Supplementary File: Statistical Appendix

Let $Z(t)$ denote a stochastic process with state space S = {0, 1, 2,3} and transitions defined by Figure A1. Hence, at time t, $Z(t)$ is 0 if the participant is alive and has normal cognition (NC), 1 if the participant is alive and has mild cognitive impairment (MCI), 2 if the participant is alive and has dementia, and 3 if the participant is deceased. Age is the time scale. Note that all of the study participants were at least 75 years old when entering the study. Let V_0 denote age at study entry (e.g., 1991), which varies by participant. The participants can be in different states at V_0 but must be alive (i.e., $Z(V_0) \neq 3$) to be eligible for selection.

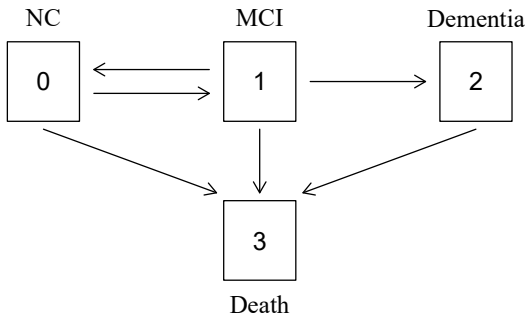

Figure A1: A four-state stochastic process for joint modelling of transitions among three cognitive states and death.
Abbreviation: NC=normal cognition; MCI=mild cognitive impairment

The multistate Markov models are defined by the transition intensities from state j to state k at a given time t such that

$$\lim_{\Delta t \to 0} \frac{\mathrm{P}(Z(t+\Delta t) = k | Z(\mathrm{t}^{-}) = \mathrm{j})}{\Delta t} = \lambda_{jk}(t)$$

where we assume the transition is reversible only between NC and MCI states so that $\lambda_{20}(t) = \lambda_{21}(t) = \lambda_{30}(t) = \lambda_{31}(t) = \lambda_{32}(t) = 0$. The transition intensity can be considered as the instantaneous time/age-specific incidence of making a particular transition, and also a generalization of the hazard function. For instance, $\lambda_{10}(t)$ is the instantaneous risk of reverse

transition from MCI to NC at age t and $\lambda_{12}(t)$ and $\lambda_{13}(t)$ represent the transition rates from MCI to dementia and death, respectively, at age t.

Proportional intensity models are most commonly used to account for the heterogeneity among participants in part by factors coded as covariates. The model is expressed in the form

$$\lambda_{jk}(t;X) = \lambda_{jk0}(t)e^{\boldsymbol{X}^T\boldsymbol{\beta}_{jk}}$$

where $\lambda_{jk0}(t)$ denotes the baseline intensity function, X denotes the vector of covariates of interest and β_{jk} is the vector of regression coefficients, $j \neq k \in \mathcal{S}$. The baseline intensity $\lambda_{jk0}(t)$ depends on time/age, and we initially used a piecewise constant model to categorize age t by five-year age groups [75,80), [80,85), [85,90), [90,95) and ≥ 95 such that $\lambda_{jk0}(t) = \alpha_{jkl}$ when t falls in the lth age interval, $l = 1,2,\ldots,5$. The final model used 90 years as the sole breakpoint given that assessment of the transition intensities between cognitive states for different five-year age groups showed that the transition rates were more similar for those 90 years and younger compared to those above 90 years of age. The exponential of the β coefficient corresponding to a covariate effect has the interpretation of the relative transition rate (RR) which is free of time/age. For example, consider the binary indicator for *APOE*-ε4 (1 for carrier and 0 for noncarrier). Then $\lambda_{10}(t;X=1)/\lambda_{10}(t;X=0) = e^{\beta_{10}}$ is simply the ratio of the reverse transition rates (MCI to NC) comparing an individual with and without at least one *APOE*-ε4 allele at a given time point. This interpretation is comparable to the concept of a hazard ratio in standard survival analysis except that in a multistate Markov model this is interpreted as a cause-specific relative rate due to the presence of the competing event of death. Given that an individual has MCI at age t, a quantity of interest is the ratio of the transition rates (i.e., the RR) for this individual to revert to NC versus progress to dementia, e.g.,

$$RR(t;X) = \frac{\lambda_{10}(t;X)}{\lambda_{12}(t;X)} = \frac{\alpha_{10}\, e^{X\beta_{10}}}{\alpha_{12l}e^{X\beta_{12}}}$$

if t belongs to the lth age category. Based on this formula, the RR of reversion from MCI to NC vs. progression to dementia depends on both the estimate of the age-specific baseline intensities and covariate effects for both types of transitions. To see the covariate effect on the relative

transition rate for an individual to revert to NC vs. progress to dementia at a given time/age, one can take the ratio of RRs at different levels of a covariate. For example, if a covariate X is a binary indicator for *APOE*-ε4 (1 for carrier and 0 for noncarrier), then the ratio of RRs is $RR(t;X=1)/RR(t;X=0) = e^{\beta_{10}-\beta_{12}}$ which does not depend on t.

Suppose that for a given participant the observations of the stochastic process are taken at $V_0, V_1, \ldots, V_m$ and the participant's status is observed until C ($C > V_m$), where V_m is the last assessment time of a living participant. Let T_3 denote the age of death; we observe $T^\dagger = \min(T_3, C)$ and $\delta = \mathrm{I}(T_3 < C)$. The likelihood for multistate Markov models with mixed types of panel observation for cognitive status and exact observation on death can be most naturally written in terms of both transition intensities $\lambda_{jk}(t;X)$ and transition probabilities $P_{jk}(s,t;X) = P(Z(t)=k \mid Z(s)=j;X)$ following Jackson.[1] Conditional on the participant starting in state $Z(V_0)$, the likelihood for the participant takes the form

$$\left[\prod_{k=0}^{m} \mathrm{P}(Z(V_{k+1})|Z(V_k);X)\right] \sum_{j\in\{0,1,2\}} P(Z(T^\dagger)=j\big|Z(V_m);X)\,\lambda_{l3}^{\delta}(T^\dagger;X)$$

where the summation in the second part accounts for the fact that the cognitive status immediately before death or censoring may not be known due to intermittent observation. Under the Markov model, the transition probabilities $P_{jk}(s,t;X)$ can be expressed as a complicated function of transition intensities by solving the Kolmogorov forward differential equation.[2] The global likelihood for observations on all participants is the product of the above terms across all of the individuals in the sample. The msm package[1] in R can be used to maximize the global likelihood.

References

1. Jackson C. Multi-state models for panel data: the msm package for R. J Stat Softw 2011;38.

2. Cox DR, Miller HD. The theory of stochastic processes. New York: Wiley; 1965.