**JASA ACS Reproducibility Initiative - Author Contributions Checklist Form**

The purpose of the Author Contributions Checklist (ACC) Form is to document the code and data supporting a manuscript, and describe how to reproduce its main results.

As of Sept. 1, 2016, the ACC Form must be included with all new submissions to JASA ACS.

This document is the initial version of the template that will be provided to authors. The JASA Associate Editors for Reproducibility will update this document with more detailed instructions and information about best practices for many of the listed requirements over time.

## Data

**Abstract (Mandatory)**

The primary data used for this paper comes from the Mayo Clinic Study of Aging. It includes demographic and biomarker data on 4742 subjects since 2004. Other data includes the CAV data set from the msm package in R.

**Availability (Mandatory)**

The actual data set analyzed from the Mayo Clinic Study of Aging cannot be made available because it contains HIPAA protected medical information. However, we include a synthetic data set, which closely resembles the real data set. The CAV data is publicly available in the msm package in R.

**Description (Mandatory if data available)**

The synthetic Mayo Clinic Study of Aging data set is in the form of a ‘.rda’ file, and further details about the data are provided in the manuscript and supplementary materials. Documentation for the CAV data set is provided within the R package documentation for msm.

## Code

**Abstract (Mandatory)**

We have written code to estimate the hidden Markov model described in the manuscript, and to carry out the two simulation studies presented in the manuscript.

**Description (Mandatory)**

All code has been written in the R programming language. Descriptions and organization of the code are made available in the supplementary material in the file submitted with the manuscript, ‘TimeCapsule\_Code\_HMMDem.tar.gz’. Included within is a ‘workflow.sh’ file that provides line-by-line Unix command line code for running the contained R script files to reproduce all results presented in the manuscript and supplementary material. The ‘.tar.gz’ file is also made publicly available at https://jonathanpw.github.io/software.html.

The code used for producing the manuscript results uses R version 3.3.2, and the following R packages:

expm (version 0.999.0), matrixStats (version 0.52.2), splines (version 3.3.2), survival (version 2.41.3), mclust (version 5.3), diagram (version 1.6.4), animation (version 2.5), msm (version 1.6.4), MASS (version 7.3.45), mvtnorm (version 1.0.5), parallel (version 3.3.2)

## Instructions for Use

**Reproducibility (Mandatory)**

All of the results/figures/tables in our paper can be reproduced using the code provided in ‘TimeCapsule\_Code\_HMMDem.tar.gz’. Follow the line-by-line instructions in the included ‘workflow.sh’ file to reproduce all results. The one caveat is that due to HIPAA restrictions we cannot provide the real data set, so reproducing the results will entail running the code on the synthetic data set that we provide. Note that this will inevitably yield slightly different estimates of the hidden Markov model parameters.

Note that parallelization was used extensively within each step of the MCMC routine to evaluate the likelihood function of the data. For the real and synthetic MCSA data 30 threads are used, and for the CAV simulation 8 threads are used. The runtimes for each seed of the MCMC routines are as follows:

Real MCSA data – approximately 3 days

Sensitivity analysis of the real MCSA data – approximately 3 days

Simulation results for synthetic MCSA data – approximately 2 days

Simulation results for toy CAV data – approximately 1 day

**Replication**

We hope to finish developing our code into an R package in the future.