Published in final edited form as:

J R Stat Soc Ser C Appl Stat. 2016 January 1; 65(1): 131–144. doi:10.1111/rssc.12110.

Causal inference with longitudinal outcomes and non-ignorable drop-out: Estimating the effect of living alone on cognitive decline

Maria Josefsson,

Department of Statistics, Umeå School of Business and Economics, Umeå University, Umeå, Sweden.

Xavier de Luna,

Department of Statistics, Umeå School of Business and Economics, Umeå University, Umeå, Sweden.

Michael J. Daniels, and

Department of Integrative Biology, Department of Statistics & Data Sciences, University of Texas at Austin.

Lars Nyberg

Department of Integrative Medical Biology & Radiation Sciences, Umeå University Umeå, Sweden.

Summary

In this paper 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 five years. One key feature of the model is the combination of propensity score matching to balance confounding variables between the two living arrangement groups –in order to reduce bias due to unbalanced covariates at baseline, with a pattern mixture model for longitudinal data –in order to deal with non-ignorable drop-out. 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 here adds to previous studies in the literature concerning the protective effect of living with someone, by proposing a modeling approach treating living arrangement as an exposure.

Keywords

Ageing; Bayesian inference; episodic memory; nonignorable missingness; pattern mixture model; propensity score matching; sensitivity analysis

Correspondence to: Maria Josefsson.

Conflict of Interest: None declared.

1. Introduction

Ageing is accompanied by an average decline in mental abilities such as processing speed, attention, and episodic memory (Schaie et al. 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 modeling framework to estimate the causal effect of living arrangement (living alone versus living with someone) on cognitive functioning using data from a large prospective cohort study, the Betula study (Nilsson et al. 1997). The cognitive outcome we use is an episodic memory score measured every five 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, van Gelder et al. 2006, Mousavi-Nasab et al. 2012, Josefsson et al. 2012, the latter also based on the Betula study). The protective effect of being married or living with some-one in relation to cognition may be explained by the fact that a partner provides, in addition to practical support, emotional support, and intellectual stimulation. While it seems reasonable to infer that loss of social/emotional support could contribute to accelerated decline, it cannot be ruled out that relations split if people start to suffer from cognitive dysfunction (perhaps ultimately affecting behavior). In any case, studying causal effects on cognitive change brings in two complications. First, in order 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 drop-out/attrition, i.e., subjects are lost to follow-up and their responses are not observed thereafter. In studies of cognitive decline in elderly populations, drop-out 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 drop-out, conventional statistical methods may be invalid and lead to severely biased estimates (Little & 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. The latter 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 & Rubin 1997, Frangakis & Rubin 2002, Frangakis et al. 2007, Chiba & VanderWeele 2011). In the earlier mentioned literature on the association between living arrangement and cognitive decline, only Josefsson et al. (2012) analysis takes into account the possibility that drop-out may be related to changes in cognitive ability. While in the latter study, living arrangement was shown to be associated to cognitive performance, we develop herein a modeling 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 & Rubin 1983*b*), to control for confounding variables, with a pattern mixture model for longitudinal data in order to deal with non-ignorable drop-out. 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 vs. living with someone) and thus reduces bias due to observed confounders (Rosenbaum & Rubin 1983*b*). A matched dataset 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.

Pattern mixture models (PMM) have been developed for the analysis of longitudinal data subject to non-ignorable drop-out (Little 1994, Hogan & Laird 1997). This approach stratifies the population by patterns of missing data (e.g. by the time of drop-out) 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 drop-out 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 we match upon, and the linear extrapolation of cognitive decline beyond drop out 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 ongoing population-based study of 4445 adults aged 35–85 years at enrollment in Sweden. The objective is to study how memory functions change over time and identify risk factors for dementia. Data have so far been collected every 5th 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 35–85 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 5th 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 enrollment. Memory was assessed 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/ceiling effects, i.e. lower/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 at baseline or older (old aged), and a cohort of 990 subjects 35–65 years of age (middle aged). The age of 65 is used as cutoff 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 cutoff also differ in that the death rate during follow up is much lower in the younger group; see Section 3 below for the methodological issues 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 enrollment.

Baseline demographic and lifestyle characteristics included age, sex, and self-reported highest level of education, (categorized into *high*: Secondary school or higher (>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 Y/N) 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 drop-outs 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 last three 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/unspecified reasons. In the event of death after a participant drops out, information regarding the participant's date of death is available.

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 formally state under which assumptions such a parameter is identified.

Let A_i denote the binary exposure, where here $A_i = 1$ if the i th individual ($i = 1, \dots, n$) is living alone at baseline and $A_i = 0$ if s/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 (log of 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 using random effects to allow for individual slopes (rate-of-change in memory score). For $i = 1, \dots, n$,

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

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

where $B_{0i}(0)$ and $B_{0i}(1)$ are random intercepts, $B_{1i}(0)$ and $B_{1i}(1)$ random slopes and $\varepsilon_{it}(0)$ and $\varepsilon_{it}(1)$ mean zero independent random errors. Let us also define two potential survival indicators for each individual: $S_i(0)$ which takes value one if individual i would survive under the follow up period if living with someone at baseline and zero otherwise, and $S_i(1)$ which takes value one if individual i would survive under the follow up period if living alone at baseline and zero otherwise. We assume throughout that the units do not interfere with each other, that is 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 & Mealli 2007, Roy et al. 2008, Gallop et al. 2009, Chiba & VanderWeele 2011, Frumento et al. 2012). We can now define a main parameter of interest for those surviving irrespective of exposure –a principal stratum causal effect (PSCE): 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 to living with someone) for those that were actually living alone at baseline? This parameter is not in general identified unless some conditions hold. There are two issues that need to be considered, first exposure is not randomised, and second that we do not observe the always survivors stratum.

Let us therefore make the following assumptions.

Assumption A

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

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

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

where $0 \leq \lambda \leq 1$.

Assumption C 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 A is called strong ignorability and holds when $\mathbf{W}_i$ contains all pre-exposure covariates related to both exposure and potential outcomes and survival. Assumption B was used in similar contexts in Roy et al. (2008, see Sec. 4.2 for a derivation) and Lee et al. (2010). Assumption B 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 commonly used in the principal stratification literature). In Assumption B we can vary λ between 0 and 1, thereby allowing for situations lying somewhere in-between these two extreme cases. Finally, Assumption C 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 based on 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

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

the first equality by Assumption A 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 Assumption B and C, the latter expression is identified up to λ and Δ_a , $a = 0, 1$. For instance, if we assume $\Delta_0 = \Delta_1 = 0$, $E[B_{1i}(a)|A_i=a, S_i=1, \mathbf{W}_i]$ 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 B; 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 due to a result by Rosenbaum & Rubin (1983b) whereby Assumption A implies

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

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 re-weighting tool instead of as an imputation device (Tsiatis 2006, O’Muircheartaigh & Hedges 2014). Matching has, however, been shown to have better finite sample properties (Waernbaum 2012).

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

$\zeta_2 = \log \frac{\Pr(A_i=1|V_i=1, \mathbf{W}_i)/\Pr(A_i=0|V_i=1, \mathbf{W}_i)}{\Pr(A_i=1|V_i=0, \mathbf{W}_i)/\Pr(A_i=0|V_i=0, \mathbf{W}_i)}$. Then Assumption A does not hold for $B_{1i}(0)$ if $\zeta_1 \zeta_2 \neq 0$. In such a case the bias implied by mistakenly assuming $\zeta_1 \zeta_2 = 0$ (Assumption A) 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 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 zero. When $\zeta_2 = 0$, V_i and A_i are then independent conditionally on $\mathbf{W}_i$, so (2) is equal to zero.

Thus, sensitivity to Assumptions B and C can be investigated by varying λ_a , $a = 0, 1$, while sensitivity to the existence of unobserved confounder (Assumption A) can be investigated by varying $\zeta_1 \zeta_2$ in (2).

4. Models and Methods

With the goal to estimate the effect of the exposure “living alone” on the ability to maintain cognitive function, we use a hierarchical modeling approach and Bayesian inference which allows us to take into account the complex dependence structure of the data and appropriately propagate uncertainty. 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 using random effects.

4.1. A propensity score model for matching

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

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

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 & Rubin 1983b). Thus, each exposed (i.e., living alone) subject is matched with a control (i.e., not living alone) subject that is closest in terms of the propensity score; we re-use 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. A linear mixed model for longitudinal matched data

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

$$\mathbf{Y}_{ij} = \mathbf{X}_{ij} \boldsymbol{\beta} + \mathbf{Z}_{ij} \mathbf{u}_i + \mathbf{Z}_{ij} \mathbf{b}_{ij} + \boldsymbol{\varepsilon}_{ij}, \quad (4)$$

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

4.3. The pattern mixture model for non-ignorable missing data

To account for non-ignorable drop-out, we specify a pattern mixture model (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., intercept and slope) which are a weighted average of the pattern-specific parameters across the three drop-out patterns. The model is based on the following 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 ($t = 1, 2, \dots, T, \beta_1$), A_{ij} (binary exposure, β_2), $A_{ij} \times \text{time}$ (β_3), M_{ij} (drop-out, factor with 3 levels, β_4, β_5), $M_{ij} \times \text{time}$ (β_6, β_7), $M_{ij} \times A_{ij}$ (β_8, β_9), and $M_{ij} \times \text{time} \times A_{ij}$ (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 $\mathbf{u}_i$ is independent of drop-out pattern, since subjects within one matched group may belong to different drop-out 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

$$\sum_{m=2}^4 \pi_{m,a} = 1, a=0, 1.$$

For the pattern mixture model, the fixed effects parameters for the population (unconditional on drop-out) are a weighted average of the drop-out specific patterns, i.e. intercept:

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

time fixed effect:

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

living alone (A_{ij}) fixed effect:

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

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

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

Section 4.5 below 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 & Schaffer 2009); i.e., conditional 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 appropriately account 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 conditional 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 appropriately account for uncertainty in the matching (and 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, $\eta(\gamma)$, that identifies pairs based on 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 , $k = 1, \dots, n_1$, is $p(\eta|\gamma, \mathbf{W}) = I\{\eta \equiv g(\gamma, \mathbf{W})\}$, where the function g determines the matching based on the covariates and the value of the parameter γ . For each draw of γ , we create the matched pairs, η , 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 that 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 defined in Sections 4.2 and 4.3 given the matched pairs for each draw of (γ, η) , $p(\beta, \mathbf{u}, \mathbf{b}, \Sigma, \Omega, \sigma^2, \pi | \mathbf{A}, \mathbf{X}, \mathbf{Z}, \mathbf{Y}, \eta, \gamma)$, using Markov chain Monte Carlo (MCMC) methods. Details on the priors follow.

For the fixed effects β , 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, $\mathbf{u}$, and subject, $\mathbf{b}$, into a set of unconstrained dependence parameters and a set of variance parameters, following the approach described in Daniels & 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 $u_{i1} \sim N(0, \sigma_{u,1}^2)$, and the random slope from group, conditional on the the random intercept is $u_{i2}|u_{i1} \sim N(\phi u_{i1}, \sigma_{u,2}^2)$. Exploration of the data suggested allowing the covariance matrix for a subject to differ by missing data pattern (which motivated using this decomposition). As such, we allowed the dependence parameter, ϕ , and variance parameters, $\sigma_{b,l}^2$, $l = 1, 2$, to vary as a function of drop-out pattern M , $b_{ij1}|M_{ij} = m \sim N(0, \sigma_{b,1m}^2)$, and $b_{ij2}|b_{ij1}, M_{ij} = m \sim N(\phi_m b_{ij1}, \sigma_{b,2m}^2)$, 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 drop-out specific group parameters to common values across patterns, $\phi_\star$ and $\delta_{b,l}$. In particular, we specify a normal prior for the dependence parameter $\phi_m \sim N(\phi_\star, \sigma_\phi)$ and gamma priors for the precision parameters, $1/\sigma_{b,lm}^2 \sim \text{Gamma}(\delta_{b,l}/\sigma_{b,l}^2, \delta_{b,l})$. The 'location' parameters of these priors are given a diffuse normal prior for $\phi_\star$ and a diffuse half-normal for $1/\sigma_{b,l}^2$, respectively. For the parameters that control shrinkage, σ_ϕ , $\delta_{b,1}$ and $\delta_{b,2}$, we specify Uniform shrinkage priors (Daniels 1999); see the 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, e.g., Gelman et al., 2006); as such, we use the shrinkage priors and half normal for these variances. Also, a sensitivity analysis showed that our results are robust to

these choices (see the Supplementary Material). Finally, we specify Dirichlet priors for $(\pi_{2,A}, \pi_{3,A}, \pi_{4,A})$, $A = 0, 1$, with hyperparameter vector $(1/3, 1/3, 1/3)$; the resulting posteriors are Dirichlet distributions.

For each sample from $p(\boldsymbol{\beta}, \mathbf{u}, \mathbf{b}, \Sigma, \Omega, \sigma^2, \pi|A, \mathbf{X}, \mathbf{Z}, \mathbf{Y}, \boldsymbol{\eta}, \boldsymbol{\gamma})$, the marginalized parameters in the PMM are computed as the weighted averages of the pattern-specific parameters as specified in (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 burnin 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 , a function of 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 (1), and $\alpha_1 + \alpha_3 + u_{2i}$ to $E(B_{1i}(1) | A_i = 1, S_i = 1, \mathbf{W}_i)$ in (1). Thus, if we assume that $\tau_0 = \tau_1 = 0$ 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 A, and $\psi_1(\mathbf{W}_i)$ in Assumption B using logistic regressions. Furthermore, averaging the argument of the expectation above over a matched sample corresponds to averaging over $p(\mathbf{W}_i | A_i=1, S_i=1)$. In order to target (9) and 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 (nonparametrically) 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 (10) is then obtained by noting

$$\text{that } p(\mathbf{W}_i | A_i=1, S_i=1) = \frac{p(S_i=1, \mathbf{W}_i | A_i=1)}{p(S_i=1 | A_i=1)}, \text{ which is equal to } \frac{\psi_1(\mathbf{W}_i)p(\mathbf{W}_i | A_i=1)}{p(S_i=1 | A_i=1)}.$$

Further, note that the numerator in (10) is equal to $\frac{p(S_i(1)=S_i(0)=1, \mathbf{W}_i | A_i=1)}{p(S_i(1)=S_i(0)=1 | A_i=1)}$, where $p(S_i(1)=S_i(0)=1, \mathbf{W}_i | A_i=1)$ is identified up to λ by Assumption B 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 described in Section 4.1, we included the following covariates: age, education, gender, feeling well, and physical activity. We specified two separate models: one for old subjects, ages 65 and older, and one for middle aged subjects, ages 35–60. To allow for non-linearity in the effect of age, the latter was categorized into the sampled age cohorts: 35, 40 ..., 60 for the middle aged subjects, and 65, 70 ..., 85 for the old group. Details on the fitted propensity score model can be found in the 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 described in Section 4.3 are found in Table 2, for both old and middle aged groups. Analyses are based on matched pairs where both the exposed and 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 (drop-outs after year 5) had lower episodic memory score at baseline compared to subjects who remained in the study (95% CI just covers zero), and had greater decline in episodic memory score over time (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 drop out indicators are negative, although with 95% CI including zero, indicating a weaker association of drop-out with episodic memory.

Posterior medians and 95% CI for the pattern-specific regression parameters marginalized over drop-out patterns, see (5–8), are given in Table 3 and the corresponding trajectories are plotted in Figure 1. For the old group, subjects living with someone had a significant decrease in memory score of 13% (original memory score scale) over the study period, 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 to subjects living with someone. The two latter 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 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), described in 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 Supplementary Material sensitivity to the omission of unobserved confounders (Assumption A), alternatives to a monotonicity assumption (Assumption B), and different nonignorable mechanisms, e.g. via changes in the slope (cognitive decline) after drop-out (Daniels & Hogan 2008). Our conclusions did not appear to be sensitive to these assumptions. The Supplementary Material also provides analyses using simpler modelling approaches (using missing 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 (5–8), although credible intervals 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 proposed approach 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 drop-out.

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 (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, and 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 provides 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 better resolve this question.

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 those living with someone may lose their partner during follow-up thereby yielding under estimation (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 the journal web site.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Anders Lundquist, two anonymous 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 grants from the Knut and Alice Wallenberg Foundation. Daniels was partially supported by U.S. NIH grants CA85295 and CA183854.

References

- Bäckman L, Small BJ, Fratiglioni L. Stability of the preclinical episodic memory deficit in alzheimer's disease. *Brain*. 2001; 124(1):96–102. [PubMed: 11133790]
- Barnard J, Frangakis CE, Hill JL, Rubin DB. Principal stratification approach to broken randomized experiments: A case study of school choice vouchers in new york city. *Journal of the American Statistical Association*. 2003; 98:299–311.
- Chiba Y, VanderWeele TJ. A simple method for principal strata effects when the outcome has been truncated due to death. *American Journal of Epidemiology*. 2011; 173(7):745–751. [PubMed: 21354986]
- Daniels MJ. A prior for the variance in hierarchical models. *Canadian Journal of Statistics*. 1999; 27(3):567–578.
- Daniels MJ, Chatterjee AS, Wang C. Bayesian model selection for incomplete data using the posterior predictive distribution. *Biometrics*. 2012; 68(4):1055–1063. [PubMed: 22551040]
- Daniels MJ.; Hogan JW. Missing data in longitudinal studies: Strategies for Bayesian modeling and sensitivity analysis. Boca Raton: Chapman and Hall/CRC; 2008.
- Daniels MJ, Zhao YD. Modelling the random effects covariance matrix in longitudinal data. *Statistics in medicine*. 2003; 22(10):1631–1647. [PubMed: 12720301]
- Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics*. 2002; 58(1):21–29. [PubMed: 11890317]
- Frangakis CE, Rubin DB, An M-W, MacKenzie E. Principal stratification designs to estimate input data missing due to death. *Biometrics*. 2007; 63(3):641–649. [PubMed: 17824995]
- Fratiglioni L, Wang H-X, Ericsson K, Maytan M, Winblad B, et al. Influence of social network on occurrence of dementia: a community-based longitudinal study. *Lancet*. 2000; 355(9212):1315–1319. [PubMed: 10776744]
- Frumento P, Mealli F, Rubin DB. Evaluating the effect of training on wages in the presence of noncompliance, nonemployment, and missing outcome data. *Journal of the American Statistical Association*. 2012; 107:450–466.
- Gallop R, Small D, Lin JY, Elliot MR, Joffe M, Ten Have TR. Mediation analysis with principal stratification. *Statistics in Medicine*. 2009; 28:1108–1130. [PubMed: 19184975]
- Gelman, A.; Carlin, J.; Stern, H.; Dunson, D.; Vehtari, A.; Rubin, D. *Bayesian Data Analysis*. Chapman and Hall/CRC; 2013.
- Gelman, A.; Hill, J. *Data analysis using regression and multi-level/hierarchical models*. Cambridge: Cambridge University Press; 2007.
- Gelman A, Meng X-L, Stern H. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*. 1996; 6:733–807.
- Gelman A, et al. Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian analysis*. 2006; 1(3):515–534.
- Harel O, Schaffer J. Partial and latent ignorability in missing-data problems. *Biometrika*. 2009; 96(1): 37–50.
- Hogan JW, Laird NM. Model-based approaches to analysing incomplete longitudinal and failure time data. *Statistics in medicine*. 1997; 16(3):259–272. [PubMed: 9004396]
- Imbens GW, Rubin DB. Bayesian inference for causal effects in randomized experiments with noncompliance. *The Annals of Statistics*. 1997; 25:305–327.
- Josefsson M, de Luna X, Pudas S, Nilsson L-G, Nyberg L. Genetic and lifestyle predictors of 15-year longitudinal change in episodic memory. *Journal of the American Geriatrics Society*. 2012; 60(12):2308–2312. [PubMed: 23110764]
- Lee K, Daniels MJ, Sargent DJ. Causal effects of treatments for informative missing data due to progression/death. *Journal of the American Statistical Association*. 2010; 105(491):912–929. [PubMed: 21318119]
- Little RJ. A class of pattern-mixture models for normal incomplete data. *Biometrika*. 1994; 81(3):471–483.
- Little, RJ.; Rubin, DB. *Statistical analysis with missing data*. second edn. New York: Wiley; 2002.

- Mattei A, Mealli F. Application of the principal stratification approach to the faenza randomized experiment on breast self-examination. *Biometrics*. 2007; 63:437–446. [PubMed: 17688496]
- Morris MC, Evans DA, Hebert LE, Bienias JL. Methodological issues in the study of cognitive decline. *American journal of epidemiology*. 1999; 149(9):789–793. [PubMed: 10221314]
- Mousavi-Nasab S-M-H, Kormi-Nouri R, Sundström A, Nilsson L-G. The effects of marital status on episodic and semantic memory in healthy middleaged and old individuals. *Scandinavian Journal of Psychology*. 2012; 53(1):1–8. [PubMed: 22092006]
- Nilsson L-G, Adolfsson R, Bäckman L, de Frias CM, Molander B, Nyberg L. Betula: A prospective cohort study on memory, health and aging. *Aging Neuropsychology and Cognition*. 2004; 11(2–3):134–148.
- Nilsson L-G, Bäckman L, Erngrund K, Nyberg L, Adolfsson R, Bucht G, Karlsson S, Widing M, Winblad B. The Betula prospective cohort study: Memory, health, and aging. *Aging, Neuropsychology, and Cognition*. 1997; 4(1):1–32.
- O’Muircheartaigh C, Hedges LV. Generalizing from unrepresentative experiments: a stratified propensity score approach. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 2014; 63(2):195–210.
- Palmer K, Bäckman L, Winblad B, Fratiglioni L. Mild cognitive impairment in the general population: Occurrence and progression to alzheimer disease. *American Journal of Geriatric Psych*. 2008; 16(7):603–611.
- Rönnlund M, Nyberg L, Bäckman L, Nilsson L-G. Stability, growth, and decline in adult life span development of declarative memory: cross-sectional and longitudinal data from a population-based study. *Psychology and aging*. 2005; 20(1):3–18. [PubMed: 15769210]
- Rosenbaum P, Rubin D. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 1983a; 45(2):212–218.
- Rosenbaum P, Rubin D. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983b; 70:41–55.
- Roy J, Hogan JW, Marcus BH. Principal stratification with predictors of compliance for randomized trials with 2 active treatments. *Biostatistics*. 2008; 9(2):277–289. [PubMed: 17681993]
- Rubin DB. Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism. *Biometrics*. 1991; 47:1213–1234. [PubMed: 1786315]
- Rubin DB. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*. 2005; 100(469):322–331.
- Schaie KW, et al. The course of adult intellectual development. *American psychologist*. 1994; 49:304–304. [PubMed: 8203802]
- Sekhon J. Multivariate and propensity score matching software with automated balance optimization: the matching package for R. *Journal of Statistical Software*. 2011; 42(7):1–52.
- Sturtz S, Ligges U, Gelman A. R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software*. 2005; 12(3):1–16. URL: <http://www.jstatsoft.org>.
- Tsiatis, A. *Semiparametric Theory and Missing Data*. Berlin: Springer; 2006.
- van Gelder BM, Tijhuis M, Kalmijn S, Giampaoli S, Nissinen A, Kromhout D. 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. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*. 2006; 61(4):P213–P219.
- Waernbaum I. Model misspecification and robustness in causal inference: comparing matching with doubly robust estimation. *Statistics in Medicine*. 2012; 31(15):1572–1581. [PubMed: 22359267]
- Zajonc T. Bayesian inference for dynamic treatment regimes: Mobility, equity, and efficiency in student tracking. *Journal of the American Statistical Association*. 2012; 107(497):80–92.
- Zigler CM. The central role of Bayes theorem for joint estimation of causal effects and propensity scores. *arXiv preprint arXiv:1308.5595*. 2013

Figure 1. Posterior median episodic memory score over 15 years, and pointwise 95% credible intervals, obtained from the pattern mixture model presented in Section 4.3, stratified by living arrangement for the middle aged and old.

Table 1

Summary statistics for observed covariates at enrollment and attrition rate stratified by living arrangement for the middle aged and old groups.

	Old		Middle aged	
	Living with someone (N=359)	Living alone (N=203)	Living with someone (N=854)	Living alone (N=136)
Mean age (SD)	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
Drop-outs %	34.5	29.1	22.1	29.4
Survivors %	59.9	57.6	96.3	91.9

Note: There were 44 (23 in Old and 21 in Middle aged) individuals with intermittent missing observations.

Table 2

Posterior medians and 95% credible intervals for parameters of the mixed model presented in Sections 4.2 and 4.3 for the middle aged and old groups.

Covariate	Old		Middle aged	
	Median	95% credible interval	Median	95% credible interval
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 $\times$ 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 $\times$ Time (β_6)	-0.20	(-0.36, -0.03)	0.03	(-0.12, 0.17)
Dropouts after year 10 $\times$ Time (β_7)	-0.06	(-0.17, 0.05)	-0.02	(-0.10, 0.06)
Living alone $\times$ Dropouts after year 5 (β_8)	-0.03	(-0.35, 0.28)	0.03	(-0.20, 0.27)
Living alone $\times$ Dropouts after year 10 (β_9)	0.11	(-0.19, 0.40)	-0.01	(-0.24, 0.26)
Living alone $\times$ Dropouts after year 5 $\times$ Time (β_{10})	0.15	(-0.07, 0.37)	-0.10	(-0.28, 0.08)
Living alone $\times$ Dropouts after year 10 $\times$ 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)
Corr	-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)
Corr	0.63	(-0.01, 0.97)	0.35	(-0.21, 0.92)

Table 3

Posterior medians and 95% credible intervals for marginal parameters, and principal stratum causal effect (PSCE), for the middle aged and old groups. PSCE is given for $\lambda=1$.

Effect	Old		Middle aged	
	Median	95% credible interval	Median	95% credible interval
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 $\times$ Time (α_3)	-0.03	(-0.10, 0.05)	-0.01	(-0.05, 0.02)
PSCE (τ)	-0.02	(-0.12, 0.10)	-	