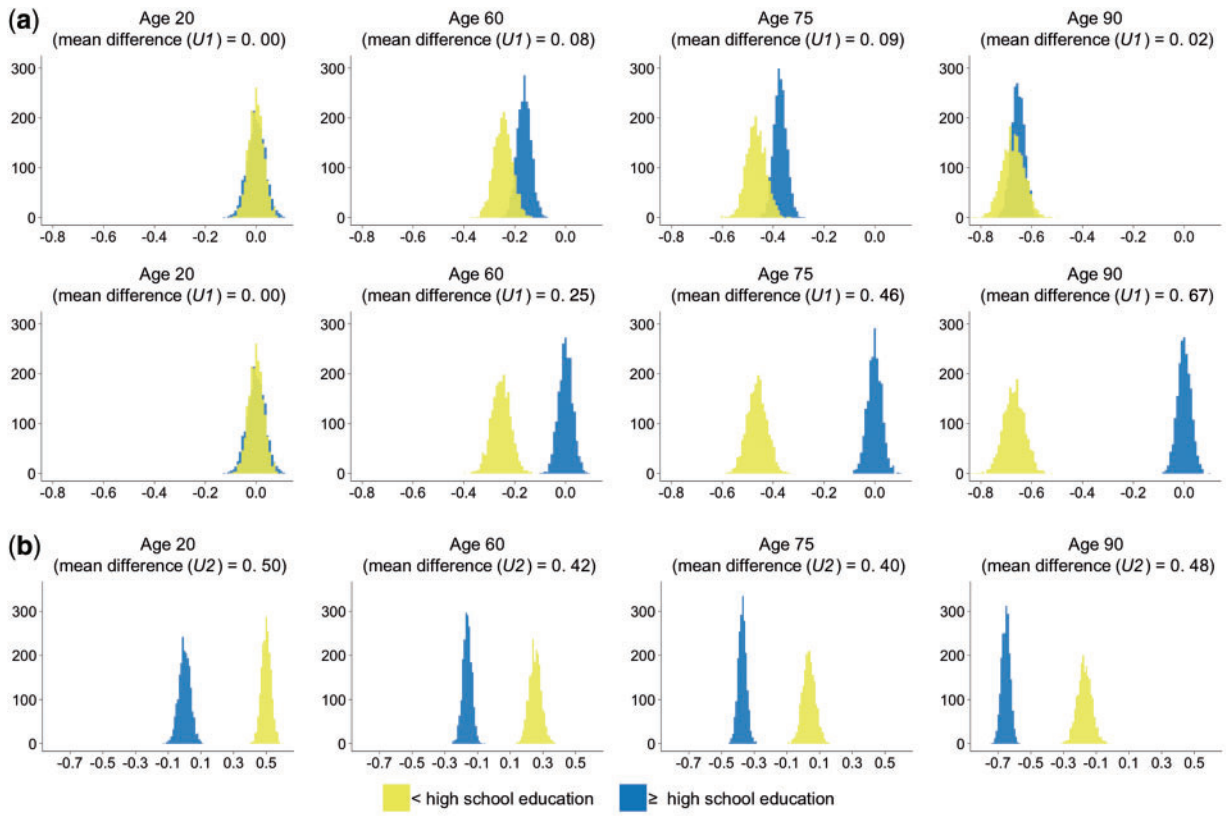

Figure 3. Histograms of *U1* (Alzheimer's disease polygenic risk score) and *U2* (occupational complexity) by education in a random sample of $n = 2000$ people at age 20 and among random samples of $n = 2000$ survivors to ages 60, 75 and 90. (a) Histograms of *U1* by education for Causal Structure 2 (collider stratification: education and *U1* influence mortality) (top row) and Causal Structure 3 (collider stratification with interaction: multiplicative interaction between education and *U1* on mortality) (bottom row). Distributions of *U1* for Causal Structure 1 (no bias anticipated scenario) and Causal Structure 4 (collider stratification with mediation: education and *U2* influence mortality) are not shown because the distributions of *U1* remained unchanged among survivors to late life in these causal structures. (b) Histograms of *U2* by education Causal Structure 4 (collider stratification with mediation: education and *U2* influence mortality). Distributions of *U2* for Causal Structures 1, 2 and 3 are not shown because the distributions of *U2* remained unchanged among survivors to late life in these causal structures.

Although the absolute magnitude of bias is not dependent on the causal effect of education on rate of cognitive decline, percentage bias is dependent on this effect. In our simulations, we assumed that low education increased rate of cognitive decline by 0.05 standardized units. This value is within the range reported by previous empirical studies of education and cognitive decline, although findings in previous studies are inconsistent, and in many studies are null.[5–10]

In simulation studies, the causal structures and parameterizations, and therefore validity of results, are based on investigator knowledge. A strength of our study is that the validity of our assumptions is empirically testable. That said, we did not explore all parameterizations of the causal structures. One potentially important causal structure that we did not assess is that cognitive decline preceding study entry directly influences mortality. Based on earlier simulation studies, it is likely that this scenario could induce sizeable bias.[28] Another strength is that our simulation code is available for researchers to estimate potential bias arising from selective survival in their own research. We previously used simulations for quantifying bias arising from selective survival after study enrolment.[28] Even when causal diagrams for selective survival before and after study enrolment are the same, the parametrization, and therefore potential magnitude of bias, differs. Both sources of attrition are common in life course epidemiology, and code can be adapted to quantify potential bias arising from selective survival from both sources. The magnitude of combined bias due to selective survival before and after study enrolment may induce greater bias than either source of attrition individually, and potentially large cumulative biases towards the null may contribute to research reporting null associations of education and cognitive decline in older adults.[5–10]

Potential bias arising from selective survival before study enrolment is widely relevant in life course epidemiology. It is especially relevant in dementia research because it is increasingly recognized that brain changes underlying cognitive decline develop over years—possibly decades—before symptom onset.[37,38] Education and other early

and midlife factors, such as childhood socioeconomic environment[7,39] and midlife cardiometabolic health,[40,41] are thought to potentially play an important role in cognitive ageing. Selective survival could lead to underestimation of possible causal effects of life course exposures on cognitive decline. This underestimation may limit identification of effective strategies to prevent or delay onset of dementia. Our simulations map survival bias in estimates of the effects of education on late-life cognitive decline to testable assumptions. The findings can inform the design and interpretation of future life course epidemiology studies.

Funding

This work was supported by National Institutes of Health [R13 AG030995, K99 AG053410, K01 AG050699, R01 AG051170; RF1 AG052132; U54 NS081760].

Acknowledgements

We gratefully acknowledge the Advanced Psychometrics Methods in Cognitive Aging Research Workshop, which facilitated the development and data analysis for this project.

Conflict of interest: None declared.

References

1. Kawachi I, Adler NE, Dow WH. Money, schooling, and health: mechanisms and causal evidence. *Ann N Y Acad Sci* 2010;**1186**: 56–68.
2. Cohen AK, Syme SL. Education: a missed opportunity for public health intervention. *Am J Public Health* 2013;**103**: 997–1001.
3. Albert MS, Jones K, Savage CR. Predictors of cognitive change in older persons: MacArthur studies of successful aging. *Psychol Aging* 1995;**10**:578–89.
4. Beydoun MA, Beydoun HA, Gamaldo AA, Teel A, Zonderman AB, Wang Y. Epidemiologic studies of modifiable factors associated with cognition and dementia: systematic review and meta-analysis. *BMC Public Health* 2014;**14**:1.
5. Glymour MM, Weuve J, Berkman LF, Kawachi I, Robins JM. When is baseline adjustment useful in analyses of change? An example with education and cognitive change. *Am J Epidemiol* 2005;**162**:267–78.
6. Zahodne LB, Glymour MM, Sparks C *et al*. Education does not slow cognitive decline with aging: 12-year evidence from the Victoria Longitudinal Study. *J Int Neuropsychol Soc* 2011;**17**: 1039–46.
7. Glymour MM, Tzourio C, Dufouil C. Is cognitive aging predicted by one's own or one's parents' educational level? Results from the Three-City Study. *Am J Epidemiol* 2012;**175**: 750–59.
8. Gottesman RF, Rawlings AM, Sharrett AR *et al*. Impact of differential attrition on the association of education with cognitive change over 20 years of follow-up: the ARIC neurocognitive study. *Am J Epidemiol* 2014;**179**:956–66.
9. Gross AL, Mungas DM, Crane PK *et al*. Effects of education and race on cognitive decline: an integrative study of generalizability versus study-specific results. *Psychol Aging* 2015;**30**: 863.
10. Schneider AL, Sharrett AR, Patel MD *et al*. Education and cognitive change over 15 years: the atherosclerosis risk in communities study. *J Am Geriatr Soc* 2012;**60**:1847–53.
11. Hummer RA, Hernandez EM. The effect of educational attainment on adult mortality in the United States. *Popul Bull* 2013; **68**:1–16.
12. Galea S, Tracy M, Hoggatt KJ, DiMaggio C, Karpati A. Estimated deaths attributable to social factors in the United States. *Am J Public Health* 2011;**101**:1456–65.
13. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004;**15**:615–25.
14. Hernán MA, Robins JM. *Selection Bias. Causal Inference*. Boca Raton, FL: Chapman & Hall/CRC, 2017.
15. Rast P, Hofer SM. Longitudinal design considerations to optimize power to detect variances and covariances among rates of change: simulation results based on actual longitudinal studies. *Psychol Methods* 2014;**19**:133–54.
16. Goldin C. America's graduation from high school: the evolution and spread of secondary schooling in the twentieth century. *J Econ Hist* 1998;**58**:345–74.
17. Arias E. United States Life Tables, 2011. *Natl Vital Stat Rep* 2015;**64**:11.
18. Tchetgen EJT, Glymour MM, Shpitser I, Weuve J. Rejoinder: to weight or not to weight? On the relation between inverse-probability weighting and principal stratification for truncation by death. *Epidemiology* 2012;**23**:132–37.
19. Tchetgen Tchetgen EJ. Identification and estimation of survivor average causal effects. *Stat Med* 2014;**33**:3601–28.
20. Tchetgen Tchetgen EJ, Phiri K, Shapiro R. A simple regression-based approach to account for survival bias in birth outcomes research. *Epidemiology* 2015;**26**:473–80.
21. Marden JR, Mayeda ER, Walter S *et al*. Using an alzheimer disease polygenic risk score to predict memory decline in black and white Americans over 14 years of follow-up. *Alzheimer Dis Assoc Disord* 2016;**30**:195–202.
22. Vivot A, Glymour MM, Tzourio C, Amouyel P, Chêne G, Dufouil C. Association of Alzheimer's related genotypes with cognitive decline in multiple domains: results from the Three-City Dijon study. *Mol Psychiatry* 2015;**20**:1173–78.
23. Kröger E, Andel R, Lindsay J, Benounissa Z, Verreault R, Laurin D. Is complexity of work associated with risk of dementia? The Canadian Study of Health and Aging. *Am J Epidemiol* 2008; **167**:820–30.
24. Then FS, Luck T, Luppa M, König H-H, Angermeyer MC, Riedel-Heller SG. Differential effects of enriched environment at work on cognitive decline in old age. *Neurology* 2015;**84**: 2169–76.
25. Greenland S. Response and follow-up bias in cohort studies. *Am J Epidemiol* 1977;**106**:184–87.
26. Hernán MA. Invited commentary: selection bias without colliders. *Am J Epidemiol* 2017;**185**:1048–50.
27. Viallon V, Dufournet M. Re: collider bias is only a partial explanation for the obesity paradox. *Epidemiology* 2017;**28**: e43–45.

28. Mayeda ER, Tchetgen Tchetgen EJ, Power MC *et al*. A simulation platform for quantifying survival bias: an application to research on determinants of cognitive decline. *Am J Epidemiol* 2016;**184**:378–87.
29. 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**:4–6.
30. Stensrud MJ, Valberg M, Røysland K, Aalen OO. Exploring selection bias by causal frailty models: the magnitude matters. *Epidemiology* 2017;**28**:379–86.
31. Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology* 2009;**20**:488–95.
32. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology* 2003;**14**: 300–06.
33. Corrada MM, Berlau DJ, Kawas CH. A population-based clinicopathological study in the oldest-old: the 90+ study. *Curr Alzheimer Res* 2012;**9**:709–17.
34. Vos SJ, Van Boxtel MP, Schiepers OJ *et al*. Modifiable risk factors for prevention of dementia in midlife, late life and the oldest-old: validation of the LIBRA Index. *J Alzheimers Dis* 2017;**58**:537–47.
35. Corrada MM, Paganini-Hill A, Berlau DJ, Kawas CH. Apolipoprotein E genotype, dementia, and mortality in the oldest old: the 90+ Study. *Alzheimers Dement* 2013;**9**:12–18.
36. Sabayan B, Gussekloo J, de Ruijter W, Westendorp RG, de Craen AJ. Framingham stroke risk score and cognitive impairment for predicting first-time stroke in the oldest old. *Stroke* 2013;**44**:1866–71.
37. Alzheimer's Association. 2016 Alzheimer's disease facts and figures. *Alzheimers Dement* 2016;**12**:459–509.
38. Villemagne VL, Burnham S, Bourgeat P *et al*. Amyloid beta deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. *Lancet Neurol* 2013;**12**:357–67.
39. Melrose RJ, Brewster P, Marquine MJ *et al*. Early life development in a multiethnic sample and the relation to late life cognition. *J Gerontol Ser B Psychol Sci Soc Sci* 2015;**70**: 519–31.
40. Whitmer R, Gustafson D, Barrett-Connor E, Haan M, Gunderson E, Yaffe K. Central obesity and increased risk of dementia more than three decades later. *Neurology* 2008;**71**:1057–64.
41. Whitmer RA, Sidney S, Selby J, Johnston SC, Yaffe K. Midlife cardiovascular risk factors and risk of dementia in late life. *Neurology* 2005;**64**:277–81.