Psychological Medicine

cambridge.org/psm

Depression and memory function – evidence from cross-lagged panel models with unit fixed effects in ELSA and HRS

Christoph Jindra, Chenlu Li, Ruby S. M. Tsang, Sarah Bauermeister and John Gallacher

Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford, OX3 7JX, United Kingdom

Original Article

Cite this article: Jindra C, Li C, Tsang RSM, Bauermeister S, Gallacher J (2022). Depression and memory function – evidence from cross-lagged panel models with unit fixed effects in ELSA and HRS. *Psychological Medicine* **52**, 1428–1436. https://doi.org/10.1017/S0033291720003037

Received: 22 November 2019
Revised: 15 June 2020
Accepted: 6 August 2020
First published online: 11 September 2020

Key words:
Cognition; cross-lagged panel models; depression; ELSA; HRS; fixed effects

Author for correspondence:
Christoph Jindra,
E-mail: christoph.jindra@psych.ox.ac.uk

Abstract

Background. Individuals with depression are often found to perform worse on cognitive tests and to have an increased risk of dementia. The causes and the direction of these associations are however not well understood. We looked at two specific hypotheses, the aetiological risk factor hypothesis and the reverse causality hypothesis.

Method. We analysed observational data from two cohorts, English Longitudinal Study of Ageing (ELSA) and Health and Retirement Study (HRS), using cross-lagged panel models with unit fixed effects. Each model was run once with depression and repeated with cognition as the dependent variable and the other variable as the main explanatory variable. All models were estimated separately for contemporaneous effects and lagged effects up to 8 years in the past. We contrasted the results with models making the random effects assumption.

Results. Evidence from the fixed effects models is mixed. We find no evidence for the reverse causality hypothesis in ELSA and HRS. While there is no evidence for the aetiological risk factors hypothesis in ELSA, results from HRS indicate some effects.

Conclusion. Our findings suggest that current levels of cognitive function do not influence future levels of depression. Results in HRS provide some evidence that current levels of depressive symptoms influence future cognition.

Introduction

Depression is associated with poor cognitive function and increased risk of dementia (e.g. Airaksinen, Larsson, Lundberg, & Forsell, 2004; Gallagher, Kiss, Lanctot, & Herrmann, 2016; González, Bowen, & Fisher, 2008; Saczynski *et al.* 2010). At least one in four older adults is found to experience a gradual decline in the affective state with increasing age, which along with age-related cognitive decline, suggests a comorbid relationship of unclear temporality (Arve, Tilvis, Lehtonen, Valvanne, & Sairanen, 1999). Furthermore, evidence suggests depressive symptoms co-occur with domain-specific cognitive deficits at follow-up such as processing speed, executive function and episodic memory (Koenig, Bhalla, & Butters, 2014; Sheline et al., 2006). Moreover, higher baseline levels of depressive symptoms are predictive of cognitive function at follow-up with a decline in delayed recall and global cognition (Panza et al., 2009) and processing speed (Bunce, Batterham, Christensen, and MacKinnon, 2014). However, the underlying causes and direction of the relationship are still not well understood and different potential mechanisms have been proposed that could explain any association between depression and cognition (Bennett & Thomas, 2014). Overburdening of neurobiological resources could be a mechanism pointing towards depression as a cause of cognitive decline (Lupien, McEwen, Gunnar, & Heim, 2009). The two variables would also be related if changes in cognitive function lead to depression as a reaction to the awareness of the cognitive decline (Vinkers, Gussekloo, Stek, Westendorp, & Van Der Mast, 2004). On the other hand, depression might be a prodrome of dementia and thus, even if temporally preceding observed changes in cognition, could still be caused by the underlying neurocognitive disorder (Zahodne, Stern, & Manly, 2014). The present study focuses on the potential reciprocal relationship between depression and cognition and thus two possible hypotheses. The first hypothesis is the aetiological risk factor hypothesis, which postulates that depression is a cause of the cognitive decline. The second hypothesis, which we refer to as the reverse causality hypothesis, holds that depression may instead be a reaction to early cognitive deficits. However, the two hypotheses are not mutually exclusive, and it is plausible that they hold simultaneously, which would imply reciprocal causality. Though difficult to compare due to differences in statistical approaches, populations, and measures, previous studies, considering the possibility of reciprocal effects, have produced mixed results. Van Den Kommer et al. (2013) report a bi-directional relationship between depression and processing speed but only find effects from depression to cognition for general cognitive function. Additional

© The Author(s) 2020. Published by Cambridge University Press

CAMBRIDGE UNIVERSITY PRESS

studies find that depression negatively impacts cognition but not the other way around. Bunce et al. (2014) report that initial levels of depressive symptoms are associated with later measures of processing speed and reaction time. Results by Zahodne et al. (2014) suggest that depressive symptoms predict memory scores. Gale, Allerhand, and Deary (2012) provide some evidence that depressive symptoms might accelerate cognitive decline, though only in people aged between 60 and 80. On the other hand, several studies report effects from cognition to depression but not the other way around. Vinkers et al. (2004) find that attention and episodic memory predict depressive symptoms, while results by Perrino, Mason, Brown, Spokane, and Szapocznik (2008) suggest this type of relationship between general cognitive function and levels of depressive symptoms. Likewise, Jajodia and Borders (2011) report that memory predicts changes in depressive symptoms and Brailean et al. (2017) find that baseline performance in the delayed recall is predictive of increases in overall depressive symptoms. Dzierzewski et al. (2015), contrarily, report no relationship between global cognitive function and depressive symptoms beyond any contemporaneous association.

We provide further evidence on the longitudinal relationship between depressive symptoms and memory function, in our case episodic memory. Specifically, we report results of an analysis of the direction of the association between depression and memory in two large population cohorts with comparable measures, ELSA and HRS. Furthermore, given that there is no strong previous evidence that allows to determine when we should expect any effect to unfold, we take advantage of a large number of waves and explore different lags. Finally, we make use of the unique advantages of these data by applying a type of dynamic panel model, the so-called cross-lagged panel models with unit fixed effects (Allison, Williams, & Moral-Benito, 2017; Nyberg, Peristera, Westerlund, Johansson, & Hanson, 2017). These are specifically designed to provide better insight into the direction of causality by controlling for past levels of the dependent variable as well as accounting for statistical issues in case of reciprocal effects. They further control for time-constant confounders with time-constant effects. Thus, these models provide researchers with a powerful tool to learn about the direction of an association between two variables over time.

Material and methods

Data

We use two longitudinal cohort studies, ELSA in England (Marmot et al., 2017; Steptoe, Breeze, Banks, & Nazroo, 2013) and HRS in the USA (Sonnega et al., 2014). ELSA participants were tested 7 times during a period of 12 years, thus every 2 years. In case of HRS, the first two waves were disregarded as memory performance was assessed differently in these waves and our baseline is wave 3. Participants in HRS were tested 10 times every 2 years during a period of 18 years. Both cohorts are nationally representative of their populations, people aged 50 years and older in ELSA and individuals over 50 years for HRS. We restrict our analysis to respondents present in the first wave that is used for the analysis of the respective survey and are 50 years or older at the time. We drop 69 respondents who self-report a diagnosis of Alzheimer's or dementia at wave 1 in ELSA. Applying the same criterion to HRS is not possible as similar information is only collected in later waves. Based on these criteria, 11453 participants are eligible at wave 1 in ELSA and 17429 at wave 3 in HRS. Ethical approval for ELSA was granted by the NHS Research Ethics Committees under the National Research and Ethics Service and consent has been obtained for all components of the study. The HRS study was approved by the University of Michigan's Institutional Review Board and participants give oral consent before each interview.

Measures

Like in Langa et al. (2009), our measure of cognition is the sum of two memory performance tests, immediate and delayed word recall (Ofstedal, Fisher, & Herzog, 2005; Steel, Huppert, McWilliams, & Melzer, 2003). Participants are presented with a list of 10 common words and asked to recall as many of them as possible immediately as well as after a short delay. The sum score ranges from 0 to 20. Respondents with missing values on any one of the two memory tests are assigned a missing value on the sum score. Previous research has shown that both measures load on to a single latent construct, episodic memory, with comparable factor loadings (McArdle, Fisher, & Kadlec, 2007), justifying the use of a sum score (Cronbach's α ranges from 0.82 to 0.86 across waves in ELSA and 0.85 to 0.87 in HRS).

Depressive symptoms are measured using an eight-item version of the Center for Epidemiologic Studies Depression Scale (CES-D), which has previously been shown to have good reliability and validity (Karim, Weisz, Bibi, & ur Rehman, 2015; Steffick, 2000). The CES-D score is the sum of six negative indicators plus two positive indicators, ranging from 0 to 8 (Cronbach's α ranges from 0.79 to 0.81 across waves in ELSA and 0.78 to 0.82 in HRS). Following Zivin et al. (2010), we assign people who have missing values on two or more of the eight items a missing value on the sum score.

Analytical strategy

We use cross-lagged panel models with unit fixed effects, a dynamic panel model, for our analysis (Allison et al., 2017; Moral-Benito, 2013; Williams, Allison, & Moral-Benito, 2018). Two features make these particularly suited for our research. The first feature is the unit fixed effects. These are different from what is sometimes referred to as fixed effects in random-effects models, where they usually describe parameters that are not allowed to vary across observations. The major difference between the approaches is that random effects models make the random effects assumption, which in order for it to hold, requires the independent variables to be uncorrelated with the unit-specific error term of the models (Allison, 2009; Halaby, 2004; Leszczensky & Wolbring, 2019). In contrast, fixed-effects models allow for correlations. There are different ways to estimate fixed-effects models, but the main advantage in each case is that it allows to control for time-invariant confounders with time-constant effects (Allison, 2009; Gunasekara, Richardson, Carter, & Blakely, 2014; Halaby, 2004). The second feature deals with problems created by reverse or reciprocal causality. Basic fixed and random effects models assume strict exogeneity, which implies that the error term is independent of all past, future and current values of the independent variable (Halaby, 2004). The assumption is however violated by default in the presence of feedback mechanisms [see Leszczensky and Wolbring (2019) for a detailed discussion]. Dynamic panel models, in contrast to static models, address reverse causality and its consequences with two changes to the standard setup. Firstly, they include

https://doi.org/10.1017/S0033291720003037 Published online by Cambridge University Press

lagged values of the dependent variable on the right-hand side of the equation to map the dynamics between the variables and account for the fact that the main independent variable might in fact be caused by the dependent variable. Secondly, they replace strict with the weaker sequential exogeneity assumption under which the values of the main independent variable can be correlated with past values of the error term, which becomes necessary if both variables are explained by each other. Together, these features make the models popular for examining the causal direction between two variables (Allison et al., 2017).

To investigate the direction of the relationship, we estimate all models twice, once with depressive symptoms and once with memory as the dependent variable (denoted y in the tables) and the other variable as the main independent variable. Given that we have no underlying theory as to when the independent variable should show an effect on the outcome, we estimate a series of models with different possible lags, ranging from contemporaneous values of the independent variable (x_t) to values eight years in the past (x_{t-4}). A recent study has cautioned that fixed effects models might be biased in the presence of misspecified lags (Vaisey & Miles, 2017). We follow one current recommendation and always include the contemporaneous effect alongside the lagged effect of interest in the dynamic models (Leszczensky & Wolbring, 2019; Nyberg et al., 2017). For each model, we use the lag of the dependent variable one wave further in the past than the largest lag of the main independent variable to ensure causal ordering. If, for example, cognition is the dependent variable and we are interested in the effect of depressive symptoms in the previous period, we use memory scores from two periods earlier as control. We include only a minimal set of controls. We control for period effects, by allowing intercepts to vary across waves and age.

There are different ways to estimate the models. We use a structural equations framework (Moral-Benito, 2013). This has two main advantages. Firstly, we can rely on the goodness of fit measures to assess model performance. Secondly, we can use full information maximum likelihood to partially account for missing data. Unless stated otherwise, all models are reported with robust standard errors. All analyses are done using Stata S.E. 15.1 and the dynamic models are estimated using the – xtdpdml – command (StataCorp, 2017; Williams et al., 2018).

We present descriptive statistics first. Subsequently, we discuss results for models in increasing complexity, starting with simple cross-sectional and pooled OLS regressions. We then continue with results for fixed and random effects models without the lagged dependent variable. Finally, we present results for the cross-lagged panel models with unit fixed effects and contrast those to models making the random effects assumption. Results are reported separately by cohort. Each regression table shows models once with cognition as the dependent variable and depressive symptoms as the independent variable and once the other way around. Results are provided separately for different lags of the main independent variable.

Results

The average age for eligible participants in ELSA at baseline is 65 (S.D. = 10.24) years. On average, they remember 9.4 (S.D. = 3.57) words and report 1.6 (S.D. = 1.99) depressive symptoms. Around 43% are still participating in wave 7. At baseline, we observe a high negative correlation of −0.45 between age and memory scores and a small positive correlation of 0.08 between age and CES-D scores. For the same individuals, memory scores and CES-D scores show a negative correlation of −0.18. The average age of eligible participants in HRS at wave 3 is 67 (S.D. = 10.59) years. At this wave, respondents report on average 1.35 (S.D. = 1.89) depressive symptoms and remember 9.86 (S.D. = 3.91) words. Around 36% of the respondents are still observed in the last wave. Correlations at baseline are similar to those from ELSA. We observe a high negative correlation between age and memory scores of −0.44 and a small positive correlation of 0.06 between age and CES-D. The correlation between memory scores and CES-D is −0.19. Looking at the development of the averages over time using available information from all individuals present at the respective waves, we find rather stable levels of CES-D scores for both cohorts (see online supplementary Table S1 in the Supplementary Materials). Results for the memory scores in ELSA are similar, but we observe a potential decrease in the average memory performance over time in HRS. Potential reasons for stable averages over time, despite increasing average age, are learning effects for the cognitive tests (Rodgers & Ofstedal, 2011) or that people with worse memory performance, or more depressive symptoms, are more likely to be lost to follow-up (Weir, Faul, & Langa, 2011).

Benchmark results for unadjusted OLS regressions are presented in online supplementary Table S2 in the Supplementary Material. All coefficients are negative, in line with the often reported finding that reporting a larger number of depressive symptoms is associated with worse performance on cognitive tests (Clark, Chamberlain, & Sahakian, 2009). To assess the impact of adjusting for time-invariant confounders, we estimate static fixed effects models (see upper panel of Table 1). We still find strong evidence against the null hypotheses of no effect for all contemporaneous variables. In case of cognition as the dependent variable, most of the lagged effects of CES-D are small in ELSA, with p-values ranging from 0.016 to 0.76. The picture is different in HRS. Here we still find strong evidence for the aetiological risk factor hypothesis for lags $t-1$ to $t-3$, thus suggesting that CES-D might have long-term effects on memory, even after considering time-constant unobserved heterogeneity. When looking at CES-D as the dependent variable, p-values for the lagged effects of cognition range from 0.18 to 0.77, thus providing no evidence for the reverse causality hypothesis. Given that fixed-effects estimates are less efficient if the random effects assumption holds, we estimate random intercept models for comparison (see lower panel of Table 1). Like for the OLS regressions, we find strong evidence against the null hypothesis in each case.

However, comparing results across the panels, we find that standard errors are similar.

As discussed, in the presence of reverse causality, cross-lagged panel models are more appropriate and results for these are presented in Table 2. Global goodness of fit measures indicates excellent fit for each model (see online supplementary Table S3 in the Supplementary Material). Looking at the models with cognition as the dependent variable (top panel of Table 2), we find strong evidence from selected models that people who perform better in previous periods tend to do better in subsequent ones. We also find relatively strong evidence for the contemporaneous effect of depression on cognition in most models. However, looking at the lagged values of depressive symptoms, support for the aetiological risk factor hypothesis is again mixed. In the case of ELSA, none of the p-values for any of the lags is small enough to allow us to conclude confidently that previous levels of

https://doi.org/10.1017/S0033291720003037 Published online by Cambridge University Press

Table 1. Unit fixed effects and random intercept models

	ELSA					HRS				
	β	s.e.	p-value	Obs.	Ind.	B	s.e.	p-value	Obs.	Ind.
Unit fixed effects models										
Y = cognition$_t$:										
CES − D$_t$	−0.062	0.010	0.0000	49514	11239	−0.065	0.007	0.0000	99716	16364
CES − D$_{t-1}$	0.003	0.011	0.7604	36924	9059	−0.025	0.007	0.0009	81053	14285
CES − D$_{t-2}$	−0.010	0.012	0.4332	28421	7801	−0.032	0.008	0.0000	67300	12839
CES − D$_{t-3}$	−0.035	0.014	0.0155	21383	6814	−0.034	0.009	0.0001	55299	11489
CES − D$_{t-4}$	−0.009	0.018	0.6276	15381	6173	−0.000	0.010	0.9745	44718	10481
Y = CES-D$_t$:										
Cog$_t$	−0.021	0.003	0.0000	49514	11239	−0.020	0.002	0.0000	99716	16364
Cog$_{t-1}$	−0.002	0.003	0.5586	36849	9052	−0.002	0.002	0.3842	81024	14284
Cog$_{t-2}$	0.002	0.004	0.5798	28330	7785	−0.002	0.002	0.3090	67245	12838
Cog$_{t-3}$	0.007	0.005	0.1806	21302	6798	−0.003	0.003	0.3054	55234	11483
Cog$_{t-4}$	−0.004	0.006	0.4767	15301	6150	−0.001	0.003	0.7675	44659	10473
Random intercept models										
Y = cognition$_t$:										
CES − D$_t$	−0.132	0.008	0.0000	49514	11239	−0.133	0.006	0.0000	99716	16364
CES − D$_{t-1}$	−0.084	0.010	0.0000	36924	9059	−0.107	0.007	0.0000	81053	14285
CES − D$_{t-2}$	−0.107	0.011	0.0000	28421	7801	−0.118	0.007	0.0000	67300	12839
CES − D$_{t-3}$	−0.132	0.012	0.0000	21383	6814	−0.120	0.008	0.0000	55299	11489
CES − D$_{t-4}$	−0.129	0.015	0.0000	15381	6173	−0.102	0.009	0.0000	44718	10481
Y = CES-D$_t$:										
Cog$_t$	−0.042	0.003	0.0000	49514	11239	−0.041	0.002	0.0000	99716	16364
Cog$_{t-1}$	−0.025	0.003	0.0000	36849	9052	−0.025	0.002	0.0000	81024	14284
Cog$_{t-2}$	−0.022	0.004	0.0000	28330	7785	−0.025	0.002	0.0000	67245	12838
Cog$_{t-3}$	−0.021	0.004	0.0000	21302	6798	−0.025	0.002	0.0000	55234	11483
Cog$_{t-4}$	−0.033	0.005	0.0000	15301	6150	−0.041	0.002	0.0000	99716	16364

Notes: Each row contains one model specification with one specific lag of the main independent variable, estimated separately for ELSA and HRS using the maximum information available. All models are adjusted for age and wave. Robust standard errors are reported and p-values are for two-sided tests.

depression influence current levels of memory performance. In case of HRS, we find relatively strong evidence for the lagged effects at $t-1$ to $t-3$, but not for the levels of depressive symptoms 8 years in the past.

Results for CES-D as the dependent variable are given in the bottom panel of Table 2. The small p-values for CES-D$_{t-1}$ in specification six, show strong evidence against the null hypothesis of no effect in both cohorts, suggesting that people with higher levels of depressive symptoms in the previous period tend to do worse in the current. Evidence for this effect at $t-2$ in specification seven is still very strong but generally weaker for the remaining lags. When inspecting the main independent variable, cognition, the models provide strong evidence that current cognition is associated with the number of depressive symptoms. With one exception in HRS, this does not apply to the lagged values of cognitive performance. Thus, again our results do not provide much support for the idea that low cognitive functioning is a long-term risk factor for depressive symptoms.

We estimated the dynamic panel models under the random effects assumption, allowing us to use likelihood ratio tests to assess which model is appropriate. Results for the models are presented in Table 3 and results for the tests in online supplementary Table S4 in the Supplementary Material. In each case, the random effects assumption is rejected. This is supported by the lack of fit of the random-effects models (see online supplementary Table S5 in the Supplementary Material). When comparing the results, we find the same picture as before. Coefficients from the random effects models indicate strong evidence against the null hypotheses for each model. However, our tests suggest that the random effects assumption does not hold, and the fixed effects models are preferred.

Discussion and conclusion

Using static and dynamic panel models with fixed effects and two population cohorts, we find no evidence for an effect of cognitive

https://doi.org/10.1017/S0033291720003037 Published online by Cambridge University Press

Table 2. Cross-lagged panel models with unit fixed effects

	ELSA				HRS			
	β	s.e.	*p*-value	*N*	β	s.e.	*p*-value	*N*
				$y = \text{cognition}_t$				
Specification 1:								
Cog_{t-1}	0.131	0.009	0.0000	11313	0.124	0.006	0.0000	17009
$\text{CES} - \text{D}_t$	−0.074	0.016	0.0000		−0.046	0.010	0.0000	
Specification 2:								
Cog_{t-2}	0.019	0.008	0.0183	11313	0.088	0.005	0.0000	17009
$\text{CES} - \text{D}_t$	−0.068	0.019	0.0004		−0.060	0.011	0.0000	
$\text{CES} - \text{D}_{t-1}$	0.017	0.015	0.2538		−0.029	0.009	0.0017	
Specification 3:								
Cog_{t-3}	−0.015	0.008	0.0580	11311	−0.002	0.005	0.6899	17009
$\text{CES} - \text{D}_t$	−0.096	0.021	0.0000		−0.055	0.011	0.0000	
$\text{CES} - \text{D}_{t-2}$	−0.026	0.015	0.0878		−0.021	0.009	0.0204	
Specification 4:								
Cog_{t-4}	0.069	0.011	0.0000	11309	0.041	0.006	0.0000	17008
$\text{CES} - \text{D}_t$	−0.070	0.029	0.0160		−0.048	0.013	0.0002	
$\text{CES} - \text{D}_{t-3}$	−0.020	0.018	0.2676		−0.031	0.010	0.0018	
Specification 5:								
Cog_{t-5}	−0.015	0.015	0.3003	11298	−0.007	0.006	0.2569	17008
$\text{CES} - \text{D}_t$	−0.049	0.051	0.3393		−0.037	0.015	0.0169	
$\text{CES} - \text{D}_{t-4}$	−0.003	0.027	0.9048		0.005	0.011	0.6481	
				$y = \text{CES-D}_t$				
Specification 6:								
$\text{CES} - \text{D}_{t-1}$	0.141	0.010	0.0000	11308	0.174	0.007	0.0000	17008
Cog_t	−0.032	0.006	0.0000		−0.026	0.003	0.0000	
Specification 7:								
$\text{CES} - \text{D}_{t-2}$	0.021	0.009	0.0163	11308	0.052	0.006	0.0000	17008
Cog_t	−0.034	0.006	0.0000		−0.026	0.003	0.0000	
Cog_{t-1}	−0.007	0.005	0.1293		−0.005	0.003	0.0872	
Specification 8:								
$\text{CES} - \text{D}_{t-3}$	−0.008	0.010	0.4132	11306	−0.006	0.006	0.3748	17008
Cog_t	−0.032	0.007	0.0000		−0.026	0.003	0.0000	
Cog_{t-2}	−0.001	0.005	0.8487		−0.004	0.003	0.1474	
Specification 9:								
$\text{CES} - \text{D}_{t-4}$	−0.020	0.013	0.1072	11304	−0.013	0.007	0.0666	17007
Cog_t	−0.036	0.009	0.0001		−0.026	0.004	0.0000	
Cog_{t-3}	0.006	0.006	0.3455		−0.006	0.003	0.0426	
Specification 10:								
$\text{CES} - \text{D}_{t-5}$	0.004	0.016	0.7989	11293	−0.016	0.008	0.0396	17007
Cog_t	−0.041	0.016	0.0124		−0.023	0.004	0.0000	
Cog_{t-4}	−0.005	0.008	0.5241		−0.005	0.003	0.1609	

Notes: All models are estimated using full information maximum likelihood. Robust standard errors are reported and *p*-values are for two-tailed tests. All models are adjusted for wave and age.

https://doi.org/10.1017/S0033291720003037 Published online by Cambridge University Press

Table 3. Cross-lagged panel models with random effects

	ELSA				HRS			
	β	S.E.	*p*-value	*N*	β	S.E.	*p*-value	*N*
				$y = \text{cognition}_t$				
Specification 1:								
Cog_{t-1}	0.271	0.013	0.0000	11313	0.196	0.007	0.0000	17009
$\text{CES} - \text{D}_t$	−0.164	0.010	0.0000		−0.169	0.007	0.0000	
Specification 2:								
Cog_{t-2}	0.204	0.011	0.0000	11313	0.195	0.006	0.0000	17009
$\text{CES} - \text{D}_t$	−0.160	0.012	0.0000		−0.147	0.007	0.0000	
$\text{CES} - \text{D}_{t-1}$	−0.056	0.011	0.0000		−0.106	0.007	0.0000	
Specification 3:								
Cog_{t-3}	0.239	0.011	0.0000	11311	0.135	0.006	0.0000	17009
$\text{CES} - \text{D}_t$	−0.159	0.013	0.0000		−0.151	0.008	0.0000	
$\text{CES} - \text{D}_{t-2}$	−0.085	0.013	0.0000		−0.098	0.008	0.0000	
Specification 4:								
Cog_{t-4}	0.347	0.012	0.0000	11309	0.202	0.006	0.0000	17008
$\text{CES} - \text{D}_t$	−0.135	0.016	0.0000		−0.145	0.009	0.0000	
$\text{CES} - \text{D}_{t-3}$	−0.084	0.015	0.0000		−0.103	0.009	0.0000	
Specification 5:								
Cog_{t-5}	0.332	0.014	0.0000	11298	0.211	0.007	0.0000	17008
$\text{CES} - \text{D}_t$	−0.145	0.022	0.0000		−0.153	0.010	0.0000	
$\text{CES} - \text{D}_{t-4}$	−0.101	0.020	0.0000		−0.081	0.010	0.0000	
				$y = \text{CES-D}_t$				
Specification 6:								
$\text{CES} - \text{D}_{t-1}$	0.303	0.017	0.0000	11308	0.246	0.008	0.0000	17008
Cog_t	−0.052	0.003	0.0000		−0.055	0.002	0.0000	
Specification 7:								
$\text{CES} - \text{D}_{t-2}$	0.215	0.014	0.0000	11308	0.149	0.007	0.0000	17008
Cog_t	−0.048	0.004	0.0000		−0.053	0.002	0.0000	
Cog_{t-1}	−0.021	0.004	0.0000		−0.027	0.002	0.0000	
Specification 8:								
$\text{CES} - \text{D}_{t-3}$	0.246	0.013	0.0000	11306	0.124	0.008	0.0000	17008
Cog_t	−0.046	0.004	0.0000		−0.053	0.003	0.0000	
Cog_{t-2}	−0.017	0.004	0.0001		−0.024	0.002	0.0000	
Specification 9:								
$\text{CES} - \text{D}_{t-4}$	0.277	0.013	0.0000	11304	0.160	0.008	0.0000	17007
Cog_t	−0.049	0.005	0.0000		−0.054	0.003	0.0000	
Cog_{t-3}	−0.013	0.005	0.0126		−0.026	0.003	0.0000	
Specification 10:								
$\text{CES} - \text{D}_{t-5}$	0.305	0.014	0.0000	11293	0.203	0.009	0.0000	17007
Cog_t	−0.045	0.006	0.0000		−0.052	0.003	0.0000	
Cog_{t-4}	−0.013	0.006	0.0397		−0.021	0.003	0.0000	

Notes: All models are estimated using full information maximum likelihood. Robust standard errors are reported and *p*-values are for two-tailed tests. All models are adjusted for wave and age.

https://doi.org/10.1017/S0033291720003037 Published online by Cambridge University Press

performance on CES-D beyond contemporaneous associations and hence no support for the reverse causality hypothesis. The absence of evidence for an effect in this direction indicates that current levels of memory function do not predict future levels of depressive symptoms in our analysis. In people with dementia, an explanation for this finding could be anosognosia, as it is sometimes argued that if people are not aware of their cognitive performance, there is no reason to expect that it would lead to depression. However, given that our population is still cognitively relatively healthy, it is unclear how much of the finding can be attributed to this. Results for cognition as the dependent variable are mixed. While we find some support for the aetiological risk factor hypothesis in HRS, the same conclusion is not supported by ELSA data. This is independent of whether we use dynamic or static panel models as long as we do not make the random effects assumption. Thus, while results for ELSA indicate that there might be no association between depressive symptoms and episodic memory beyond contemporaneous associations, results for HRS provide some support for an effect of depression on memory scores. Results in HRS are in line with two potential mechanisms. It could imply that depression has a detrimental effect on memory, for example via the overburdening of neurobiological resources, or that depression is a prodrome of cognitive decline. However, as our models control for previous levels of cognition and as we do not find evidence for an effect in the direction from cognition to depression, the latter seems less likely. A remaining question is why results differ between the cohorts. One factor is the sample size. On the other hand, there could be effect heterogeneity due to, e.g., different management of depression across populations. Well managed treatment of depression might sever the link between depression and cognition and differences in health care systems could partially explain differences between the cohorts. Future research is needed to better understand the causes of the differences. Given that our analysis shows that the effects sizes for unit fixed effects models are substantially smaller than those from the random effects counterparts, we interpret this as some support for the notion that there is a substantial overlap in risk factors for cognition and depression, e.g. genetics.

Some previous studies explicitly address the question of the direction of the relationship between depression and cognition. Our results for ELSA are similar to those by Dzierzewski et al. (2015) who find no relationship aside from contemporaneous associations. Using ELSA data, Gale et al. (2012) report an association between depressive symptoms at baseline and rates of change in general cognition. While this effect is only found for one age group and is in line with our findings for HRS, they differ from our results for ELSA. While there are differences in methodology and exact research question, we believe the main reason behind the difference is the use of unit fixed effects and thus the ability to control for time-invariant confounders. Results for HRS are in line with those by Bunce et al. (2014) and Zahodne et al. (2014). They are further like parts of Van Den Kommer et al. (2013), which find an effect from depression to general cognitive function in the Longitudinal Aging Study Amsterdam. On the other hand, our results contrast with several studies. Vinkers et al. (2004) find that baseline values for different cognitive measures predict depressive symptoms in a sample of Dutch participants. Analysing a sample of Hispanics, Perrino et al. (2008) report that general cognitive function is predictive of depressive symptoms the following year for each of their follow-ups. Analysing the same data as Van Den Kommer et al. (2013) using a cross-domain latent growth curve model, Brailean et al. (2017) conclude that baseline performance on the delayed recall task predicts the rate of change in depressed affect over time. Furthermore, they find that changes in processing speed are associated with changes in depressive symptoms. Finally, the most obvious difference is between our results and those from Jajodia and Borders (2011), who, using complex methods for longitudinal data, conclude that memory performance predicts change in depressive symptoms 2 years later using HRS data. One possible explanation for differences in the conclusions is the use of sampling weights in their analysis, which is not routinely implemented for our approach. Furthermore, the authors use delayed word recall only, while we use the sum of immediate and delayed word recall. However, re-analysing our data using delayed recall only does not change the results. On the other hand, the authors note that their model uses within and between variation and assumes that the two converge. Thus, in our opinion, the more likely reason for differences in the conclusion is that they are caused using a fixed-effects approach in our case which relies on within variation only. We assume that in most cases the use of a fixed-effects approach is the likely reason for differences between ours and previous studies. However, in some cases, differences might be caused by using a model that is designed for situations in which we expect reciprocal causality.

Our study has several strengths. We use data with directly comparable repeated measurements from two large population cohorts and cross-lagged panel models with unit fixed effects which are specifically designed for situations where we expect reciprocal causality. Doing so protects us against two threats to valid causal inference, unobserved time-constant heterogeneity and reverse causation, and allows us to base our conclusions on two independent data. As results from these complex models might be sensitive, we increase the confidence in our conclusions by additionally reporting simpler models. Having observations for seven waves in ELSA and ten in HRS allows us to investigate different possible lags and build a comprehensive picture of the time component of the relationship.

There are several limitations. Substantively, conclusions are limited towards the simple hypotheses from above. The postulated mechanisms in the literature are sometimes more complex. For example, our models do not consider whether the length of a depressive episode matters (Dotson, Resnick, & Zonderman, 2008; John et al., 2019). Additionally, the effect of depression might vary across cognitive domains (Zaninotto, Batty, Allerhand, & Deary, 2018). Given that some previous studies also report differences across depression subgroups (Airaksinen et al., 2004), one should look at measures of depression that are better suited to distinguish strength and type. Similarly, we are constrained to the short form of the CES-D scale and a more detailed version could produce a different picture. Finally, results for the aetiological risk factor hypothesis are not consistent, allowing only weak conclusions.

There are also statistical limitations. We observe relatively stable population averages over time for those remaining in the studies, suggesting bias due to attrition. Within-estimators should be less susceptible to this as they look at within-individual change and disregard between information. Additionally, we use FIML to account for missing data. However, FIML still assumes that data are missing at random (MAR). Due to the complexity of the models, we did not include auxiliary variables (Enders, 2010), beyond the time-varying controls in the sensitivity analysis (see below). Thus, the MAR assumption might be violated. FIML additionally requires the data to follow a multivariate normal

https://doi.org/10.1017/S0033291720003037 Published online by Cambridge University Press

distribution, which is violated for CES-D. However, Williams et al. (2018) argue that maximum likelihood is still consistent and we followed their advice to use robust standard errors.

As mentioned above, recent studies have discussed bias due to misspecifications of the dynamic processes in fixed-effects models (Leszczensky & Wolbring, 2019; Vaisey & Miles, 2017). Given that we are looking at the effect from different time periods and base our conclusion on the overall picture, this may be less of a concern in our case. For sensitivity analysis, we estimated the models by stepwise inclusion of the lags of the main variable (see online supplementary Tables S6 and S7 in the Supplementary Material). Results lead to the same conclusion but the threat from misspecification remains.

Similarly, in the presence of complex causal dynamics, including time-varying confounders, other statistical approaches, like the inverse probability of treatment weighting or matching, might be more appropriate (Gunasekara et al., 2014; Imai & Kim, 2019). However, these usually rely on a selection-on-observable assumption but often do not control for fixed effects and are consequently open to another line of criticism. The underlying question is what is deemed more important, bias due to unobserved time-constant confounding or dynamics. As we are particularly interested in controlling for time-invariant confounding, a fixed-effects approach seems more appropriate in our case, despite being unable to explore complex dynamics. Given that time-varying confounding will always be present, we estimated the models using a set of time-varying controls (see online supplementary Tables S8 and S9 in the Supplementary Material) and the conclusions remain the same.

In summary, using dynamic panel models with fixed effects in two populations, we fail to find evidence for the notion that cognitive function has any long-term impact on depression, while we find some evidence for the aetiological risk factor hypothesis in HRS but not in ELSA.

Supplementary material. For supplementary material accompanying this paper visit https://doi.org/10.1017/S0033291720003037.

Acknowledgements. The HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and conducted by the University of Michigan. The English Longitudinal Study of Ageing was developed by a team of researchers based at the University College London, NatCen Social Research, and the Institute for Fiscal Studies. The data were collected by NatCen Social Research. The funding is currently provided by the National Institute of Aging (R01AG017644), and a consortium of UK government departments coordinated by the National Institute for Health Research. ELSA data were analysed on Dementias Platform UK (Bauermeister et al., 2020). Dementia Platform UK funded this project through MRC grant ref MR/L023784/2.

Conflict of interest. None.

References

Airaksinen, E., Larsson, M., Lundberg, I., & Forsell, Y. (2004). Cognitive functions in depressive disorders: Evidence from a population-based study. *Psychological Medicine*, *34*(1), 83–91.

Allison, P. D. (2009). *Fixed effects regression models*. Los Angeles: SAGE.

Allison, P. D., Williams, R., & Moral-Benito, E. (2017). Maximum likelihood for cross-lagged panel models with fixed effects. *Socius: Sociological Research for a Dynamic World*, *3*, 1–17.

Arve, S., Tilvis, R. S., Lehtonen, A., Valvanne, J., & Sairanen, S. (1999). Coexistence of lowered mood and cognitive impairment of elderly people in five birth cohorts. *Aging Clinical and Experimental Research*, *11*(2), 90–95.

Bauermeister, S., Orton, C., Thompson, S., Barker, R. A., Bauermeister, J. R., BenShlomo, Y., … Gallacher, J. E. J. (2020). The dementias platform UK (DPUK) data portal. *European Journal of Epidemiology*, *35*, 601–611.

Bennett, S., & Thomas, A. J. (2014). Depression and dementia: Cause, consequence or coincidence? *Maturitas*, *79*(2), 184–190.

Brailean, A., Aartsen, M. J., Muniz-Terrera, G., Prince, M., Prina, A. M., Comijs, H. C., … Beekman, A. (2017). Longitudinal associations between late-life depression dimensions and cognitive functioning: A cross-domain latent growth curve analysis. *Psychological Medicine*, *47*(4), 690–702.

Bunce, D., Batterham, P. J., Christensen, H., & MacKinnon, A. J. (2014). Causal associations between depression symptoms and cognition in a community-based cohort of older adults. *American Journal of Geriatric Psychiatry*, *22*(12), 1583–1591.

Clark, L., Chamberlain, S. R., & Sahakian, B. J. (2009). Neurocognitive mechanisms in depression: Implications for treatment. *Annual Review of Neuroscience*, *32*(1), 57–74.

Dotson, V. M., Resnick, S. M., & Zonderman, A. B. (2008). Differential association of concurrent, baseline, and average depressive symptoms with cognitive decline in older adults. *American Journal of Geriatric Psychiatry*, *16*(4), 318–330.

Dzierzewski, J. M., Potter, G. G., Jones, R. N., Rostant, O. S., Ayotte, B., Yang, F. M., … Steffens, C, D. (2015). Cognitive functioning throughout the treatment history of clinical late-life depression. *International Journal of Geriatric Psychiatry*, *30*(10), 1076–1084.

Enders, C. K. (2010). *Applied missing data analysis*. New York and London: The Guildford Press.

Gale, C. R., Allerhand, M., & Deary, I. J. (2012). Is there a bidirectional relationship between depressive symptoms and cognitive ability in older people? A prospective study using the English Longitudinal Study of Ageing. *Psychological Medicine*, *42*(10), 2057–2069.

Gallagher, D., Kiss, A., Lanctot, K., & Herrmann, N. (2016). Depressive symptoms and cognitive decline: A longitudinal analysis of potentially modifiable risk factors in community dwelling older adults. *Journal of Affective Disorders*, *190*, 235–240.

González, H. M., Bowen, M. E., & Fisher, G. G. (2008). Memory decline and depressive symptoms in a nationally representative sample of older adults: The Health and Retirement Study (1998–2004). *Dementia and Geriatric Cognitive Disorders*, *25*(3), 266–271.

Gunasekara, F. I., Richardson, K., Carter, K., & Blakely, T. (2014). Fixed effects analysis of repeated measures data. *International Journal of Epidemiology*, *43*(1), 264–269.

Halaby, C. N. (2004). Panel models in sociological research: Theory into practice. *Annual Review of Sociology*, *30*(1), 507–544.

Imai, K., & Kim, I. S. (2019). When should we use linear fixed effects regression models for causal inference with panel data? *American Journal of Political Science*, *63*(2), 467–490.

Jajodia, A., & Borders, A. (2011). Memory predicts changes in depressive symptoms in older adults: A bidirectional longitudinal analysis. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *66 B*(5), 571–581.

John, A., James, S.-N., Patel, U., Rusted, J., Richards, M., & Gaysina, D. (2019). Longitudinal associations of affective symptoms with mid-life cognitive function: Evidence from a British birth cohort. *The British Journal of Psychiatry*, *215*(5), 675–682.

Karim, J., Weisz, R., Bibi, Z., & ur Rehman, S. (2015). Validation of the eight-item Center for Epidemiologic Studies Depression Scale (CES-D) among older adults. *Current Psychology*, *34*(4), 681–692.

Koenig, A. M., Bhalla, R. K., & Butters, M. A. (2014). Cognitive functioning and late-life depression. *Journal of the International Neuropsychological Society*, *20*(5), 461–467.

Kommer, V. D., Comijs, T. N., Aartsen, H. C., Huisman, M. J., Deeg, M., and Beekman, D. J., & T, A. (2013). Depression and cognition: How do they interrelate in old age? *American Journal of Geriatric Psychiatry*, *21*(4), 398–410.

Langa, K. M., Llewellyn, D. J., Lang, I. A., Weir, D. R., Wallace, R. B., Kabeto, M. U., & Huppert, F. A. (2009). Cognitive health among older adults in the United States and in England. *BMC Geriatrics*, *9*(1), 1–11.

Leszczensky, L., & Wolbring, T. (2019). How to deal with reverse causality using panel data? Recommendations for researchers based on a simulation

https://doi.org/10.1017/S0033291720003037 Published online by Cambridge University Press

study. *Sociological Methods and Research* (Online First), 1–29. https://doi.org/10.1177/0049124119882473.

Lupien, S. J., McEwen, B. S., Gunnar, M. R., & Heim, C. (2009). Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nature Reviews Neuroscience*, *10*(6), 434–445.

Marmot, M., Oldfield, Z., Clemens, S., Blake, M., Phelps, A., Nazroo, J., … Oskala, A. (2017). English Longitudinal Study of Ageing: Waves 0–7, 1998–2015. [data collection]. SN: 5050, UK Data Service, 27 edition.

McArdle, J. J., Fisher, G. G., & Kadlec, K. M. (2007). Latent variable analyses of age trends of cognition in the Health and Retirement Study, 1992–2004. *Psychology and Aging*, *22*(3), 525–545.

Moral-Benito, E. (2013). Likelihood-based estimation of dynamic panels with predetermined regressors. *Journal of Business and Economic Statistics*, *31*(4), 451–472.

Nyberg, A., Peristera, P., Westerlund, H., Johansson, G., & Hanson, L. L. (2017). Does job promotion affect men's and women's health differently? Dynamic panel models with fixed effects. *International Journal of Epidemiology*, *46*(4), 1137–1146.

Ofstedal, M. B., Fisher, G. G., & Herzog, R. A. (2005). Documentation of cognitive functioning measures in the Health and Retirement Study. Ann Arbor, Michigan: Institute for Social Research, University of Michigan.

Panza, F., D'Introno, A., Colacicco, A. M., Capurso, C., Del Parigi, A., Caselli, R. J., … Solfrizzi, V. (2009). Temporal relationship between depressive symptoms and cognitive impairment: The Italian Longitudinal Study on Aging. *Journal of Alzheimer's Disease*, *17*(4), 899–911.

Perrino, T., Mason, C. A., Brown, S. C., Spokane, A., & Szapocznik, J. (2008). Functioning and depressive symptoms among hispanic older adults. *Journal of Gerontology: Psychological Sciences*, *63B*(5), 309–317.

Rodgers, W. L., & Ofstedal, M. B. (2011). Trends in scores on tests of cognitive ability in the elderly U.S. Population, 1993. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *58*(6), S348–S349.

Saczynski, J. S., Beiser, A., Seshadri, S., Auerbach, S., Wolf, P. A., & Au, R. (2010). Depressive symptoms and risk of dementia. *Neurology*, *75*(1), 35 LP–35 41.

Sheline, Y. I., Barch, D. M., Garcia, K., Gersing, K., Pieper, C., Welsh-Bohmer, K., … Doraiswamy, P. M. (2006). Cognitive function in late life depression: Relationships to depression severity, cerebrovascular risk factors and processing speed. *Biological Psychiatry*, *60*(1), 58–65.

Sonnega, A., Faul, J. D., Ofstedal, M. B., Langa, K. M., Phillips, J. W. R., & Weir, D. R. (2014). Cohort profile: The Health and Retirement Study (HRS). *International Journal of Epidemiology*, *43*(2), 576–585.

StataCorp. (2017). *Stata statistical software: Release 15*. College Station, TX: StataCorp LLC.

Steel, N., Huppert, F. A., McWilliams, B., & Melzer, D. (2003). Physical and cognitive function. In Marmot, M. G., Blundell, M., & Nazroo, L. C. (eds), *Health, wealth and lifestyles of the older population in England: The 2002 English Longitudinal Study of Ageing*, chapter 7. London: Institute for Fiscal Studies, pp. 249–271.

Steffick, D. E. (2000). Documentation of affective functioning measures in the Health and Retirement Study. Ann Arbor, Michigan: Institute for Social Research, University of Michigan. https://hrs.isr.umich.edu/sites/default/files/biblio/dr-005.pdf.

Steptoe, A., Breeze, E., Banks, J., & Nazroo, J. (2013). Cohort profile: The English longitudinal study of ageing. *International Journal of Epidemiology*, *42*, 1640–1648.

Vaisey, S., & Miles, A. (2017). What you can - and can't - do with three-wave panel data. *Sociological Methods and Research*, *46*(1), 44–67.

Vinkers, D. J., Gussekloo, J., Stek, M. L., Westendorp, R. G., & Van Der Mast, R. C. (2004). Temporal relation between depression and cognitive impairment in old age: Prospective population based study. *British Medical Journal*, *329*(7471), 881–883.

Weir, D., Faul, J., & Langa, K. (2011). Proxy interviews and bias in cognition measures due to non-response in longitudinal studies: A comparison of HRS and ELSA. *Longitudinal and Life Course Studies*, *2*(2), 170–184.

Williams, R., Allison, P. D., & Moral-Benito, E. (2018). Linear dynamic panel-data estimation using maximum likelihood and structural equation modeling. *The Stata Journal*, *18*(2), 293–326.

Zahodne, L. B., Stern, Y., & Manly, J. J. (2014). Depressive symptoms precede memory decline, but not vice versa, in non-demented older adults. *Journal of the American Geriatrics Society*, *62*(1), 130–134.

Zaninotto, P., Batty, G. D., Allerhand, M., & Deary, I. J. (2018). Cognitive function trajectories and their determinants in older people: 8 years of follow-up in the English Longitudinal Study of Ageing. *Journal of Epidemiology and Community Health*, *72*(8), 685–694.

Zivin, K., Llewellyn, D. J., Lang, I. A., Vijan, S., Kabeto, M. U., Miller, E. M., & Langa, K. M. (2010). Depression among older adults in the United States and England. *American Journal of Geriatric Psychiatry*, *18*(11), 1036–1044.

https://doi.org/10.1017/S0033291720003037 Published online by Cambridge University Press