How Selection Over Time Contributes to the Inconsistency of the Association between Sex/Gender and Cognitive Decline across Cognitive Aging Cohorts

Anaïs Rouanet, Justina Avila-Rieger, Aline Dugravot, Jérémie Lespinasse, Rachel Stuckwisch, Richard Merrick, Emma Anderson, Leann Long, Catherine Helmer, Hélène Jacqmin-Gadda, Carole Dufouil, Suzanne Judd, Jennifer Manly, Séverine Sabia, Alden Gross and Cécile Proust-Lima.

Correspondence to Dr. Anaïs Rouanet, Université de Bordeaux, Inserm U1219-BPH, Equipe de Biostatistique, 146 rue Léo-Saignat CS 61292, 33076 Bordeaux Cedex, France (e-mail: anais.rouanet@u-bordeaux.fr), + 33 (0)5 57 57 64 43.

Author affiliations: University of Bordeaux, INSERM, Bordeaux Population Health Research Center, U1219, F-33000 Bordeaux, France (Anaïs Rouanet, Jérémie Lespinasse, Catherine Helmer, Hélène Jacqmin-Gadda, Carole Dufouil, Cécile Proust-Lima); Taub Institute for Research on Alzheimer's Disease and the Aging Brain, College of Physicians and Surgeons, Columbia University, New York, NY 10032, USA (Justina Avila-Rieger, Jennifer Manly); Gertrude H. Sergievsky Center, College of Physicians and Surgeons, Columbia University, New York, NY 10032, USA (Justina Avila-Rieger, Jennifer Manly); Department of Neurology, College of

Physicians and Surgeons, Columbia University, New York, NY 10032, USA (Justina Avila-Rieger, Jennifer Manly); Université de Paris, Inserm U1153, Epidemiology of Ageing and Neurodegenerative diseases, Paris, France (Aline Dugravot, Séverine Sabia); CHU de Bordeaux, Pole Santé Publique, France (Jérémie Lespinasse, Carole Dufouil); Department of Biostatistics, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA (Rachel Stuckwisch, Leann Long, Suzanne Judd); Department of Public Health and Primary Care, University of Cambridge, UK (Richard Merrick); Medical Research Council Integrative Epidemiology Unit, Department of Population Health Sciences within the Bristol Medical School, University of Bristol, UK (Emma Anderson); Department of Epidemiology and Public Health, University College London, UK (Séverine Sabia); Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, and Johns Hopkins University Center on Aging and Health, Baltimore, MD, USA (Alden Gross).

This work was funded by the Agence Nationale de la Recherche (grant ANR-18-CE36-0004-01), the British Heart Foundation (grant RG/16/11/32334), the Medical Research Council (grants R024227, S011676, MR/P014437/1), the National Institute of Neurological Disorders and Stroke (grant U01 NS041588), the National Institute on Aging (grants K01AG050699, R01AG056477, R01AG056486, RF1AG062553), the National Institute of Health (grant NIH-NIA/R13AG064971).

Conflicts of interest: none declared.

Running head: [Sex/Gender & Cognitive Aging: impact of selection]

Abstract

The sex/gender and aging-related cognitive decline association remains poorly understood due to inconsistencies in findings. Such heterogeneity could be attributable to the cognitive functions studied and study population characteristics, but also to a differential selection by drop-out and death between men and women. This work aims to evaluate the impact of selection by drop-out and death on the association between sex/gender and cognitive decline. We first compared the most frequently used statistical methods for longitudinal data, targeting either population estimands (marginal models estimated by Generalized Estimating Equations) or subject-specific estimands (mixed/joint models estimated by likelihood maximization) on eight aging studies: six population-based (ACTIVE(1996-2009), Paquid(1988-2014), REGARDS(2003-2016), 3-City(1999-2016), WHICAP(1992-2017), Whitehall II(2007-2016)) and two clinic-based (ADNI(2004-2017), MEMENTO(2011-2016)) studies. We illustrated the differences in the estimands of the sex/gender association with cognitive decline in selected examples and highlighted the critical role of differential selection by drop-out and death. By using the same estimand, we then contrasted the sex/gender association across cohorts and cognitive measures suggesting residual differential sex/gender association depending on the targeted cognitive measure (memory or animal fluency) and the initial cohort selection. We recommend focusing on subject-specific estimands in the alive population for assessing sex/gender differences while handling differential selection over time.

Keywords (8): aging cohorts, cognition, death, drop-out, longitudinal models, selection, sex/gender.

Abbreviations: LMM: Linear mixed model; JM: Joint model; GEE: Generalized Estimating Equations; wGEE: weighted Generalized Estimating Equations; DAR death at random; DCAR death completely at random; DNAR death not at random; MAR missing at random; MCAR missing completely at random; MNAR missing not at random.

Examining sex/gender as a risk factor for Alzheimer's disease (AD), dementia, and aging-related cognitive decline is critical to better understand their underlying mechanisms and develop tailored prevention strategies. A sizable body of literature has investigated the association between sex/gender and either AD diagnosis [1][2][3] or cognitive decline [4][5][6][7], without reaching a consensus. This work aims to explore some directions that may explain such inconsistency in sex/gender associations by focusing specifically on aging-related cognitive decline.

There has been a growing public health recommendation to differentiate sex, biologically defined, and gender, the social construct [8]. Cognitive differences by sex may be biologically driven [1][9][10][11] while cognitive disparities between genders may result from differences in cognitive reserve [12], lifestyle or sociocultural factors [1][13]. Throughout this article, we refer to the "sex/gender" concept [14][15], combining the effect of sex and gender.

Heterogeneity in findings regarding sex/gender differences in aging-related cognitive decline is likely multifactorial. First, cohort characteristics and initial selection are an inherent source of discrepancy across cohorts. The lack of representativeness of the study populations may lead to spurious conclusions if sex/gender is associated with sample selection. Including men (or women) with different levels of education, ages, social backgrounds or ethnicities may widen the cognitive differences by sex/gender across studies at baseline and over follow-up.

Second, heterogeneity in findings may also result from the cognitive domains investigated [4][16] and the neuropsychological tests used to measure cognition [17][18][19][20]. As an example, episodic memory and verbal fluency, particularly involved in the pathological process toward dementia [21][22], seem to be oppositely associated with sex/gender [4][23][24] [25][26]. Neuropsychological tests also have inherent specific characteristics. Capitani et al. [17] argued

that gender was differentially associated with semantic fluency tasks, such as the fruit and tool categories. The varying testing conditions (length of the test, language, potential cues) may also make comparisons across studies less straightforward.

Finally, the selection over time induced by attrition by death and drop-out in aging cohorts is a major driver of heterogeneity since attrition commonly differs by sex/gender. Indeed, the survival bias observed in cognitive aging studies [27] can lead to men showing better preserved cognition compared to less selected women at the same age [14][28][29], as men and participants with low cognitive level tend to die earlier and drop out of studies at a higher rate. Handling this attrition properly is therefore critical when investigating the association between sex/gender and aging-related cognitive decline.

The present study, initiated during the 2018 annual meeting of MELODEM (MEthods in LOngitudinal DEMentia research) initiative [30], aimed to evaluate to what extent selection over time due to attrition by death and drop-out and its statistical handling could explain the heterogeneity in associations between sex/gender and aging-related cognitive decline. We leveraged data from eight large longitudinal cognitive aging cohorts to illustrate this impact on different attrition scenarios. We specifically compared two families of state-of-the-art longitudinal statistical methods, namely:

- population models, especially marginal models estimated by Generalized Estimating Equations (GEE), which target the change in the population mean [31];

- subject-specific models, especially mixed and joint models estimated by likelihood maximization, which target the individual change [32][33].

Although not handling them statistically, we also acknowledged the roles of initial selection and cognitive domains by illustrating the heterogeneity of associations across the eight cohorts and two important domains in cognitive aging, verbal fluency and episodic memory.

METHODS

Cohorts

This work relied on six population-based prospective cohorts: Advanced Cognitive Training for Independent and Vital Elderly (ACTIVE)[34], Personnes Âgées QUID (Paquid)[35], REasons for Geographic And Racial Differences in Stroke (REGARDS)[36], Bordeaux-Three City (3-City)[37], Washington Heights-Inwood Columbia Aging Project (WHICAP)[38], and Whitehall II[39]. We also considered two clinic-based prospective cohorts: Alzheimer's Disease Neuroimaging Initiative (ADNI)[40] and the Memento study (MEMENTO)[41]. The eight cohorts are described in Web Table 1 and Web Appendix 1.

To reduce heterogeneity across cohorts while limiting the selection criteria and ensuring a relevant target population for the study of aging-related cognitive decline, we included all the participants aged 65 years and older at baseline, with no prevalent dementia, with information at baseline on sex/gender, education level, age at entry, and with at least one measure of the cognitive test(s) under consideration at baseline. In addition, as the ACTIVE study was originally a randomized, controlled, single-masked trial, we only considered the control group as a representative sample of community residents. We considered the individuals from the Bordeaux area only in the 3-City study as the memory test under consideration was collected only in this

city. Finally, we included the four collection waves from ADNI study: ADNI1, ADNIGO, ADNI2 and ADNI3.

Cognitive domains

We focused primarily on episodic memory, central in AD research [21][24]. Memory was assessed using different tests across cohorts: a free recall test in Whitehall II [42], Rey's auditory verbal learning test [43] in ACTIVE and ADNI, Word List Learning test in REGARDS, Free and Cued Selective Reminding Test [44] in 3-City and MEMENTO, Selective Reminding Test [45] in WHICAP. For all these tests, participants were read a list of words and asked to recall as many as possible, at different occasions. However, the number of words and trials varied across tests: the respective lists included 20, 15, 10, 16, and 12 words and the total scores were the sum of the free recalls over 1, 5, 3, 3, and 6 trials, respectively. To further illustrate the importance of the potential differences according to cognitive domains, we considered verbal fluency in an additional analysis, as also involved in the pre-dementia process [21][25][26]. Participants had to name orally (or in writing in Whitehall II) as many animals as possible within 15 seconds (Paquid), 30 seconds (3-City), 60 seconds (ADNI, WHICAP, REGARDS, Whitehall II) or 120 seconds (MEMENTO).

Missingness and death processes

The missingness process in longitudinal studies has been categorized through three mechanisms [46]:

- Data are missing completely at random (MCAR) if the probability of drop-out at any time depends on the observed covariates only.

- Data are missing at random (MAR) if the probability of drop-out at any time *t* depends on the values of the marker observed before time *t*.

- Data are missing not at random (MNAR) if the probability of drop-out at any time *t* further depends on unobserved values such as the marker value at time *t* or current slope.

In aging studies, another source of attrition is death. Rouanet et al. [47] extended the missing data assumptions to the death mechanism by differentiating:

- Death completely at random (DCAR) if the probability to die at any time depends on the covariates only.

- Death at random (DAR) if the probability to die at any time *t* depends on the values of the marker observed before the current time *t*.

- Death not at random (DNAR) if the probability to die at any time further depends on unobserved values such as the current marker value or current slope.

Subject-specific and Population Estimands and Models

Two main types of models are used for analyzing longitudinal data in epidemiology[48]:

- subject-specific models such as linear mixed models (LMM) and joint models (JM);

- population-averaged models such as marginal models estimated by GEE.

Subject-specific methods ascertain individual mean trajectories and provide estimates of the individual change associated with covariates (adjusted for individual random effects that can be interpreted as the unobserved characteristics). In contrast, population-averaged methods estimate the population outcome mean and yield estimates of the population-level average change

associated with covariates (averaged over all unmeasured characteristics). Subject-specific and population-averaged estimands are identical only in the linear context with complete or MCAR/DCAR data.

When the follow-up can be terminated by drop-out or death at random (MAR/DAR) or not (MNAR/DNAR), which is likely in cohorts of elderly, some authors argued that subject-specific models impute data after death, making their estimates only interpretable for immortal subjects [49][50]. Rouanet et al. [47] showed that in a linear case where measurement errors are independent from the random effects and future time-to-death (which is a reasonable assumption), subject-specific estimates are also interpretable for subjects alive.

For estimating the association of sex/gender on cognitive decline with an etiologic purpose, we claim that subject-specific estimands for subjects alive are more relevant than population-averaged estimands in the population alive. Indeed, we are interested in the mean outcome difference between a man and a woman who are alive and share the same observed characteristics (covariates other than sex/gender) and unobserved ones (individual random effects). In contrast, the population-averaged estimand among the population alive represents the difference between the mean outcome value among men alive at some point, averaged over all unmeasured characteristics, and the mean outcome value among women alive at the same time, also averaged over all unmeasured characteristics. If death is more selective for men than women, the population-averaged estimand in the population alive would result from this survival bias. Hereafter, we consider that the estimand of interest in our context is the subject-specific effect of sex/gender for subjects alive. Under M(C)AR/D(C)AR assumptions, linear mixed models, usually estimated by Maximum Likelihood, are robust and provide unbiased estimates. When either drop-out or death is not at random (MNAR or DNAR), joint models for the outcome and the informative process

(drop-out or death, respectively) will give robust estimates of the subject-specific effect for subjects alive, provided that the association between the two variables is well-specified [47].

In the population-averaged approach, marginal models with an independence working correlation structure, estimated by GEE, estimate the covariate effect on the outcome mean among the population currently observed. This population being selected by both drop-out and death, the corresponding population-averaged estimate will be equal to the subject-specific target estimand only under the MCAR and DCAR assumptions. In order to make the marginal approach robust to MAR data, Dufouil et al. [51] proposed a weighting method where each observation is weighted by the inverse probability to be observed at the current time, given the subject is currently alive and given his/her covariates and past observed marker values. This method, noted wGEE, therefore corrects for the selection by drop-out by overweighting subjects alive likely to drop out. It estimates the population-averaged covariate effect among subjects currently alive, that is equal to the subject-specific effect for subjects alive only under M(C)AR/DCAR assumptions.

Table 2 summarizes the target estimands of each method and the assumptions under which each estimate corresponds to the unbiased subject-specific estimate for subjects alive.

Statistical analysis

All population-averaged and subject-specific models included the same linear regression at the population level (Web Appendix 2): the cognitive score was regressed on delay in the study (in decades) with a quadratic time trend, and was adjusted for age at baseline (centered in 75 years and in decades) and for a practice effect (time-dependent variable equal to 1 at the first visit and 0 afterwards) [52]. We also accounted for the effects of sex/gender (1 for male, 0 for female),

education level and their interactions with the functions of time. Education level was binarized, taking the value 1 for: obtaining primary school certificate in Paquid; >12 years of education (Baccalaureat or greater) in 3-City and MEMENTO; 'high school or greater' in WHICAP; 'A-level or more' in Whitehall II; 'some college or above' in REGARDS; > 14 years of education in ADNI and ACTIVE. Correlated individual random effects on each function of time were added for LMM and JM. Models were also adjusted for the entry wave in WHICAP and for ethnicity in REGARDS, WHICAP, and Whitehall II (see Web Appendix 1 for more details on these variables).

For the joint model, we defined a composite endpoint as the first event of either drop-out, dementia (when available) or death. To limit the potential bias due to interval censoring of time-to-dementia, we censored participants at their last visit if they died later than 3 years after their last visit. We specified a proportional hazards sub-model for the instantaneous hazard of the event adjusted for centered aged at baseline, sex/gender, education (and entry wave in WHICAP and ethnicity in REGARDS, WHICAP and Whitehall II), with a baseline hazard function modelled by cubic splines with 5 knots placed at the quantiles of the event times. The correlation between the cognitive test and the time-to-event was accounted for by adjusting the survival sub-model for the current true value of the test and current slope.

The wGEE weights were computed using a logistic model for the probability of drop-out among subjects alive, adjusted for sex/gender, education level, centered age at baseline and the previously observed outcome value (plus entry wave in WHICAP and ethnicity in REGARDS, WHICAP and Whitehall II).

The LMM, JM, GEE and wGEE models were run with the R packages nlme [53], JM [54], geeM [55] and the combination of weightQuant [56] and geeM, respectively. An example of the script is proposed for replication[57].

RESULTS

Description of the cohort samples

The study samples are described in Table 1 and Web Tables 2 and 3. Sample sizes varied from 606 in ACTIVE to 12,376 in REGARDS, however the age ranges were comparable across studies (Active: [65-93], ADNI: [65-99], MEMENTO: [65-93], Paquid: [65-105], REGARDS: [65-98], 3-City: [65-100], WHICAP: [65-103], Whitehall II: [65-79]). The heterogeneity in the follow-up length did not translate into a high variability in the number of measurements per subject, as studies with shorter duration of follow-up (Whitehall II, REGARDS, MEMENTO) presented more regular visits and/or less attrition. Finally, the cohorts exhibited large differences in terms of education levels and sex/gender proportions (male frequency ranging from 71% in Whitehall II to 27.4% in ACTIVE), demonstrating different initial selection across cohorts (Characteristics by sex/gender in Web table 3).

Contrasts across statistical methods and estimands

We illustrate the differences in the statistical methods on three cohorts that differ according to the drop-out and death mechanisms in Figures 1-3; the same figures for the other cohorts are available in the supplementary materials (Web Figures 1-4; Web Figures 5-11 for verbal fluency). For each

method, we report the trajectories estimated for men and women in a reference group (A) and the sex/gender estimates (with the 95% confidence intervals) as a function of the duration in the study (B). The non-parametric cumulative probabilities of drop-out (C) and death (D) over time in the study are also presented by sex/gender.

In the ACTIVE cohort (Figure 1), which has a substantial attrition by drop-out and death over time (probabilities above 25% at 10 years), estimated memory declines differ across methods (A). GEE estimates the slowest apparent decline as it targets a population-averaged estimand among individuals still in the cohort, and those who remain longer are likely to be healthier on average than those who drop out or die (not MCAR/DCAR mechanism). The wGEE, targeting the population-averaged evolutions among subjects alive, corrects for drop-out selection. It shows steeper declines than GEE, because participants who drop out are cognitively more impaired and in the wGEE method, participants alive with a higher probability of drop-out (possibly explained by lower cognitive scores) are upweighted. Finally, the LMM/JM estimated trajectories can either be interpreted as population-averaged trajectories in the immortal population or as subject-specific trajectories for a subject alive with random effects equal to 0 [47]. They show even steeper declines than wGEE trajectories as they also account for selection by death (DAR in LMM; DNAR in JM) and participants who die are likely to be more cognitively impaired than those who survive.

Sex/gender differences in ACTIVE (B) are comparable across the four methods at baseline (all in favor of women performing better) as every participant is alive and observed, as required in the inclusion criteria. We observe no difference between GEE and wGEE estimates over time, the drop-out probability being the same among men and women (C). However, differences between the population-averaged estimates (GEE and wGEE) and the subject-specific estimates for any

subject alive (LMM and JM) arise over time: at 5 and 9 years, the population-averaged estimates are more in favor of men. Indeed, men who stay alive are likely to be healthier since the risk of death is higher among both men (D) and subjects with low cognitive levels.

In contrast, the Whitehall II study (Figure 2) presents a very low probability of death among both men and women (D), but a selective drop-out in favor of men mainly at three times (right after baseline, around 4 years and after 8 years, C). All the predicted memory trajectories are more similar as they all lay within the 95% confidence band of the LMM estimated evolution, represented in shaded areas (A). In particular, due to the very low mortality, the population-averaged trajectory among subjects alive (wGEE) and the subject-specific mean trajectory (LMM) are very close. All the sex/gender estimates evolve similarly, showing a significant cognitive difference in favor of women at baseline which gets closer to the null value over time (B).

Finally, the WHICAP study (Figure 3), which is open and has a longer follow-up, exhibited larger probabilities of drop-out and death over time. This translates into wider differences between the estimated trajectories of GEE and wGEE on the one hand, and between wGEE and LMM on the other hand (A). The difference between LMM and JM models suggests that missing data might not be at random. Despite large rates of drop-out and death, the sex/gender estimates in favor of women (B) remain relatively close and stable across time and methods. This might be due to opposite sex/gender attrition patterns: higher drop-out among women and higher death rate among men.

Contrasts across cognitive functions and cohorts

Due to the small differences between the LMM and JM estimates observed in the illustrations above, we focused on the subject-specific LMM estimates to assess the cognitive differences according to sex/gender across cohorts (Table 3). Here, cognitive tests were standardized according to the mean and standard deviation at baseline to make effect sizes comparable across cohorts. Overall, women had higher memory scores regardless of the cohorts and sex/gender differences seemed to attenuate with time (but were still significant at 9 years), although some differences in effect sizes remained across cohorts. When exploring verbal fluency, sex/gender differences were smaller and in favor of men at baseline, except in clinic-based cohorts (ADNI, MEMENTO). This might be due to the stricter initial selection in the latter (see Web Table 2).

DISCUSSION

This longitudinal study comparing eight diverse cohorts presents two key findings. First, there were substantial differences across studies in the probabilities of drop-out and death, which affected predicted cognitive trajectories and estimates of sex/gender differences during follow-up. Secondly, even after controlling for attrition, sex/gender associations still differed according to the cognitive function of interest and the initial cohort selection.

The choice of the appropriate statistical method is crucial when assessing the sex/gender association with cognition. We argued that subject-specific models, which target the change in the individual mean for any subject alive, are more suitable for studying this association. Indeed, from an etiological perspective, the estimand needs to contrast the effect for individuals that are similar

except for sex/gender and subject-specific models target the cognitive difference between a man and a woman alive sharing the same other observed and unobserved (i.e. random effects) characteristics. When using population-averaged methods with GEE and wGEE models, the drop-out and/or death mechanisms over time may bias this estimand estimation, as illustrated here on chosen examples. Indeed, in elderly populations, the missingness mechanism likely depends on cognition (either observed cognition in MAR/DAR or unobserved cognitive information in MNAR/DNAR) but is also likely differential according to sex/gender with men tending to have a higher risk of death while women tend to have a higher risk of dementia [28][29] [58]. It should be noted that, although the LMM/JM covariate effects are interpreted for any subject alive, the predicted mean trajectories can only be displayed for a given subject (i.e. with a given value of the random effects). Finally, by estimating the covariate effect on the outcome mean among the dynamic population of subjects alive at each time, population-averaged models (or marginal models) are more suitable to define norms of cognitive scores in the population alive [56].

Once the heterogeneity in sex/gender and cognition association due to attrition by death and drop-out was handled, we still observed residual differences in the sex/gender association across cohorts. For instance, in the two clinic-based studies, women performed better at verbal fluency at baseline. One potential reason for differences between observational and clinic-based cohorts relies on differences between men and women in their tendency to consult healthcare services as suggested by later diagnosis of Alzheimer's Disease observed among men [59]. As a consequence, men who take part in clinical cognitive studies may be at a more advanced stage of cognitive impairment than women at inclusion. Thus, heterogeneity of findings across cohorts

highlights the potentially large impact of initial sample selection on the result estimates, which must be interpreted carefully.

As expected, we also found very different sex/gender associations according to the cognitive domains investigated: women generally performed better at memory while men generally performed better at animal naming. In a secondary analysis, we even confirmed the variability among semantic categories: in the Paquid cohort, women performed better than men in the fruit category (Web Figure 12). These differences illustrate the complexity and specificity of sex/gender association with cognitive domains/functions, even when the other sources of heterogeneity have been handled.

This work leveraged large high-quality cohort data and used state-of-the-art longitudinal methods to demonstrate how attrition over follow-up contributes in making the associations of sex/gender with cognitive decline so inconsistent in the literature. However, it also has limitations. First, we focused on only two cognitive domains particularly relevant to AD research and available in most cohorts, as our aim was to illustrate the differences, not to give an exhaustive view of the heterogeneity in association across all cognitive domains. Second, although we focused on similar tests across cohorts (same memory test, same semantic category), differences remained in test stimuli and administration procedures (e.g., number of words for memory tests, duration and scoring for animal naming) across cohorts. We reported standardized estimates to account for effect sizes (Table 3), however it is unlikely to completely eliminate variability due to cohort-specific characteristics of the tests. Third, although the samples shared the same inclusion criteria, cohorts were very different in design and target population (see Table 1 and Web Table 1), which is inherent to aging research. Fourth, we considered a relatively simple parametric regression.

However additional analyses using more flexible functions of time (B-splines) led to very similar results (not shown). We adjusted for only a few major confounders. However, a more comprehensive adjustment may refine comparisons across cohorts but is unlikely to change conclusions regarding the comparison across statistical methods in a given cohort. Finally, we considered a composite event in the joint model, combining drop-out, dementia (when available) and death, described by an attrition mechanism potentially not at random. The three causes of attrition being unlikely defined by the same mechanism, a joint model with competing risks might be more appropriate but is unlikely to change our main conclusions regarding subject-specific and population-averaged differences.

Our work highlighted three sources of heterogeneity in sex/gender association with cognition: (i) the comparability of cohorts at baseline; (i) the cognitive dimension under study; (iii) selection over time and its statistical handling. We focused here on the appropriate estimand to handle selection by death and drop-out, as a prerequisite for any analysis on sex/gender association with cognitive decline. However, in future research, the methods employed here could also be combined with (i) inverse probability weighting methods [60] to further handle initial selection and make populations comparable across cohorts; (ii) latent processes approaches [4] to target the effect of sex/gender on the cognitive process underlying different cognitive tests, while accounting for differential test-specific associations.

Beyond the sex/gender association with cognitive decline studied in this work, our take-home message is that subject-specific models (LMM/JM) should be favored in etiologic research for correctly handling selection by drop-out and death when assessing a covariate effect on a longitudinal outcome.

Acknowledgments:

Author list: Anaïs Rouanet, Justina Avila-Rieger, Aline Dugravot, Jérémie Lespinasse, Rachel Stuckwisch, Richard Merrick, Emma Anderson, Leann Long, Catherine Helmer, Hélène Jacqmin-Gadda, Carole Dufouil, Suzanne Judd, Jennifer Manly, Séverine Sabia, Alden Gross and Cécile Proust-Lima.

Author affiliations: University of Bordeaux, INSERM, Bordeaux Population Health Research Center, U1219, F-33000 Bordeaux, France (Anaïs Rouanet, Jérémie Lespinasse, Catherine Helmer, Hélène Jacqmin-Gadda, Carole Dufouil, Cécile Proust-Lima); Taub Institute for Research on Alzheimer's Disease and the Aging Brain, College of Physicians and Surgeons, Columbia University, New York, NY 10032, USA (Justina Avila-Rieger, Jennifer Manly); Gertrude H. Sergievsky Center, College of Physicians and Surgeons, Columbia University, New York, NY 10032, USA (Justina Avila-Rieger, Jennifer Manly); Department of Neurology, College of Physicians and Surgeons, Columbia University, New York, NY 10032, USA (Justina Avila-Rieger, Jennifer Manly); Université de Paris, Inserm U1153, Epidemiology of Ageing and Neurodegenerative diseases, Paris, France (Aline Dugravot, Séverine Sabia); CHU de Bordeaux, Pole Santé Publique, France (Jérémie Lespinasse, Carole Dufouil); Department of Biostatistics, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA (Rachel Stuckwisch, Leann Long, Suzanne Judd); Department of Public Health and Primary Care, University of Cambridge, UK (Richard Merrick); Medical Research Council Integrative Epidemiology Unit, Department of Population Health Sciences within the Bristol Medical School, University of Bristol, UK (Emma Anderson); Department of Epidemiology and Public Health, University College London, UK (Séverine Sabia); Department of Epidemiology, Johns Hopkins

Bloomberg School of Public Health, and Johns Hopkins University Center on Aging and Health, Baltimore, MD, USA (Alden Gross).

This work was funded by the Agence Nationale de la Recherche (grant ANR-18-CE36-0004-01), the British Heart Foundation (grant RG/16/11/32334), the Medical Research Council (grants R024227, S011676, MR/P014437/1), the National Institute of Neurological Disorders and Stroke (grant U01 NS041588), the National Institute on Aging (grants K01AG050699, R01AG056477, R01AG056486, RF1AG062553), the National Institute of Health (grant NIH-NIA/R13AG064971).

Data availability: The access to the 8 cohorts was granted by the respective Principal Investigators: Alden Gross for ACTIVE, Catherine Helmer for Paquid, Suzanne Judds for REGARDS, Carole Dufouil for MEMENTO and Three-City, Jennifer Manly for WHICAP, Séverine Sabia for Whitehall II, and ADNI is publicly available.

Conflicts of interest: none declared.

[1] M. Mielke, P. Vemuri, and W. Rocca, "Clinical epidemiology of Alzheimer's disease: assessing sex and gender differences," *Clin. Epidemiol.*, p. 37, Jan. 2014.
[2] A. Altmann, L. Tian, V. W. Henderson, and M. D. Greicius, "Sex modifies the APOE-related risk of developing Alzheimer disease," *Ann. Neurol.*, vol. 75, no. 4, pp. 563–573, Apr. 2014.
[3] S. Kim *et al.*, "Gender differences in risk factors for transition from mild cognitive impairment to Alzheimer's disease: A CREDOS study," *Compr. Psychiatry*, vol. 62, pp. 114–122, Oct. 2015.
[4] C. Proust-Lima, H. Amieva, L. Letenneur, J.-M. Orgogozo, H. Jacqmin-Gadda, and J.-F. Dartigues, "Gender and Education Impact on Brain Aging : A General Cognitive Factor Approach," *Psychol. Aging*, vol. 23, no. 3, pp. 608–620, 2008.
[5] K. R. Laws, K. Irvine, and T. M. Gale, "Sex differences in cognitive impairment in Alzheimer's disease," *World J. Psychiatry*, vol. 6, no. 1, p. 54, 2016.
[6] K. Irvine, K. R. Laws, T. M. Gale, and T. K. Kondel, "Greater cognitive deterioration in women than men with Alzheimer's disease: A meta analysis," *J. Clin. Exp. Neuropsychol.*, vol. 34, no. 9, pp. 989–998, Nov. 2012.
[7] S. B. Maitland, R. C. Intrieri, W. K. Schaie, and S. L. Willis, "Gender Differences and Changes in Cognitive Abilities Across the Adult Life Span," *Aging, Neuropsychol. Cogn.*, vol. 7, no. 1, pp. 32–53, Mar. 2000.
[8] C. Tannenbaum, L. Greaves, and I. D. Graham, "Why sex and gender matter in implementation research," *BMC Med. Res. Methodol.*, vol. 16, no. 1, p. 145, Dec. 2016.
[9] M. Ancelin and K. Ritchie, "Lifelong Endocrine Fluctuations and Related Cognitive Disorders," *Curr. Pharm. Des.*, vol. 11, no. 32, pp. 4229–4252, Dec. 2005.
[10] D. W. Fisher, D. A. Bennett, and H. Dong, "Sexual dimorphism in predisposition to Alzheimer's disease," *Neurobiol. Aging*, vol. 70, pp. 308–324, Oct. 2018.
[11] R. S. Vest and C. J. Pike, "Gender, sex steroid hormones, and Alzheimer's disease," *Horm. Behav.*, vol. 63, no. 2, pp. 301–307, Feb. 2013.
[12] E. E. Sundermann, P. M. Maki, L. H. Rubin, R. B. Lipton, S. Landau, and A. Biegon, "Female advantage in verbal memory," *Neurology*, vol. 87, no. 18, pp. 1916–1924, Nov. 2016.
[13] A. Herlitz and J. Rehnman, "Sex Differences in Episodic Memory," *Curr. Dir. Psychol. Sci.*, vol. 17, no. 1, pp. 52–56, Feb. 2008.
[14] E. R. Mayeda, "Examining sex/gender differences in risk of Alzheimer's disease and related dementias: Challenges and future directions," *Am. J. Epidemiol.*, 2019.
[15] J. F. Avila, J. M. J. Vonk, S. P. Verney, and K. Witkiewitz, "Sex / gender differences in cognitive trajectories vary as a function of race / ethnicity," pp. 1–8, 2019.
[16] E. M. Weiss, G. Kemmler, E. A. Deisenhammer, W. W. Fleischhacker, and M. Delazer, "Sex differences in cognitive functions," *Pers. Individ. Dif.*, vol. 35, no. 4, pp. 863–875, Sep. 2003.
[17] E. Capitani, M. Laiacona, and R. Barbarotto, "Gender Affects Word Retrieval of Certain Categories in Semantic Fluency Tasks," *Cortex*, vol. 35, no. 2, pp. 273–278, 1999.
[18] N. Raoux, M. Le Goff, S. Auriacombe, J.-F. Dartigues, and H. Amieva, "Fluences verbales sémantiques et littérales : normes en population générale chez des sujets âgés de 70 ans et plus issus de la cohorte PAQUID," *Rev. Neurol. (Paris).*, vol. 166, no. 6–7, pp. 594–605, Jun. 2010.
[19] C. Thomas-Antérion, S. Honoré, H. Cougny, C. Grosmaitre, and B. Laurent, "Apport de

l'épreuve d'évocation lexicale du Set test dans le dépistage de la maladie d'Alzheimer," *Rev. Neurol. (Paris).*, vol. 160, pp. 1377–1382, 2001.
[20] D. Cardebat, B. Doyon, M. Puel, P. Goulet, and Y. Joanette, "[Formal and semantic lexical evocation in normal subjects. Performance and dynamics of production as a function of sex, age and educational level].," *Acta Neurol. Belg.*, vol. 90, no. 4, pp. 207–17, 1990.
[21] H. Amieva *et al.*, "The 9 year cognitive decline before dementia of the Alzheimer type: a prospective population-based study," *Brain*, vol. 128, no. 5, pp. 1093–1101, May 2005.
[22] F. Di Stefano *et al.*, "Prediction of Alzheimer's Disease Dementia: Data from the GuidAge Prevention Trial," *J. Alzheimer's Dis.*, vol. 48, no. 3, pp. 793–804, Oct. 2015.
[23] N. M. Payton *et al.*, "Rates of cognitive decline during the preclinical phase and their use in predicting future dementia," *Alzheimer's Dement.*, vol. 16, no. S5, Dec. 2020.
[24] M. Albert, "The neuropsychology of the development of Alzheimer's disease," in *The handbook of aging and cognition*, F. I. . Craik and T. A. Salthouse, Eds. Psychology Press., 2008, pp. 97–132.
[25] J. L. Cummings, "The one-minute mental status examination," *Neurology*, vol. 62, no. 4, pp. 534–535, Feb. 2004.
[26] J. D. Henry, J. R. Crawford, and L. H. Phillips, "Verbal fluency performance in dementia of the Alzheimer's type: a meta-analysis," *Neuropsychologia*, vol. 42, no. 9, pp. 1212–1222, Jan. 2004.
[27] M. A. Hernán, S. Hernández-Díaz, and J. M. Robins, "A Structural Approach to Selection Bias," *Epidemiology*, vol. 15, no. 5, pp. 615–625, Sep. 2004.
[28] R. F. Buckley, M. Waller, C. L. Masters, and A. Dobson, "To What Extent Does Age at Death Account for Sex Differences in Rates of Mortality From Alzheimer Disease?," *Am. J. Epidemiol.*, vol. 188, no. 7, pp. 1213–1223, Jul. 2019.
[29] G. Chêne *et al.*, "Gender and incidence of dementia in the Framingham Heart Study from mid-adult life," *Alzheimer's Dement.*, vol. 11, no. 3, pp. 310–320, Mar. 2015.
[30] J. Weuve *et al.*, "Guidelines for reporting methodological challenges and evaluating potential bias in dementia research," *Alzheimer's Dement.*, vol. 11, no. 9, pp. 1098–1109, Sep. 2015.
[31] N. M. Laird and J. H. Ware, "Random-Effects Models for Longitudinal Data," *Biometrics*, vol. 38, no. 4, pp. 963–974, 1982.
[32] K.-Y. Liang and S. L. Zeger, "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, vol. 73, no. 1, pp. 13–22, 1986.
[33] D. Rizopoulos, *Joint models for longitudinal and time-to-event data: With applications in R*. Boca Raton, Florida: Chapman & Hall/CRC., 2012.
[34] J. B. Jobe *et al.*, "ACTIVE: A cognitive intervention trial to promote independence in older adults," *Control. Clin. Trials*, vol. 22, no. 4, pp. 453–479, Aug. 2001.
[35] L. Letenneur, D. Commenges, J.-F. Dartigues, and P. Barberger-Gateau, "Incidence of dementia and Alzheimer's disease in elderly community residents of south-western France.," *Int. J. Epidemiol.*, vol. 23, no. 6, pp. 1256–1261, 1994.
[36] V. J. Howard *et al.*, "The Reasons for Geographic and Racial Differences in Stroke Study: Objectives and Design," *Neuroepidemiology*, vol. 25, no. 3, pp. 135–143, 2005.
[37] 3C Study Group, "Vascular factors and risk of dementia: design of the Three-City Study and baseline characteristics of the study population.," *Neuroepidemiology*, vol. 22, no. 6, pp. 316–325, 2003.

[38] M.-X. Tang *et al.*, "Incidence of AD in African-Americans, Caribbean Hispanics, and Caucasians in northern Manhattan," *Neurology*, vol. 56, no. 1, pp. 49–56, Jan. 2001.
[39] M. G. Marmot *et al.*, "Health inequalities among British civil servants: the Whitehall II study," *Lancet*, vol. 337, no. 8754, pp. 1387–1393, Jun. 1991.
[40] S. G. Mueller *et al.*, "Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI).," *Alzheimers. Dement.*, vol. 1, no. 1, pp. 55–66, 2005.
[41] C. Dufouil, B. Dubois, B. Vellas, F. Pasquier, F. Blanc, and J. Hugon, "Cognitive and imaging markers in non- demented subjects attending a memory clinic : study design and baseline findings of the MEMENTO cohort," *Alzheimer's Res. Ther.*, vol. 9, no. 1, p. 67, 2017.
[42] A. Singh-Manoux *et al.*, "Timing of onset of cognitive decline: results from Whitehall II prospective cohort study," *BMJ*, vol. 344, no. jan04 4, pp. d7622–d7622, Jan. 2012.
[43] A. Rey, *The clinical psychological examination*, Paris: Pre. 1964.
[44] E. Grober, H. Buschke, H. Crystal, S. Bang, and R. Dresner, "Screening for dementia by memory testing," *Neurology*, vol. 38, no. 6, pp. 900–900, Jun. 1988.
[45] H. Buschke and P. A. Fuld, "Evaluating storage, retention, and retrieval in disordered memory and learning," *Neurology*, vol. 24, no. 11, pp. 1019–1019, Nov. 1974.
[46] R. J. A. Little and D. Rubin, *Statistical analysis with missing data*. Wiley, New York, 1987.
[47] A. Rouanet, C. Helmer, J.-F. Dartigues, and H. Jacqmin-Gadda, "Interpretation of mixed models and marginal models with cohort attrition due to death and drop-out," *Stat. Methods Med. Res.*, vol. 28, no. 2, pp. 343--356, 2017.
[48] D. Commenges and H. Jacqmin-Gadda, *Dynamical Biostatistical Models*, 1st ed. Boca Raton, Florida: Chapman and Hall/CRC, 2015.
[49] B. F. Kurland and P. J. Heagerty, "Directly parameterized regression conditioning on being alive: analysis of longitudinal data truncated by deaths," *Biostatistics*, vol. 6, no. 2, pp. 241–258, Apr. 2005.
[50] B. F. Kurland, L. L. Johnson, B. L. Egleston, and P. H. Diehr, "Longitudinal Data with Follow-up Truncated by Death: Match the Analysis Method to Research Aims," *Stat. Sci.*, vol. 24, no. 2, pp. 211–222, May 2009.
[51] C. Dufouil, C. Brayne, and D. Clayton, "Analysis of longitudinal studies with death and drop-out: a case study," *Stat. Med.*, vol. 23, no. 14, pp. 2215–2226, Jul. 2004.
[52] A. Vivot *et al.*, "Jump, Hop, or Skip: Modeling Practice Effects in Studies of Determinants of Cognitive Change in Older Adults," *Am. J. Epidemiol.*, vol. 183, no. 4, pp. 302–314, Feb. 2016.
[53] J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, and R. C. Team, "nlme: Linear and Nonlinear Mixed Effects Model," *R Packag. version*, vol. 3, no. 4, p. 109, 2006.
[54] D. Rizopoulos, "JM : An R Package for the Joint Modelling of Longitudinal and Time-to-event data," vol. 35, no. 9, 2010.
[55] L. S. McDaniel, N. C. Henderson, and P. J. Rathouz, "Fast pure R implementation of GEE: application of the Matrix package," *R J.*, vol. 5, no. 1, pp. 181–187, 2013.
[56] H. Jacqmin-gadda, A. Rouanet, and V. Philipps, "Quantile regression for incomplete longitudinal data with selection by death Statistics in Medicine," *Stat. Med.*, pp. 1–14, 2020.
[57] "Script for Sex/Gender analysis." [Online]. Available: https://github.com/anarouanet/Sex-

gender-analysis.
[58] C. R. Beam, C. Kaneshiro, J. Y. Jang, C. A. Reynolds, N. L. Pedersen, and M. Gatz, "Differences Between Women and Men in Incidence Rates of Dementia and Alzheimer's Disease," *J. Alzheimer's Dis.*, vol. 64, no. 4, pp. 1077–1083, Jul. 2018.
[59] T. Rapp, "Patients' diagnosis decisions in Alzheimer's disease: The influence of family factors," *Soc. Sci. Med.*, vol. 118, pp. 9–16, Oct. 2014.
[60] E. A. Nohr and Z. Liew, "How to investigate and adjust for selection bias in cohort studies," *Acta Obstet. Gynecol. Scand.*, vol. 97, no. 4, pp. 407–416, Apr. 2018.

Table 1. Description of the 8 study samples (ACTIVE (1996-2009), ADNI (2004-2017), MEMENTO (2011-2016), REGARDS (2003-2016), Paquid (1988-2014), 3-City (1999-2016), WHICAP (1992-2017), Whitehall II (2007-2016)).

Characteristics	Cohorts															
	ACTIVE (*n* = 606)		ADNI (*n* = 1,483)		MEMENTO (*n* = 1,752)		Paquid (*n* = 3,386)		REGARDS (*n* = 12,376)		3-City (*n* = 1,487)		WHICAP (*n* = 5,398)		Whitehall II (*n* = 2,897)	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Drop-out	140	23.1	1,424	96	476	27.2	1,769	52.2	3,137	25.3	958	64.4	3,575	66.2	768	26.5
Dementia	NA	NA	394	26.6	248	14.2	828	24.5	NA	NA	273	18.4	790	14.6	33	1.1
Incident death before drop-out	241	39.8	46	3.1	63	3.6	1,465	43.3	3,279	26.5	595	40	1,222	22.6	168	5.8
Male	166	27.4	813	54.8	683	39.0	1,439	42.5	5,792	46.8	563	37.9	1,743	32.3	2,056	71
High Education[a]	156	25.7	1,048	70.7	688	39.3	2,251	66.5	7,759	62.7	288	19.4	1,570	29.1	1,465	50.6
MMSE at baseline:																
<24	36	5.9	2	0.1	73	4.2	680	20.1	NA	NA	46	3.1	NA	NA	20	0.7
24-27	252	41.5	565	38.1	1,168	66.7	1,296	38.3	NA	NA	483	32.5	NA	NA	608	21
28-30	319	52.6	916	61.8	509	29.1	1,408	41.6	NA	NA	957	64.4	NA	NA	2,317	80

ACTIVE: Advanced Cognitive Training for Independent and Vital Elderly; ADNI: Alzheimers Disease Neuroimaging; Paquid: Personnes Âgées QUID; REGARDS: REasons for Geographic And Racial Differences in Stroke; 3-City: Three-city study; WHICAP: Washington Heights-Inwood Columbia Aging Project

[a] High education is defined as follows: obtaining primary school certificate in Paquid, >12 years of education (Baccalaureate or greater) in 3City and MEMENTO, (high school of greater) in WHICAP, 'A-level or more' in Whitehall II, 'college graduate or above' in REGARDS, > 14 years of education in ADNI and ACTIVE.

Table 2. Comparison of population models[a] and subject-specific models[b] in terms of target estimands[c] and drop-out and death mechanisms under which the method estimate corresponds to the unbiased subject-specific estimate for subjects alive.

Model	Estimand		Level		Drop-out mechanism			Death mechanism		
	Observed	Alive	PA	SS	MCAR	MAR	MNAR	DCAR	DAR	DNAR
GEE	X		X		X			X		
wGEE		X	X		X	X		X		
LMM		X		X	X	X		X	X	
JM		X		X	X	X	X	X	X	X

Abbreviations: wGEE: weighted Generalized Estimating Equations; LMM: linear mixed model; JM: joint model; PA: population-averaged; SS: subject-specific; DAR: death at random; DCAR: death completely at random; DNAR: death not at random; MAR: missing at random; MCAR: missing completely at random; MNAR: missing not at random.

[a]Generalized Estimating Equations and weighted Generalized Estimating Equations;
[b] Linear mixed model and joint model;
[c] population-averaged or subject-specific population observed or alive;

Table 3. Mean difference in Animal naming task and memory test according to sex/gender (for male compared to female) predicted by a linear mixed model at baseline, 5 and 9 years, with 95% confidence intervals in brackets, in ACTIVE (1996-2009), ADNI (2004-2017), MEMENTO (2011-2016), Paquid (1988-2014), REGARDS (2003-2016), 3-City (1999-2016), WHICAP (1992-2017), Whitehall II (2007-2016).

Cohort	Duration of Study Period, years	Baseline		5 years		9 years	
		Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
Memory							
ACTIVE[a]	13 years	-0.52	-0.66, -0.38	-0.48	-0.68, -0.29	-0.42	-0.69, -0.14
ADNI[a]	13 years	-0.55	-0.64, -0.46	-0.53	-0.65, -0.41	-0.49	-0.75, -0.24
MEMENTO[a]	5 years	-0.34	-0.43, -0.24	-0.34	-0.47, -0.20		
REGARDS[b]	13 years	-0.44	-0.46, -0.42	-0.40	-0.42, -0.38	-0.41	-0.44, -0.38
3-City[a]	17 years	-0.42	-0.52, -0.33	-0.28	-0.40, -0.17	-0.26	-0.40, -0.12
WHICAP[c]	25 years	-0.26	-0.30, -0.21	-0.25	-0.32, -0.18	-0.25	-0.35, -0.14
Whitehall II[b]	9 years	-0.14	-0.22, -0.06	-0.08	-0.17, 0.01	0.02	-0.14, 0.19
Verbal fluency							
ADNI[a]	13 years	-0.10	-0.19, -0.01	-0.18	-0.30, -0.06	-0.29	-0.48, -0.11
MEMENTO[a]	5 years	-0.07	-0.17, 0.02	-0.17	-0.29, -0.06		
Paquid[a]	26 years	0.09	0.03, 0.15	0.04	-0.02, 0.10	0.02	-0.05, 0.10
REGARDS[b]	13 years	0.12	0.10, 0.14	0.06	0.04, 0.08	0.06	0.03, 0.08
3-City[a]	17 years	0.14	0.04, 0.24	0.11	0.02, 0.21	0.06	-0.04, 0.16
WHICAP[c]	25 years	0.06	0.01, 0.12	-0.02	-0.09, 0.06	-0.07	-0.17, 0.02
Whitehall II[b]	9 years	0.06	-0.02, 0.13	0.07	-0.01, 0.14	0.04	-0.10, 0.19

CI: Confidence Interval; ACTIVE: Advanced Cognitive Training for Independent and Vital Elderly; ADNI: Alzheimers Disease Neuroimaging; Paquid: Personnes Âgées QUID; REGARDS: REasons for Geographic And Racial Differences in Stroke; 3-City: Three-city study; WHICAP: Washington Heights-Inwood Columbia Aging Project

[a] Model adjusted for education level, practice effect and age at entry in the cohort.

[b] Model adjusted for education level, practice effect, age at entry in the cohort and ethnicity.

[c] Model adjusted for education level, practice effect, age at entry in the cohort, ethnicity and entry wave.

For this table, cognitive tests were standardized according to mean and standard deviation at baseline to make effect sizes comparable.

Figure 1. Comparison of the statistical approaches on memory in the ACTIVE study (1996-2009): A) Mean trajectories estimated by GEE, weighted GEE (wGEE), linear mixed model (LMM) and joint model (JM) for males (light gray) and females (dark gray) entered at 75 years old, with low level of education; B) Male vs female difference in memory estimated by the four methods, at baseline, 5 and 9 years; C) non-parametric Aalen-Johansen estimates of the probability to drop out for males (light gray) and females (dark gray); D) non-parametric Aalen-Johansen estimates of the probability to die for males (light gray) and females (dark gray). Shaded areas represent 95% confidence intervals of LMM in A) and of Aalen-Johansen estimates in C) and D).

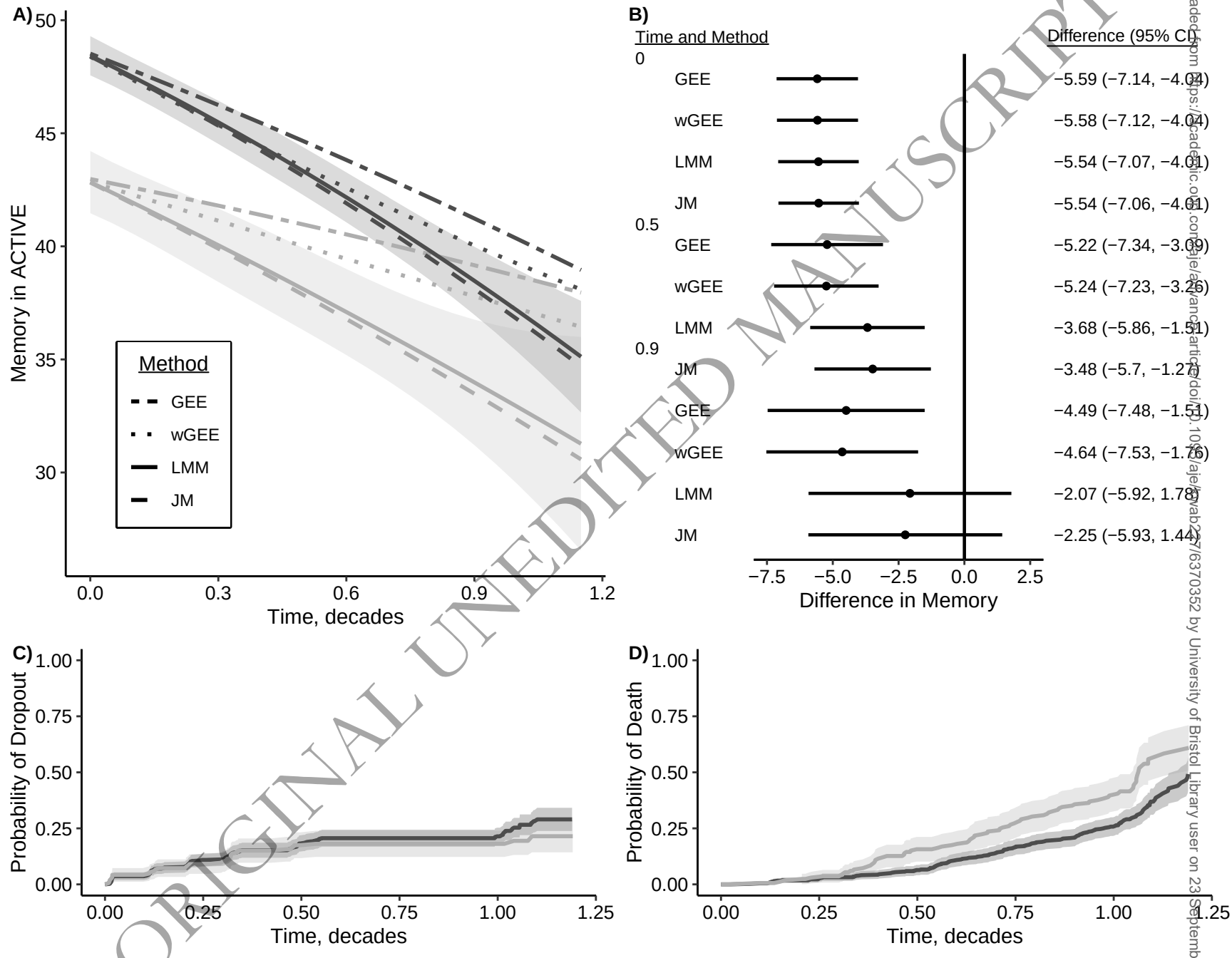

Figure 2. Comparison of the statistical approaches on memory in the Whitehall II study (2007-2016): A) Mean trajectories estimated by GEE, weighted GEE (wGEE), linear mixed model (LMM) and joint model (JM) for Caucasian males (light gray) and females (dark gray) entered at 75 years old, with low level of education; B) Male vs female difference in memory estimated by the four methods, at baseline, 5 and 9 years; C) non-parametric Aalen-Johansen estimates of the probability to drop out for males (light gray) and females (dark gray); D) non-parametric Aalen-Johansen estimates of the probability to die for males (light gray) and females (dark gray). Shaded areas represent 95% confidence intervals of LMM in A) and of Aalen-Johansen estimates in C) and D).

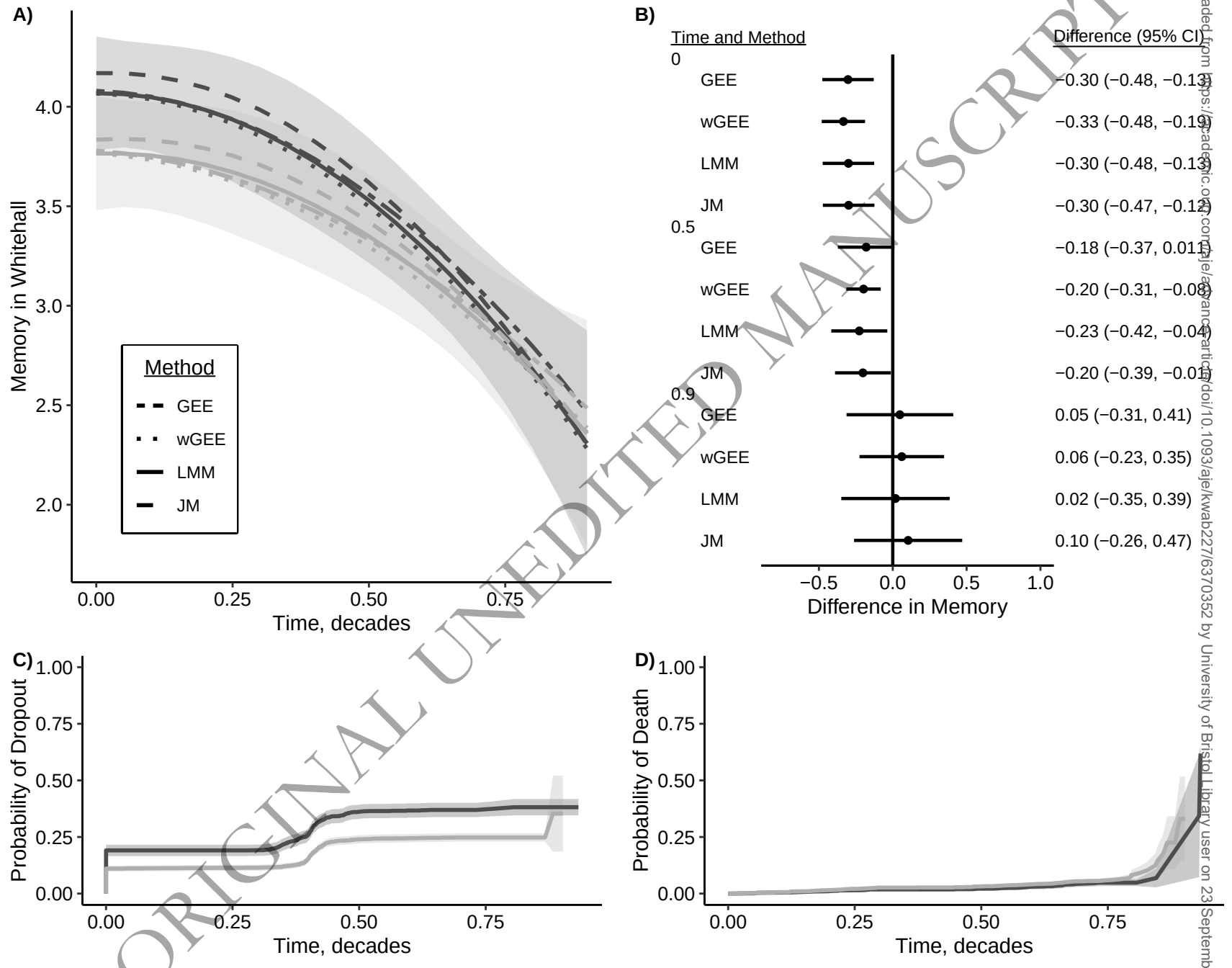

Figure 3. Comparison of the statistical approaches on memory in the WHICAP study (1992-2017): A) Mean trajectories estimated by GEE, weighted GEE (wGEE), linear mixed model (LMM) and joint model (JM) for non-Hispanic White males (light gray) or females (dark gray) entered at 75 years old in the 1992 recruitment wave, with low level of education; B) Male vs female difference in memory estimated by the four methods, at baseline, 5, 9 and 15 years; C) non-parametric Aalen-Johansen estimates of the probability to drop out for males (light gray) and females (dark gray); D) non-parametric Aalen-Johansen estimates of the probability to die for males (light gray) and females (dark gray). Shaded areas represent 95% confidence intervals of LMM in A) and of Aalen-Johansen estimates in C) and D).