ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Applied Statistics
Series C

Appl. Statist. (2016)
65, *Part* 1, *pp.* 131–144

Causal inference with longitudinal outcomes and non-ignorable dropout: estimating the effect of living alone on cognitive decline

Maria Josefsson and Xavier de Luna,

Umeå University, Sweden

Michael J. Daniels

University of Texas at Austin, USA

and Lars Nyberg

Umeå University, Sweden

[Received December 2013. Final revision March 2015]

Summary. We develop a model to estimate the causal effect of living arrangement (living alone *versus* living with someone) on cognitive decline based on a 15-year prospective cohort study, where episodic memory function is measured every 5 years. One key feature of the model is the combination of propensity score matching to balance confounding variables between the two living arrangement groups—to reduce bias due to unbalanced covariates at baseline, with a pattern–mixture model for longitudinal data—to deal with non-ignorable dropout. A fully Bayesian approach allows us to convey the uncertainty in the estimation of the propensity score and subsequent matching in the inference of the causal effect of interest. The analysis conducted adds to previous studies in the literature concerning the protective effect of living with someone, by proposing a modelling approach treating living arrangement as an exposure.

Keywords: Aging; Bayesian inference; Episodic memory; Non-ignorable missingness; Pattern–mixture model; Propensity score matching; Sensitivity

1. Introduction

Aging is accompanied by an average decline in mental abilities such as processing speed, attention and episodic memory (Schaie, 1994; Rönnlund *et al*., 2005). Accelerated decline in memory and other higher order cognitive functions has been much studied and linked to pathology such as dementia (Bäckman *et al*., 2001; Palmer *et al*., 2008). Preventive strategies against cognitive decline and early recognition of cognitive diseases are therefore of great interest.

In this paper our objective is to develop a modelling framework to estimate the causal effect of living arrangement (living alone *versus* living with someone) on cognitive functioning by using data from a large prospective cohort study: the Betula study (Nilsson *et al*., 1997). The cognitive outcome that we use is an episodic memory score measured every 5 years with up to 15 years of follow-up. Earlier studies have shown an association between living arrangement, social networks, marital status and cognitive decline and dementia (e.g. Fratiglioni *et al*. (2000),

Address for correspondence: Maria Josefsson, Department of Statistics, Umeå School of Business and Economics, Umeå University, Umeå 90187, Sweden.
E-mail: maria.josefsson@umu.se

 0035–9254/16/65131

van Gelder *et al*. (2006), Mousavi-Nasab *et al*. (2012) and Josefsson *et al*. (2012), the last also based on the Betula study). The protective effect of being married or living with someone in relation to cognition may be explained by the fact that a partner provides, in addition to practical support, emotional support and intellectual stimulation. Although it seems reasonable to infer that loss of social or emotional support could contribute to accelerated decline, it cannot be ruled out that relationships split if people start to suffer from cognitive dysfunction (perhaps ultimately affecting behaviour). In any case, studying causal effects on cognitive change brings in two complications. First, to reduce bias when inferring causal effects from observational data, the analysis must balance confounding variables between the two groups compared (here living alone and with someone) (Rubin, 2005). The second complication is that longitudinal data of older cohorts are prone to dropout or attrition, i.e. subjects are lost to follow-up and their responses are not observed thereafter. In studies of cognitive decline in elderly populations, dropout is often due to illness, e.g. dementia, or death, and may be directly related to the cognitive outcome of interest (Josefsson *et al*., 2012). In the presence of such non-ignorable dropout, conventional statistical methods may be invalid and lead to severely biased estimates (Little and Rubin, 2002). In addition, non-response due to illness and due to death must be treated differently since episodic memory score is not well defined after death. This is tackled by considering as the parameter of interest the causal effect on the subpopulation (stratum) of those always surviving irrespective of exposure (e.g. Imbens and Rubin (1997), Frangakis and Rubin (2002), Frangakis *et al*. (2007) and Chiba and VanderWeele (2011)). In the earlier-mentioned literature on the association between living arrangement and cognitive decline, only the analysis of Josefsson *et al*. (2012) took into account the possibility that dropout may be related to changes in cognitive ability. Whereas, in that study, living arrangement was shown to be associated with cognitive performance, we develop herein a modelling approach which focuses on the living arrangement variable as an exposure whose effect on cognitive decline we want to infer.

We propose a fully Bayesian approach to tackle the aforementioned complications in our study of living arrangement and its effect on cognitive decline. Specifically, our modelling approach extends previous work by combining methods for propensity score matching (Rosenbaum and Rubin, 1983a), to control for confounding variables, with a pattern–mixture model (PMM) for longitudinal data to deal with non-ignorable dropout. Matching on the propensity score, i.e. the probability of living alone given background variables, is a widely used procedure for balancing the distribution of variables in two groups that need to be compared (here living alone *versus* living with someone) and thus reduces bias due to observed confounders (Rosenbaum and Rubin, 1983a). A matched data set is formed by connecting each individual living alone with an individual living with someone having the most similar propensity score from the original data. Using a Bayesian approach allows us to convey the uncertainty in the estimation of the propensity score and subsequent matching in our final inferences.

PMMs have been developed for the analysis of longitudinal data subject to non-ignorable dropout (Little, 1994; Hogan and Laird, 1997). This approach stratifies the population by patterns of missing data (e.g. by the time of dropout) and the distribution of the complete population is a mixture over these patterns. We model the matched data with a mixed effect model for the episodic memory outcome (embedded within a PMM), where random effects are used to take into account the dependence structure due to both time and matching. The final inference is obtained from a mixture over dropout patterns.

Finally, we show how to investigate the sensitivity of the inference obtained to several of the assumptions made, including the possible omission of confounders in the covariate set that we match on, and the linear extrapolation of cognitive decline beyond dropout time points.

The paper is organized as follows. The Betula study and data are described in Section 2. In Section 3 parameters of interest and identification assumptions are presented. The model and methods are presented in Section 4. The results of our analysis using the Betula study and a concluding discussion follow in Sections 5 and 6 respectively.

2. The Betula project

The Betula Prospective Cohort Study (Nilsson *et al.*, 1997, 2004) is an on-going population-based study of 4445 adults aged 35–85 years at enrolment in Sweden. The objective is to study how memory functions change over time and to identify risk factors for dementia. Data have so far been collected every fifth year since 1988.

1966 participants were recruited at the first two measurement times (1988–1990 and 1993–1995). At recruitment the samples consisted of approximately 1000 individuals, evenly distributed over 11 age cohorts with 5-year intervals ranging from 35 to 85 years of age, and approximately balanced on sex. We focus on those participants who have been followed longitudinally for 15 years, and cognitively tested every fifth year. Each of the four visits included a health examination, questionnaires and an extensive battery of cognitive tests. The Regional Ethical Vetting Board at Umeå University approved this study and informed consent was obtained from all participants.

We have 1552 individuals with two or more measurements. Of these, 1213 individuals were living with someone and 339 living alone at enrolment. Memory was assessed by using the number of correct answers from a composite of three validated episodic memory tasks sensitive to mild cognitive deficits (Nilsson *et al.*, 1997). The episodic memory test scores can range between 0 and 76, with a higher score indicating better episodic memory (Josefsson *et al.*, 2012). To reduce skewness and the influence of floor or ceiling effects, i.e. lower or upper boundaries of the memory score (Morris *et al.*, 1999), a natural logarithmic transformation was applied to the outcome after adding 1.

Subjects were further subdivided into a cohort consisting of 562 subjects aged 65 years at baseline or older (old aged), and a cohort of 990 subjects 35–65 years of age (middle aged). The age of 65 years is used as the cut-off because episodic memory is known from previous studies (Rönnlund *et al.*, 2005) to show little decline before that age. The two groups defined by the 65 year-of-age cut-off also differ in that the death rate during follow-up is much lower in the younger group; see Section 3 for the methodological issues that are related to attrition by death. In the old cohort 203 subjects were living alone, and in the middle-age cohort 136 were living alone at enrolment.

Baseline demographic and lifestyle characteristics included age, sex and self-reported highest level of education (categorized into *high*, secondary school or higher (more than 9 years of education), *intermediate* (7–9 years) or *low*, elementary school (6–7 years of education)). Baseline health characteristics were self-reported health (feeling well, yes or no) and whether the participant had engaged in some physical activities during the previous 3 months. Table 1 provides summary statistics of these baseline covariates and the proportion of dropouts and survivors. For the old cohort, individuals living alone were older and were more likely to be female, less educated, having no physical activity in the previous 3 months and not feeling well. In the middle-aged cohort, the main difference was more females living alone.

Only 45% of the participants completed the fourth test wave. Reasons for attrition included withdrawal, moving out of the area, sickness and dementia, death, no contact and unknown or unspecified reasons. In the event of death after a participant drops out, information regarding the participant's date of death is available.

Table 1. Summary statistics for observed covariates at enrolment and attrition rate stratified by living arrangement for the middle-aged and old groups†

Statistic	*Results for old group*		*Results for middle-aged group*	
	Living with someone (N = 359)	*Living alone (N = 203)*	*Living with someone (N = 854)*	*Living alone (N = 136)*
Mean age (years) (standard deviation)	70.9 (5.6)	73.4 (5.8)	48.4 (8.0)	49.8 (8.4)
Education: high (%)	15.9	12.8	52.9	52.2
Education: intermediate (%)	29.8	20.2	27.1	24.3
Education: low (%)	54.3	67.0	20.0	23.5
Physical activity (%)	88.9	81.3	94.4	91.2
Females (%)	42.1	83.7	51.4	61.0
Feeling well (%)	80.2	68.5	80.6	73.5
Dropouts (%)	34.5	29.1	22.1	29.4
Survivors (%)	59.9	57.6	96.3	91.9

†There were 44 (23 in the old and 21 in the middle-aged group) individuals with intermittent missing observations.

3. Parameter of interest and identification assumptions

We use the potential outcome framework for causal inference to define the parameter of interest—a causal effect of 'living alone' on the ability to maintain cognitive function, and to state formally under which assumptions such a parameter is identified.

Let A_i denote the binary exposure, where here $A_i = 1$ if the ith individual ($i = 1, \ldots, n$) is living alone at baseline and $A_i = 0$ if she or he is living with someone. Let $\mathbf{W}_i$ be a $p \times 1$ vector of baseline covariates including an intercept and potential confounders, age, education, sex, physical activity and feeling well. Each individual has two sets of potential outcomes: $\mathbf{Y}_i(0) = \{Y_{it}(0); t \in \mathbb{R}^+\}$, vector of longitudinal outcomes (logarithm of the episodic memory score) if living with someone at baseline $t = 0$ and $\mathbf{Y}_i(1) = \{Y_{it}(1); t \in \mathbb{R}^+\}$ longitudinal outcome if living alone at baseline. These potential outcomes are further modelled by using random effects to allow for individual slopes (rate of change in memory score). For $i = 1, \ldots, n$,

$$Y_{it}(0) = B_{0i}(0) + B_{1i}(0)t + \epsilon_{it}(0),$$
$$Y_{it}(1) = B_{0i}(1) + B_{1i}(1)t + \epsilon_{it}(1),$$

where $B_{0i}(0)$ and $B_{0i}(1)$ are random intercepts, $B_{1i}(0)$ and $B_{1i}(1)$ random slopes and $\epsilon_{it}(0)$ and $\epsilon_{it}(0)$ mean 0 independent random errors. Let us also define two potential survival indicators for each individual: $S_i(0)$, which takes value 1 if individual i would survive under the follow-up period if living with someone at baseline and 0 otherwise, and $S_i(1)$, which takes value 1 if individual i would survive under the follow-up period if living alone at baseline and 0 otherwise. We assume throughout that the units do not interfere with each other, i.e. the exposure of a given individual does not affect the potential outcomes and survival of the others, and that the observed outcome and survival can be written as $\mathbf{Y}_i = \{A_i Y_{it}(1) + (1 - A_i) Y_{it}(0); t \in \mathbb{R}^+\}$ and $S_i = A_i S_i(1) + (1 - A_i) S_i(0)$ (Rubin, 1991).

Because cognition is not defined after death we define a causal parameter for those surviving both under and without exposure, i.e. such that $S_i(0) = S_i(1) = 1$. This is the always-survivors stratum: one of four strata obtained by varying the values taken by $S_i(0)$ and $S_i(1)$ (Barnard *et al.*, 2003; Frangakis *et al.*, 2007; Mattei and Mealli, 2007; Roy *et al.*, 2008; Gallop *et al.*, 2009;

Chiba and VanderWeele, 2011; Frumento *et al.*, 2012). We can now define a main parameter of interest for those surviving irrespectively of exposure—a principal stratum causal effect: the average effect (over those living alone at baseline and always surviving) of living alone at baseline on the rate of change (slope) in memory score,

$$\tau = E[B_{1i}(1) - B_{1i}(0) | A_i = 1, S_i(0) = S_i(1) = 1].$$

Conditioning on $A = 1$ puts the focus on the average effect over the exposed population, thereby answering the question what is the effect of living alone at baseline (compared with living with someone) for those who were actually living alone at baseline? This parameter is not in general identified unless some conditions hold. Two issues need to be considered: first, exposure is not randomized and, second, that we do not observe the always-survivors stratum.

Let us therefore make the following assumptions.

Assumption 1.

$$A_i \perp\!\!\!\perp (B_{1i}(0), B_{1i}(1), S_i(0), S_i(1)) | \mathbf{W}_i, \text{ and } 0 < \Pr(A_i = 1 | \mathbf{W}_i) < 1.$$

Assumption 2. Let $\Pr\{S_i(a) = 1 | A_i = 1, \mathbf{W}_i\} = \psi_a(\mathbf{W}_i)$, $a = 0, 1$, and

$$\Pr\{S_i(0) = 1 | S_i(1) = 1, A_i = 1, \mathbf{W}_i\} = \psi_0(\mathbf{W}_i) + \lambda \left[\min\left\{ 1, \frac{\psi_0(\mathbf{W}_i)}{\psi_1(\mathbf{W}_i)} \right\} - \psi_0(\mathbf{W}_i) \right], \qquad 0 \leqslant \lambda \leqslant 1.$$

Assumption 3. For $a = 0, 1$,

$$E[B_{1i}(a) | A_i = 1, S_i(1) = S_i(0) = 1, \mathbf{W}_i] - E[B_{1i}(a) | A_i = 1, S_i(1) = 1, S_i(0) = 0, \mathbf{W}_i] = \Delta_a.$$

Assumption 1 is called strong ignorability and holds when $\mathbf{W}_i$ contains all pre-exposure covariates that are related to both exposure and potential outcomes and survival. Assumption 2 was used in similar contexts in Roy *et al.* (2008) (see section 4.2 for a derivation) and Lee *et al.* (2010). Assumption 2 imposes restrictions on $\Pr\{S_i(0) = 1 | S_i(1) = 1, A_i = 1, \mathbf{W}_i\}$ which are different depending on the value of λ. Thus, if $\lambda = 0$ then $\Pr\{S_i(0) = 1 | S_i(1) = 1, A_i = 1, \mathbf{W}_i\} = \Pr\{S_i(0) = 1 | A_i = 1, \mathbf{W}_i\}$ (a conditional independence statement) and if $\lambda = 1$ then $\Pr\{S_i(0) = 1 | S_i(1) = 1, A_i = 1, \mathbf{W}_i\} = 1$ or as large as possible while compatible with the marginals (the latter is a weaker form of monotonicity assumption which is commonly used in the principal stratification literature). In assumption 2 we can vary λ between 0 and 1, thereby allowing for situations lying somewhere in between these two extreme cases. Finally, assumption 3 says that the differences in slopes when comparing different strata defined by $S_i(0)$ and $S_i(1)$ does not depend on the pre-exposure covariates $\mathbf{W}_i$.

Let us now consider the identifiability of the parameter τ on the basis of these assumptions. We have, for $a = 0, 1$,

$$\begin{aligned} E[B_{1i}(a) | A_i = 1, S_i(a) = 1, \mathbf{W}_i] = \sum_k & E[B_{1i}(a) | A_i = 1, S_i(a) = 1, S_i(1-a) = k, \mathbf{W}_i] \\ & \times \Pr\{S_i(1-a) = k | S_i(a) = 1, A_i = 1, \mathbf{W}_i\}. \end{aligned}$$

Noting that

$$\begin{aligned} E[B_{1i}(a) | A_i = 1, S_i(a) = 1, \mathbf{W}_i] &= E[B_{1i}(a) | A_i = a, S_i(a) = 1, \mathbf{W}_i] \\ &= E[B_{1i}(a) | A_i = a, S_i = 1, \mathbf{W}_i], \end{aligned}$$

the first equality by assumption 1 for $a = 0$, we have

$$E[B_{1i}(a)|A_i=1,S_i(1)=S_i(0)=1,\mathbf{W}_i]$$
$$=\frac{E[B_{1i}(a)|A_i=a,S_i=1,\mathbf{W}_i]-\Delta_a(\mathbf{W}_i)\Pr\{S_i(1-a)=0|S_i(a)=1,A_i=1,\mathbf{W}_i\}}{\Pr\{S_i(1-a)=1|S_i(a)=1,A_i=1,\mathbf{W}_i\}}. \quad (1)$$

By assumptions 2 and 3, this expression is identified up to λ and Δ_a, $a=0,1$. For instance, if we assume that $\Delta_0=\Delta_1=0$, $E[B_{1i}(a)|A_i=a,S_i=1,\mathbf{W}_i]$ and $\Pr\{S_i(1-a)=1|S_i(a)=1,A_i=1,\mathbf{W}_i\}$ can be modelled and fitted to the observed data, the latter up to parameter λ by assumption 2; see Section 4.5 for further details.

Let us further denote the propensity score $e(\mathbf{W}_i;\gamma)=\Pr(A_i=1|\mathbf{W}_i;\gamma)$, here assumed to be known up to parameter γ. The propensity score plays a central role in the causal inference literature because of a result by Rosenbaum and Rubin (1983a) whereby assumption 1 implies that

$$A_i \perp\!\!\!\perp (B_{1i}(0), B_{1i}(1), S_i(0), S_i(1))|e(\mathbf{W}_i;\gamma),$$

and thus the analysis can be carried out by conditioning on the propensity score instead of the vector $\mathbf{W}_i$. A popular implementation of this theoretical result is matching as described in the next section. Note that one of several alternatives to matching is to use inverse probability weighting estimators, whereby the propensity score is used as a reweighting tool instead of as an imputation device (Tsiatis, 2006; O'Muircheartaigh and Hedges, 2014). Matching has, however, been shown to have better finite sample properties (Waernbaum, 2012).

The existence of an unobserved confounder such that assumption 1 does not hold can never be excluded. To characterize the bias that could result, we follow Rosenbaum and Rubin (1983b) and let $B_{1i}(0)=\tilde{B}_{1i}(0)+\zeta_1 V_i$, where V_i, a hypothetical confounder, is Bernoulli distributed with probability of success $\frac{1}{2}$, such that assumption 1 holds when replacing $B_{1i}(0)$ with $\tilde{B}_{1i}(0)$. We assume that $V_i \perp\!\!\!\perp \mathbf{W}_i$ and $V_i \perp\!\!\!\perp S_i(0)|A_i,\mathbf{W}_i$. The dependence between V_i and A_i conditionally on $\mathbf{W}_i$ is further described by

$$\zeta_2=\log\left\{\frac{\Pr(A_i=1|V_i=1,\mathbf{W}_i)/\Pr(A_i=0|V_i=1,\mathbf{W}_i)}{\Pr(A_i=1|V_i=0,\mathbf{W}_i)/\Pr(A_i=0|V_i=0,\mathbf{W}_i)}\right\}.$$

Then assumption 1 does not hold for $B_{1i}(0)$ if $\zeta_1\zeta_2\neq 0$. In such a case the bias that is implied by mistakenly assuming $\zeta_1\zeta_2=0$ (assumption 1) is obtained by comparing

$$E[B_{1i}(0)|A_i=1,S_i(0)=1,\mathbf{W}_i]=E[\tilde{B}_{1i}(0)|A_i=0,S_i(0)=1,\mathbf{W}_i]+\zeta_1 E[V_i|A_i=1,\mathbf{W}_i],$$

and

$$E[B_{1i}(0)|A_i=0,S_i(0)=1,\mathbf{W}_i]=E[\tilde{B}_{1i}(0)|A_i=0,S_i(0)=1,\mathbf{W}_i]+\zeta_1 E[V_i|A_i=0,\mathbf{W}_i],$$

where we have used the assumptions that were made on $\tilde{B}_{1i}(0)$ and V_i. Thus, the bias due to $\zeta_1\zeta_2\neq 0$ is

$$\zeta_1 E[E[V_i|\mathbf{W}_i,A_i=1]-E[V_i|\mathbf{W}_i,A_i=0]|A_i=1]. \quad (2)$$

If $\zeta_1\zeta_2=0$, i.e. either $\zeta_1=0$ or $\zeta_2=0$, then the bias (2) is 0. When $\zeta_2=0$, V_i and A_i are then independent conditionally on $\mathbf{W}_i$, so expression (2) is equal to 0.

Thus, sensitivity to assumptions 2 and 3 can be investigated by varying λ and Δ_a, $a=0,1$, whereas sensitivity to the existence of unobserved confounder (assumption 1) can be investigated by varying $\zeta_1\zeta_2$ in expression (2).

4. Models and methods

With the goal of estimating the effect of the exposure 'living alone' on the ability to maintain

cognitive function, we use a hierarchical modelling approach and Bayesian inference which allows us to take into account the complex dependence structure of the data and to propagate uncertainty appropriately. We first introduce a Bayesian approach for propensity score matching. We then introduce a mixed effect model for the observed longitudinal outcome that accounts for non-ignorable non-response and accounts for the dependence structure in the data by using random effects.

4.1. Propensity score model for matching

We use a logistic regression model for the propensity score $e(\mathbf{W}_i;\gamma)$:

$$\log\left\{\frac{e(\mathbf{W}_i;\gamma)}{1-e(\mathbf{W}_i;\gamma)}\right\}=\gamma'\mathbf{W}_i. \tag{3}$$

We then use matching on $e(\mathbf{W}_i;\hat{\gamma})$, where $\hat{\gamma}$ is a consistent estimator of γ, to control for $\mathbf{W}_i$ (Rosenbaum and Rubin, 1983a). Thus, each exposed (i.e. living alone) subject is matched with a control (i.e. not living alone) subject who is closest in terms of the propensity score; we reuse controls here such that each control can be matched to multiple cases. For the models in the next section, we define matched groups to be the control and all exposed cases that are matched to it.

4.2. Linear mixed model for longitudinal matched data

We consider a linear mixed model (Gelman and Hill, 2007) that allows for subject and (matched) group variance components. The subject level intercept and slope take into account the temporal dependence and the group level takes into account the dependence that is introduced through the matching process. The model that we consider is

$$\mathbf{Y}_{ij}=\mathbf{X}_{ij}\boldsymbol{\beta}+\mathbf{Z}_{ij}\mathbf{u}_i+\mathbf{Z}_{ij}\mathbf{b}_{ij}+\boldsymbol{\epsilon}_{ij}, \tag{4}$$

where $\mathbf{Y}_{ij}=(Y_{ij1},\ldots,Y_{ijT})'$ is the vector of observed longitudinal outcomes (the logarithm of episodic memory score) for subject j ($j=1,\ldots,N_i$) in (matched) group i ($i=1,\ldots,N$). $\mathbf{X}_{ij}$ is a design matrix (intercept and time from enrolment and A_{ij}, and their interactions) and $\boldsymbol{\beta}$ is the vector of corresponding fixed effect coefficients. $\mathbf{Z}_{ij}$ is the random-effects design matrix (intercept and time from enrolment) with random effects $\mathbf{u}_i$ and $\mathbf{b}_{ij}$ for matched group and individual respectively. We assume that $\mathbf{u}_i\sim N(\mathbf{0},\Sigma)$, $\mathbf{b}_{ij}\sim N(\mathbf{0},\Omega)$ and $\boldsymbol{\epsilon}_{ijt}\sim N(\mathbf{0},\sigma^2 I)$, where I denotes the identity matrix, and that all are mutually independent.

4.3. Pattern–mixture model for non-ignorable missing data

To account for non-ignorable dropout, we specify a PMM that stratifies the population by the subject's last measurement wave, denoted by M; M takes values 2, 3 or 4, where 4 is the last follow-up (year 15). The quantities of interest in this model are the population-average regression parameters (e.g. the intercept and slope) which are a weighted average of the pattern-specific parameters across the three dropout patterns. The model is based on the factorization

$$p(\mathbf{Y}_{ij},\mathbf{u}_i,\mathbf{b}_{ij},M_{ij}|\mathbf{X}_{ij})=p(\mathbf{Y}_{ij}|\mathbf{u}_i,\mathbf{b}_{ij},\mathbf{X}_{ij},M_{ij})\,p(\mathbf{u}_i)\,p(\mathbf{b}_{ij}|M_{ij})\,p(M_{ij}).$$

This mixture model is specified in four components: the first and second were specified in Section 4.2, but now with an augmented design matrix $\mathbf{X}_{ij}$ that includes the intercept β_0, time ($t=1,2,\ldots,T$, β_1), A_{ij} (binary exposure, β_2), $A_{ij}\times$ time (β_3), M_{ij} (dropout, a factor with three levels, β_4, β_5), $M_{ij}\times$ time (β_6, β_7), $M_{ij}\times A_{ij}$ (β_8, β_9) and $M_{ij}\times$ time $\times A_{ij}$ (β_{10}, β_{11}).

For the third component, we specify $\mathbf{b}_{ij}|M_{ij}=m \sim N(\mathbf{0}, \Omega_m)$, $m=2,3,4$. We also assume that $\mathbf{u}_i$ is independent of dropout pattern, since subjects within one matched group may belong to different dropout patterns. The last term, the marginal distribution of M_{ij}, follows a multinomial distribution with parameters $\pi_{m,a}=\Pr(M_{ij}=m|A_{ij}=a)$, where $\Sigma_{m=2}^{4}\pi_{m,a}=1, a=0,1$.

For the PMM, the fixed effects parameters for the population (unconditionally on drop-out) are a weighted average of the dropout-specific patterns, i.e. intercept

$$\alpha_0=\beta_0\pi_{4,0}+(\beta_0+\beta_4)\pi_{2,0}+(\beta_0+\beta_5)\pi_{3,0}, \tag{5}$$

time fixed effect

$$\alpha_1=\beta_1\pi_{4,0}+(\beta_1+\beta_6)\pi_{2,0}+(\beta_1+\beta_7)\pi_{3,0}, \tag{6}$$

living alone (A_{ij}) fixed effect

$$\alpha_2=\beta_2\pi_{4,1}+(\beta_2+\beta_8)\pi_{2,1}+(\beta_2+\beta_9)\pi_{3,1} \tag{7}$$

and the interaction of living alone and time ($A_{ij}\times$ time),

$$\alpha_3=\beta_3\pi_{4,1}+(\beta_3+\beta_{10})\pi_{2,1}+(\beta_3+\beta_{11})\pi_{3,1}. \tag{8}$$

Section 4.5 will relate these parameters to the causal effect of interest τ.

Intermittent missing data (Table 1) are implicitly handled in the mixture model under an assumption of partial ignorability (Harel and Schaffer, 2009), i.e., conditionally on pattern, intermittent missingness is assumed ignorable.

4.4. Bayesian inference: model, prior specification and sampling algorithms

We want to make inference in the mixed model conditional on matched sets (based on the propensity score). However, we want our approach to account appropriately for the uncertainty in the propensity score (and the subsequent matched sets). As such we conduct the analysis in two pieces. We first fit a propensity score model to obtain the posterior distribution of its parameters (and the corresponding matched sets). We can then draw such parameters values from these posteriors, yielding for each such draw a new matched set. Then, conditionally on each matched set, we fit the mixed model. Note that we are not using a joint likelihood as the two components above do not factor (Zigler, 2013); instead we are doing the inference in two stages to account appropriately for uncertainty in the matching (and to avoid feedback) as was recommended in Zigler (2013).

First, we describe sampling from the posterior distribution of the parameters in the propensity score model with a subsequent matching algorithm. To be more formal, we define a parameter $\boldsymbol{\eta}(\gamma)$ that identifies pairs on the basis of the propensity score, and which is a deterministic function of γ and $\mathbf{W}$. This parameter is of dimension n_1: the number of exposed individuals in the study. The implicit prior on component η_k, $k=1,\ldots,n_1$, is $p(\boldsymbol{\eta}|\gamma,\mathbf{W})=I\{\boldsymbol{\eta}\equiv g(\gamma,\mathbf{W})\}$, where the function g determines the matching on the basis of the covariates and the value of the parameter γ. For each draw of γ, we create the matched pairs by $\boldsymbol{\eta}$ by using the matching function $g(\gamma,\mathbf{W})$. The matching function matches with respect to the propensity score $e(\mathbf{W}_j;\gamma)$, i.e. each exposed subject is matched with a control subject who is closest to it in terms of the propensity score (with replacement). Thus, for each draw from the posterior for γ, subjects may have different matches. We specified a diffuse normal prior for the parameter γ.

In a second step, we sample from the posterior distribution of the parameters from the linear mixed models that were defined in Sections 4.2 and 4.3 given the matched pairs for each draw of

$(\boldsymbol{\gamma}, \boldsymbol{\eta})$, $p(\boldsymbol{\beta}, \mathbf{u}, \mathbf{b}, \Sigma, \Omega, \sigma^2, \boldsymbol{\pi} | \mathbf{A}, \mathbf{X}, \mathbf{Z}, \mathbf{Y}, \boldsymbol{\eta}, \boldsymbol{\gamma})$, using Markov chain Monte Carlo methods. Details on the priors follow.

For the fixed effects $\boldsymbol{\beta}$, we specify diffuse independent normal priors, with zero mean and large variances; for the precision parameter $1/\sigma^2$, we specify a non-informative gamma prior with shape and scale parameters equal to 0.001 (Gelman *et al.*, 2013).

We decompose the random-effects covariance matrices for the group **u** and subject **b** into a set of unconstrained dependence parameters and a set of variance parameters, following the approach that was described in Daniels and Zhao (2003). From this decomposition, we obtain a more flexible prior specification and, via the parameters, we can easily allow the matrix to depend on covariates. The random intercept for group is modelled as $u_{i1} \sim N(0, \sigma^2_{u,1})$, and the random slope from group, conditionally on the random intercept, is $u_{i2}|u_{i1} \sim N(\varphi u_{i1}, \sigma^2_{u,2})$. Exploration of the data suggested allowing the covariance matrix for a subject to differ by missing data pattern (which motivated the use of this decomposition). As such, we allowed the dependence parameter ϕ and variance parameters $\sigma^2_{b,l}$, $l = 1, 2$, to vary as a function of dropout pattern M, $b_{ij1}|M_{ij} = m \sim N(0, \sigma^2_{b,1m})$, and $b_{ij2}|b_{ij1}, M_{ij} = m \sim N(\phi_m b_{ij1}, \sigma^2_{b,2m})$, for $m = 2, 3, 4$.

For the parameters of the group level random-effect covariance matrix, we specify a diffuse normal prior for the dependence parameter, and diffuse half-normal priors for the variances. For the subject level random-effects covariance matrix, we specify priors that shrink the dropout-specific group parameters to common values across patterns, ϕ^* and $\delta_{b,l}$. In particular, we specify a normal prior for the dependence parameter $\phi_m \sim N(\phi^* \sigma_\phi)$ and gamma priors for the precision parameters, $1/\sigma^2_{b,lm} \sim$ gamma$(\delta_{b,l}/\sigma^2_{b,l}, \delta_{b,l})$. The 'location' parameters of these priors are given a diffuse normal prior for ϕ^* and a diffuse half-normal prior for $1/\sigma^2_{b,l}$. For the parameters that control shrinkage, σ_ϕ, $\delta_{b,1}$ and $\delta_{b,2}$, we specify uniform shrinkage priors (Daniels, 1999); see the on-line supplementary material for the exact form of these priors. In terms of the variance at the first level, a standard choice is a diffuse inverse gamma prior. However, such priors are not a good choice for random-effects variances (see, for example, Gelman (2006)); as such, we use the shrinkage priors and half-normal prior for these variances. Also, a sensitivity analysis showed that our results are robust to these choices (see the on-line supplementary material). Finally, we specify Dirichlet priors for $(\pi_{2,A}, \pi_{3,A}, \pi_{4,A})$, $A = 0, 1$, with hyperparameter vector $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$; the resulting posteriors are Dirichlet distributions.

For each sample from $p(\boldsymbol{\beta}, \mathbf{u}, \mathbf{b}, \Sigma, \Omega, \sigma^2, \boldsymbol{\pi} | A, \mathbf{X}, \mathbf{Z}, \mathbf{Y}, \boldsymbol{\eta}, \boldsymbol{\gamma})$, the marginalized parameters in the PMM are computed as the weighted averages of the pattern-specific parameters as specified in equations (5)–(8).

To obtain samples from the posterior distributions, we use OpenBUGS called from R (`R2OpenBUGS` function; Sturtz *et al.* (2005)). For the matching function g we use the `Match` function in R (`Matching` package (Sekhon, 2011)).

For our analysis, we ran chains of length 100 000 for each set of matched pairs (after a burn-in of 20 000 and thinned by 20). Chains were carefully monitored for convergence and mixing via trace plots. Samples from 200 matched sets were used to obtain the target posterior distribution and to ensure negligible Monte Carlo error.

4.5. Causal effect revisited

As defined in Section 3, the causal parameter of interest in the current study is τ. We re-express τ, which is a function of Δ_a, $a = 0, 1$, and λ, by noting that $\alpha_1 + u_{2i}$ corresponds to $E[B_{1i}(0)|A_i = 1, S_i = 1, \mathbf{W}_i]$ in equation (1), and $\alpha_1 + \alpha_3 + u_{2i}$ to $E[B_{1i}(1)|A_i = 1, S_i = 1, \mathbf{W}_i]$ in equation (1). Thus, if we assume that $\Delta_0 = \Delta_1 = 0$ then we obtain (with a slight abuse of index notation)

$$\tau = E\left[\frac{\alpha_1+\alpha_3+u_{2i}}{\Pr\{S_i(0)=1|S_i(1)=1,A_i=1,\mathbf{W}_i\}} - \frac{\alpha_1+u_{2i}}{\Pr\{S_i(1)=1|S_i(0)=1,A_i=1,\mathbf{W}_i\}}\middle| A_i=1, S_i(1)=S_i(0)=1\right], \quad (9)$$

where the denominators are obtained by modelling and fitting the functions $\psi_0(\mathbf{W}_i)=\Pr\{S_i(0)=1|A_i=1,\mathbf{W}_i\}$ by assumption 1, and $\psi_1(\mathbf{W}_i)$ in assumption 2 by using logistic regressions. Furthermore, averaging the argument of the expectation above over a matched sample corresponds to averaging over $p(\mathbf{W}_i|A_i=1,S_i=1)$. To target equation (9) and to average over $p\{\mathbf{W}_i|A_i=1,S_i(1)=S_i(0)=1\}$, we take a weighted average of a matched sample with weights

$$\frac{p\{\mathbf{W}_i|A_i=1,S_i(1)=S_i(0)=1\}}{p(\mathbf{W}_i|A_i=1,S_i=1)}. \quad (10)$$

To estimate these weights we (non-parametrically) model $p(\mathbf{W}_i|A_i=1)$, where $\mathbf{W}_i$ involve only categorical covariates, using a multinomial distribution with an equal probability Dirichlet prior on the probability vector. The denominator in expression (10) is then obtained by noting that

$$p(\mathbf{W}_i|A_i=1,S_i=1)=\frac{p(S_i=1,\mathbf{W}_i|A_i=1)}{p(S_i=1|A_i=1)},$$

which is equal to $\psi_1(\mathbf{W}_i)\,p(\mathbf{W}_i|A_i=1)/p(S_i=1|A_i=1)$. Further, note that the numerator in expression (10) is equal to $p\{S_i(1)=S_i(0)=1,\mathbf{W}_i|A_i=1\}/p\{S_i(1)=S_i(0)=1|A_i=1\}$, where $p\{S_i(1)=S_i(0)=1,\mathbf{W}_i|A_i=1\}$ is identified up to λ by assumption 2 and $p\{S_i(1)=S_i(0)=1|A_i=1\}$ is obtained by averaging $p\{S_i(1)=S_i(0)=1,\mathbf{W}_i|A_i=1\}$ over $p(\mathbf{W}_i|A_i=1)$.

5. Results

For the propensity score model that was described in Section 4.1, we included the covariates age, education, gender, feeling well and physical activity. We specified two separate models: one for old subjects, ages 65 years and older, and one for middle-aged subjects, ages 35–60 years. To allow for non-linearity in the effect of age, the latter was categorized into the sampled age cohorts 35, 40,...,60 years for the middle-aged subjects, and 65, 70,..., 85 years for the old group. Details on the fitted propensity score model can be found in the on-line supplementary material, where χ^2-tests are also conducted to examine covariate balance across the two exposure groups after matching.

Posterior summaries of the parameters in the linear mixed model that was described in Section 4.3 are found in Table 2, for both the old and the middle-aged groups. Analyses are based on matched pairs where both the exposed and the control subject survived past the fourth test wave (analysis on survivors; see Section 3). For the old group, subjects who dropped out after the second test wave (dropouts after year 5) had lower episodic memory score at baseline compared with subjects who remained in the study (the 95% credible interval (CI) just covers zero) and had greater decline in episodic memory score over time (the 95% CI excludes zero). Similar tendencies are observed for those dropping out after year 10 (after the third test wave). In the middle-aged group, several dropout indicators are negative, although with 95% CI including zero, indicating a weaker association of dropout with episodic memory.

Posterior medians and 95% CIs for the pattern-specific regression parameters marginalized over dropout patterns (see equations (5)–(8)) are given in Table 3 and the corresponding trajectories are plotted in Fig. 1. For the old group, subjects living with someone had a significant decrease in memory score of 13% (on the original memory score scale) over the study period,

Table 2. Posterior medians and 95% CIs for parameters of the mixed model presented in Sections 4.2 and 4.3 for the middle-aged and old groups

Covariate	*Results for old group*		*Results for middle-aged group*	
	Median	*95% CI*	*Median*	*95% CI*
Intercept (β_0)	3.58	(3.46, 3.70)	3.70	(3.64, 3.76)
Time (β_1)	−0.06	(−0.12, −0.01)	−0.01	(−0.04, 0.01)
Living alone (β_2)	−0.11	(−0.26, 0.04)	−0.02	(−0.09, 0.05)
Living alone × time (β_3)	−0.06	(−0.13, 0.01)	0.00	(−0.02, 0.03)
Dropouts after year 5 (β_4)	−0.21	(−0.44, 0.02)	−0.07	(−0.26, 0.12)
Dropouts after year 10 (β_5)	−0.19	(−0.40, 0.02)	−0.05	(−0.28, 0.15)
Dropouts after year 5 × time (β_6)	−0.20	(−0.36, −0.03)	0.03	(−0.12, 0.17)
Dropouts after year 10 × time (β_7)	−0.06	(−0.17, 0.05)	−0.02	(−0.10, 0.06)
Living alone × dropouts after year 5 (β_8)	−0.03	(−0.35, 0.28)	0.03	(−0.20, 0.27)
Living alone × dropouts after year 10 (β_9)	0.11	(−0.19, 0.40)	−0.01	(−0.24, 0.26)
Living alone × dropouts after year 5 × time (β_{10})	0.15	(−0.07, 0.37)	−0.10	(−0.28, 0.08)
Living alone × dropouts after year 10 × time (β_{11})	−0.01	(−0.16, 0.14)	−0.00	(−0.10, 0.09)
Variance components				
σ^2	0.03	(0.02, 0.04)	0.02	(0.01, 0.02)
Group level				
Intercept	0.16	(0.06, 0.25)	0.08	(0.01, 0.14)
Time	0.05	(0.02, 0.09)	0.04	(0.01, 0.06)
Correlation	−0.22	(−0.95, 0.81)	0.03	(−0.87, 0.91)
Subject level				
Intercept	0.22	(0.15, 0.32)	0.19	(0.14, 0.30)
Time	0.10	(0.07, 0.17)	0.05	(0.02, 0.07)
Correlation	0.63	(−0.01, 0.97)	0.35	(−0.21, 0.92)

95% CI (−20%,−7%). The marginal score at baseline and also the slope, among subjects living alone, was smaller, 9%, 95% CI (−20%,2%), and 3%, 95% CI (−10%,5%) respectively, compared with subjects living with someone. The last two effects were negative with posterior probabilities of 0.954 and 0.757 respectively. Finally, the principal stratum causal effect of living alone on the rate of change in cognition, τ, was −2% with 95% CI (−12%,10%). For the middle-aged group, the differences between the two living arrangement groups were smaller, and no principal stratum causal effect was computed since the death rate was marginal (only 42 died during the follow-up period, including only 11 living with someone; thus, fitting such a model for $\psi_1(\mathbf{W}_i)$ led to non-convergence). Exploratory model checks and an omnibus posterior predictive check (Gelman *et al.*, 1996) appropriate for incomplete data (Daniels *et al.*, 2012), which are described in the on-line supplementary material, did not indicate any problems with model fit.

Finally, several aspects of our approach allow for sensitivity analyses. We thus examined in the on-line supplementary material sensitivity to the omission of unobserved confounders (assumption 1), alternatives to a monotonicity assumption (assumption 2) and various non-ignorable missingness mechanisms, e.g. via changes in the slope (cognitive decline) after dropout (Daniels and Hogan, 2008). Our conclusions did not appear to be sensitive to these assumptions. The

Table 3. Posterior medians and 95% CIs for marginal parameters, and principal stratum causal effect, for the middle-aged and old groups†

Effect	*Results for old group*		*Results for middle-aged group*	
	Median	*95% CI*	*Median*	*95% CI*
Intercept (α_0)	3.46	(3.36, 3.56)	3.68	(3.63, 3.74)
Time (α_1)	−0.13	(−0.20, −0.07)	−0.01	(−0.04, 0.01)
Living alone (α_2)	−0.09	(−0.20, 0.02)	−0.02	(−0.08, 0.05)
Living alone × time (α_3)	−0.03	(−0.10, 0.05)	−0.01	(−0.05, 0.02)
Principal stratum causal effect (τ)	−0.02	(−0.12, 0.10)	—	

†The principal stratum causal effect is given for $\lambda = 1$.

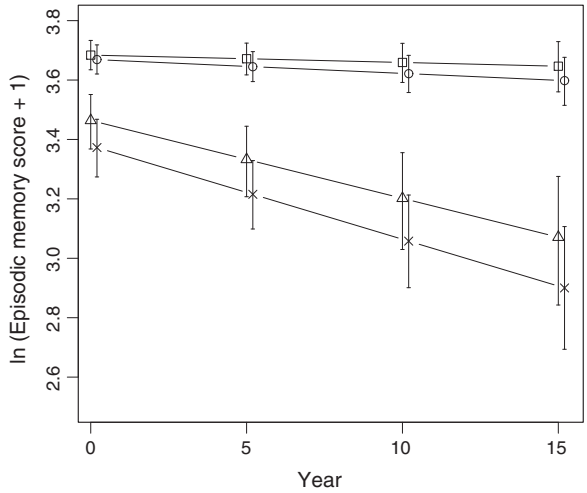

Fig. 1. Posterior median episodic memory score over 15 years, and pointwise 95% CIs, obtained from the PMM presented in Section 4.3, stratified by living arrangement for the middle aged and old: □, living with someone, middle aged; ○, living alone, middle aged; Δ, living with someone, old; ×, living alone, old

supplementary material also provides analyses using simpler modelling approaches (using the missingness at random assumption and ignoring uncertainty in propensity score fit) for comparison. Under the assumption of missingness at random, an underestimation of the negative time trend in cognition was observed, and an analysis considering the estimated propensity score fixed resulted in similar median scores for equations (5)–(8), although the CIs were tighter.

6. Discussion

We have presented a framework for estimating the effect of living alone on episodic memory function over time, based on a longitudinal study. The approach proposed used a propensity score matching method to balance baseline covariates for the two comparison groups (living alone and with someone), followed by a pattern–mixture approach based on mixed models to account for the complex dependence and non-ignorable dropout.

The results of our analyses do not provide evidence of a negative effect of living alone on episodic memory. We obtain negative, although non-significant (the 95% CI covers zero), effects

on the rate of change in memory scores. Thus, our findings cannot confirm nor challenge results from previous studies, where an association between living arrangement and cognitive decline was noted (van Gelder *et al*., 2006; Mousavi-Nasab *et al*., 2012). Stronger negative effects were found in Mousavi-Nasab *et al*. (2012), where the focus was on married *versus* non-married subjects; there is evidence in the literature (e.g. van Gelder *et al*. (2006)) that spouses provide stronger social support and cognitive stimulation than other forms of cohabitation. The inconclusive results of our study may be due to a lack of power, such that a larger cohort and/or more informative priors may be needed to resolve this question better.

We have compared individuals with different living arrangement at baseline, as opposed to allowing the exposure to vary over time (Zajonc (2012), and references therein), implying that those living with someone may lose their partner during follow-up, thereby yielding underestimation (in absolute value) if the actual effect of living alone is negative.

7. Software

See `www.betula.su.se` for information regarding access to the Betula data. BUGS code is downloadable from

```
http://wileyonlinelibrary.com/journal/rss-datasets
```

Acknowledgements

We thank Anders Lundquist, two referees, the Associate Editor and the Joint Editor for helpful comments that have improved the paper. We are also grateful to the staff of the Betula study. This work was funded by the Swedish Research Council as well as by grants from the Knut and Alice Wallenberg Foundation. Daniels was partially supported by US National Institutes of Health grants CA85295 and CA183854.

References

Bäckman, L., Small, B. J. and Fratiglioni, L. (2001) Stability of the preclinical episodic memory deficit in Alzheimer's disease. *Brain*, **124**, 96–102.

Barnard, J., Frangakis, C. E., Hill, J. L. and Rubin, D. B. (2003) Principal stratification approach to broken randomized experiments: a case study of school choice vouchers in New York City. *J. Am. Statist. Ass.*, **98**, 299–311.

Chiba, Y. and VanderWeele, T. J. (2011) A simple method for principal strata effects when the outcome has been truncated due to death. *Am. J. Epidem.*, **173**, 745–751.

Daniels, M. J. (1999) A prior for the variance in hierarchical models. *Can. J. Statist.*, **27**, 567–578.

Daniels, M. J., Chatterjee, A. S. and Wang, C. (2012) Bayesian model selection for incomplete data using the posterior predictive distribution. *Biometrics*, **68**, 1055–1063.

Daniels, M. J. and Hogan, J. W. (2008) *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. Boca Raton: Chapman and Hall–CRC.

Daniels, M. J. and Zhao, Y. D. (2003) Modelling the random effects covariance matrix in longitudinal data. *Statist. Med.*, **22**, 1631–1647.

Frangakis, C. E. and Rubin, D. B. (2002) Principal stratification in causal inference. *Biometrics*, **58**, 21–29.

Frangakis, C. E., Rubin, D. B., An, M.-W. and MacKenzie, E. (2007) Principal stratification designs to estimate input data missing due to death. *Biometrics*, **63**, 641–649.

Fratiglioni, L., Wang, H.-X., Ericsson, K., Maytan, M. and Winblad, B. (2000) Influence of social network on occurrence of dementia: a community-based longitudinal study. *Lancet*, **355**, 1315–1319.

Frumento, P., Mealli, F. and Rubin, D. B. (2012) Evaluating the effect of training on wages in the presence of noncompliance, nonemployment, and missing outcome data. *J. Am. Statist. Ass.*, **107**, 450–466.

Gallop, R., Small, D., Lin, J. Y., Elliot, M. R., Joffe, M. and Ten Have, T. R. (2009) Mediation analysis with principal stratification. *Statist. Med.*, **28**, 1108–1130.

van Gelder, B. M., Tijhuis, M., Kalmijn, S., Giampaoli, S., Nissinen, A. and Kromhout, D. (2006) Marital status and living situation during a 5-year period are associated with a subsequent 10-year cognitive decline in older men: the FINE Study. *J. Gerontol.* B, **61**, P213–P219.

Gelman, A. (2006) Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Baysn Anal.*, **1**, 515–534.
Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A. and Rubin, D. (2013) *Bayesian Data Analysis*. Boca Raton: Chapman and Hall–CRC.
Gelman, A. and Hill, J. (2007) *Data Analysis using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.
Gelman, A., Meng, X.-L. and Stern, H. (1996) Posterior predictive assessment of model fitness via realized discrepancies. *Statist. Sin.*, **6**, 733–807.
Harel, O. and Schaffer, J. (2009) Partial and latent ignorability in missing-data problems. *Biometrika*, **96**, 37–50.
Hogan, J. W. and Laird, N. M. (1997) Model-based approaches to analysing incomplete longitudinal and failure time data. *Statist. Med.*, **16**, 259–272.
Imbens, G. W. and Rubin, D. B. (1997) Bayesian inference for causal effects in randomized experiments with noncompliance. *Ann. Statist.*, **25**, 305–327.
Josefsson, M., de Luna, X., Pudas, S., Nilsson, L.-G. and Nyberg, L. (2012) Genetic and lifestyle predictors of 15-year longitudinal change in episodic memory. *J. Am. Geriatr. Soc.*, **60**, 2308–2312.
Lee, K., Daniels, M. J. and Sargent, D. J. (2010) Causal effects of treatments for informative missing data due to progression/death. *J. Am. Statist. Ass.*, **105**, 912–929.
Little, R. J. (1994) A class of pattern-mixture models for normal incomplete data. *Biometrika*, **81**, 471–483.
Little, R. J. and Rubin, D. B. (2002) *Statistical Analysis with Missing Data*, 2nd edn. New York: Wiley.
Mattei, A. and Mealli, F. (2007) Application of the principal stratification approach to the Faenza randomized experiment on breast self-examination. *Biometrics*, **63**, 437–446.
Morris, M. C., Evans, D. A., Hebert, L. E. and Bienias, J. L. (1999) Methodological issues in the study of cognitive decline. *Am. J. Epidem.*, **149**, 789–793.
Mousavi-Nasab, S.-M.-H., Kormi-Nouri, R., Sundström, A. and Nilsson, L.-G. (2012) The effects of marital status on episodic and semantic memory in healthy middle-aged and old individuals. *Scand. J. Psychol.*, **53**, 1–8.
Nilsson, L.-G., Adolfsson, R., Bäckman, L., de Frias, C. M., Molander, B. and Nyberg, L. (2004) Betula: a prospective cohort study on memory, health and aging. *Agng Neurpsychol. Cogn*, **11**, 134–148.
Nilsson, L.-G., Bäckman, L., Erngrund, K., Nyberg, L., Adolfsson, R., Bucht, G., Karlsson, S., Widing, M. and Winblad, B. (1997) The Betula prospective cohort study: memory, health, and aging. *Agng Neurpsychol. Cogn*, **4**, 1–32.
O'Muircheartaigh, C. and Hedges, L. V. (2014) Generalizing from unrepresentative experiments: a stratified propensity score approach. *Appl. Statist.*, **63**, 195–210.
Palmer, K., Bäckman, L., Winblad, B. and Fratiglioni, L. (2008) Mild cognitive impairment in the general population: occurrence and progression to Alzheimer disease. *Am. J. Geriatr. Psychol.*, **16**, 603–611.
Rönnlund, M., Nyberg, L., Bäckman, L. and Nilsson, L.-G. (2005) Stability, growth, and decline in adult life span development of declarative memory: cross-sectional and longitudinal data from a population-based study. *Psychol. Agng*, **20**, 3–18.
Rosenbaum, P. and Rubin, D. (1983a) The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70**, 41–55.
Rosenbaum, P. R. and Rubin, D. B. (1983b) Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *J. R. Statist. Soc.* B, **45**, 212–218.
Roy, J., Hogan, J. W. and Marcus, B. H. (2008) Principal stratification with predictors of compliance for randomized trials with 2 active treatments. *Biostatistics*, **9**, 277–289.
Rubin, D. B. (1991) Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism. *Biometrics*, **47**, 1213–1234.
Rubin, D. B. (2005) Causal inference using potential outcomes: design, modeling, decisions. *J. Am. Statist. Ass.*, **100**, 322–331.
Schaie, K. W. (1994) The course of adult intellectual development. *Am. Psychol.*, **49**, 304–313.
Sekhon, J. (2011) Multivariate and propensity score matching software with automated balance optimization: the matching package for R. *J. Statist. Softwr.*, **42**, 1–52.
Sturtz, S., Ligges, U. and Gelman, A. (2005) R2WinBUGS: a package for running WinBUGS from R. *J. Statist. Softwr.*, **12**, 1–16.
Tsiatis, A. (2006) *Semiparametric Theory and Missing Data*. Berlin: Springer.
Waernbaum, I. (2012) Model misspecification and robustness in causal inference: comparing matching with doubly robust estimation. *Statist. Med.*, **31**, 1572–1581.
Zajonc, T. (2012) Bayesian inference for dynamic treatment regimes: mobility, equity, and efficiency in student tracking. *J. Am. Statist. Ass.*, **107**, 80–92.
Zigler, C. M. (2013) The central role of Bayes theorem for joint estimation of causal effects and propensity scores. *Preprint arXiv:1308.5595.* Harvard School of Public Health, Boston.

Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

'Supplementary material for the paper "Causal inference with longitudinal outcomes and non-ignorable drop-out: Estimating the effect of living alone on cognitive decline"'.