A Bayesian Latent Class Mixture Modeling Framework for Algorithmic Dementia Classification in Population-Representative Studies

Crystal Shaw[1,2], Elizabeth Rose Mayeda[2], Donatello Telesca[1]
Thomas R. Belin[1]
[1]Fielding School of Public Health, Department of Biostatistics, University of California, Los Angeles, 650 Charles E. Young Drive South, Los Angeles, CA 90095
[2]Fielding School of Public Health, Department of Epidemiology, University of California, Los Angeles, 650 Charles E. Young Drive South, Los Angeles, CA 90095

Abstract
Gold-standard dementia ascertainment, involving neuropsychological testing, clinical exams, and diagnosis adjudication by a panel of clinicians, is resource-intensive and infeasible in large population-representative studies, presenting a barrier to understanding population-level burden and determinants of dementia. Algorithmic dementia classification methods are an alternative, but lack key measures used for clinical dementia ascertainment (e.g., detailed neuropsychological assessments). The first step to strengthening algorithmic dementia classification in large cohorts is developing frameworks that can make use of these important measures. We developed a Bayesian latent class mixture modeling framework for algorithmic dementia classification that incorporates neuropsychological measures and sociodemographic, health, and health behavior information.

Methods were demonstrated and validated using detailed neuropsychological measures from the Aging, Demographics, and Memory Study (ADAMS, n = 520), a substudy of the US Health and Retirement Study that included detailed neuropsychological assessments and clinical dementia adjudication (unimpaired n=211, cognitive impairment due to other conditions (other) n=86, mild cognitive impairment n = 65, dementia n=158). Utilizing a portion of the ADAMS data as a training sample (n=364), Bayesian methods were used to fit latent class mixture models and generate synthetic datasets with subgroups distinguished based on impairment characteristics, a process that overlaps with what has been termed “algorithmic dementia classification” in the epidemiologic and aging literature. The remaining hold-out portion of the ADAMS data (n=156) was used to evaluate the model’s external validity.

We generated one thousand synthetic version of ADAMS training and hold-out samples. Data quality checks showed that synthetic samples reproduced characteristics of the ADAMS cohort (e.g., similar covariate distributions between synthetic datasets and observed data). Algorithmic dementia classification within the ADAMS training sample yielded 95% credible intervals that captured the observed count in all ADAMS impairment classes (unimpaired, other, MCI, dementia) with the mean count for the dementia class possessing the largest discrepancy compared with ADAMS observed counts (25 individuals out of 364 in the sample). Algorithmic dementia classification in the ADAMS hold-out sample yielded 95% credible intervals that captured the observed count in all ADAMS impairment classes with the largest mean-count discrepancy being

five individuals in the MCI class (out of 156 in the sample). Thus, our model for algorithmic dementia classification was highly concordant with ADAMS clinical diagnoses. This work demonstrates opportunities for improving algorithmic dementia classification in large cohorts where clinical diagnoses are unavailable.

Key Words: Bayesian models, latent class mixture models, algorithmic dementia classification, HRS, ADAMS, HCAP

1. Introduction

Dementia is characterized by cognitive impairment severe enough to impact functional ability; it is heterogenous in presentation, making clinical diagnosis challenging (Alzheimer's Association, 2021). Accepted gold-standard dementia diagnosis involves hours of neuropsychological testing, a clinical exam, informant interview, medical history, and consensus diagnosis by an expert panel (Langa et al., 2005; Mayeux et al., 2011; McKhann et al., 2011). This time- and resource- intensive procedure is infeasible in population-based surveys, which are of considerable public health interest to develop dementia prevention strategies and reduce disparities in incidence and treatment across subgroups. Algorithmic dementia classification methods have been developed to predict individuals' probability of dementia in these large studies (Gross et al., 2017; Kasper et al., 2013; Prina et al., 2019), but the lack of available data on neuropsychological testing in population-based surveys has been a persistent limitation of existing dementia classification strategies (Gianattasio et al., 2019). The aim of the present work is to develop an algorithmic dementia classification framework incorporating additional predictors known to be important in gold-standard dementia diagnosis, specifically detailed neuropsychological measures that are available in a substudy of a population-based survey.

The Health and Retirement Study (HRS) is a large, ongoing (1992-present) population-representative study in which a variety of algorithmic dementia classification methods have been used to predict participants' probability of dementia (Gianattasio et al., 2019). HRS collects measures associated with aging and thus, has extensive information available on sociodemographic characteristics, lifestyle and health variables, and general cognitive assessments (Sonnega et al., 2014). Though HRS does not have detailed neuropsychological assessments available for all participants, the Aging, Demographics, and Memory Study (ADAMS [2001-2009]) (Langa et al., 2005) and the Harmonized Cognitive Assessment Protocol (HCAP [2016-present]) (Weir et al., 2016) collected neuropsychological test data for a subset of HRS participants. In this present paper, we develop methods that lay the groundwork for translating information from detailed neuropsychological assessments available in studies like ADAMS or HCAP to population-representative surveys like HRS to strengthen algorithmic dementia classification in population-representative samples.

2. Methods

2.1 Data Example: Aging, Demographics, and Memory Study (ADAMS)

Methods in the present paper were validated using ADAMS, a substudy of HRS that included detailed neuropsychological assessments and clinical dementia adjudication (Heeringa et al., 2009). There are four waves of ADAMS data (A-D) spanning the years

2001-2009. ADAMS was a stratified random sample of n=856 HRS participants aged 70 or older with strata determined by performance on HRS general cognitive assessments in 2001 (cognitively normal, borderline impaired, low functioning). All ADAMS participants received a clinical impairment diagnosis, and participants were followed until death or dementia diagnosis.

Wave A ADAMS data was used for the present analysis. The sample was restricted to a complete-case subset of participants for which all relevant covariate measures were available (n=520). We collapsed ADAMS impairment diagnoses into four general categories: (1) Unimpaired (n=211), (2) Other (cognitive impairment due to conditions other than dementia, e.g., depression or traumatic brain injury, n=56), (3) Mild Cognitive Impairment (MCI, n=65), and (4) Dementia (n=158). The complete-case data was split into 70% training (n=364) and 30% hold-out (n=156) for internal and external model validation. All analyses were done using R version 4.0.3 (R Core Team, 2020).

2.2 Important Predictors of ADAMS Cognitive Impairment Diagnoses

An initial step in model formulation was understanding important predictors of ADAMS adjudicated impairment classes. Rather than defining one model for the multi-level categorical outcome (Unimpaired vs. Other vs. MCI vs. Dementia), we used multi-part models which specify different logistic regression models at each stage of separation. Three models were used to distinguish between impairment classes: (1) Unimpaired vs. Impaired, (2) Other vs. MCI or Dementia, and (3) MCI vs. Dementia, where models (2) and (3) were conditional on individuals being classified as being impaired or having MCI or Dementia, respectively. Modeling impairment classes this way naturally accommodates non-linear relationships and different subsets of predictors (Olsen & Schafer, 2001).

A complete list of candidate predictors for the multi-part models is provided in **Appendix Table A1**. Ordinal categorical variables were analyzed on a continuous scale. Important predictors of ADAMS diagnosed impairment class were chosen by fitting multi-part models in the ADAMS training sample (n = 364). Variables were entered sequentially starting with consideration of fully observed variables as predictors, and variables that led to more than 25% missing observations in the model were not considered. As an initial simplifying assumption to identify important predictors, predictors of cognitive status that were significant at the p = 0.05 level were retained in the models. See **Discussion** for plans to update this process.

Let $G_i, i = 1, \dots, 364$ denote the ADAMS adjudicated impairment class (group) for each individual in the ADAMS training sample,

$$G_i = \begin{cases} 1 & \text{if participant } i \text{ is Unimpaired} \\ 2 & \text{if participant } i \text{ has Other impairment} \\ 3 & \text{if participant } i \text{ has MCI} \\ 4 & \text{if participant } i \text{ has Dementia} \end{cases} \quad (1)$$

Letting X denote the vector of candidate predictor variables including an intercept and $\boldsymbol{\gamma_1}, \boldsymbol{\gamma_2}, \boldsymbol{\gamma_3}$ being vectors of regression coefficients, the following logistic regression models were fit in the ADAMS training sample:

$$logit(P(G = 1|X)) = \boldsymbol{\gamma_1} X \quad (2)$$

$$logit(P(\mathrm{G} = 2|X, G \neq 1)) = \boldsymbol{\gamma_2} X \quad (3)$$

$$logit(P(G = 3\,|X, G \neq 1, G \neq 2) = \boldsymbol{\gamma_3} X. \quad (4)$$

Table 1 lists the specific predictors with non-zero regression coefficients reflecting that $\boldsymbol{\gamma_1}, \boldsymbol{\gamma_2}, \boldsymbol{\gamma_3}$ are distinct from one another.

Ideally, all important predictors of impairment would be available for participants we aim to classify. Crucial measures are often unavailable in the larger sample where we aim to predict impairment but are available in a subset of the study. For example, the Total Mini-Mental State Exam (MMSE) Score, which was a consistently important predictor across the multi-part models (**Table 1**) was only available in ADAMS, not HRS. A reasonable strategy might be to impute this missing variable for the rest of HRS using multiple imputation (Rubin, 1996; Stef van Buuren, 2019) or semi-supervised learning methods (Zhang et al., 2019). MMSE is a notoriously skewed variable, though, and modeling attempts had difficulty recovering observed values, especially in the tails of the distribution.

Table 1: Variables included in multi-part models for predicting different stages of impairment.

Model 1 *Unimpaired vs. Impaired*	*Model 2* *Other* vs. MCI† or Dementia*	*Model 3* *MCI† vs. Dementia*
Total MMSE‡ Score	Total MMSE‡ Score	Total MMSE‡ Score
Immediate Word Recall	Immediate Word Recall	Immediate Word Recall
Age	Age	
Race/Ethnicity		
Serial 7s		
Word List Recall (yes)		
Logical Memory I		
Average Proxy Cognition		
	Delayed Word Recall	
		IADLs#
		BMI
		Stroke History (yes/no)

*Other: Cognitive impairment due to other conditions (e.g., depression, traumatic brain injury)
†MCI: Mild Cognitive Impairment
‡MMSE: Mini-mental state exam
#IADL: Instrumental Activities of Daily Living

2.3 Bayesian Latent Class Mixture Model

Moving from modeling MMSE using a single distribution to a mixture of distributions was motivated by the clinical practice of classifying individuals into different impairment groups. One of the more challenging steps in modeling data using mixture distributions is choosing the number of distributions to use. The 4-class mixture was motivated by ADAMS impairment categories and the practice of classifying individuals as having dementia, MCI, or no cognitive impairment in cohort studies with gold-standard adjudication (Bennett et al., 2012; Demirovic et al., 2003; Knopman et al., 2016; Lopez et al., 2012; Manly et al., 2005; Plassman et al., 2007; Trittschuh et al., 2011; Wilson et al., 2010). Further, visualizing the distribution of MMSE scores stratified by ADAMS

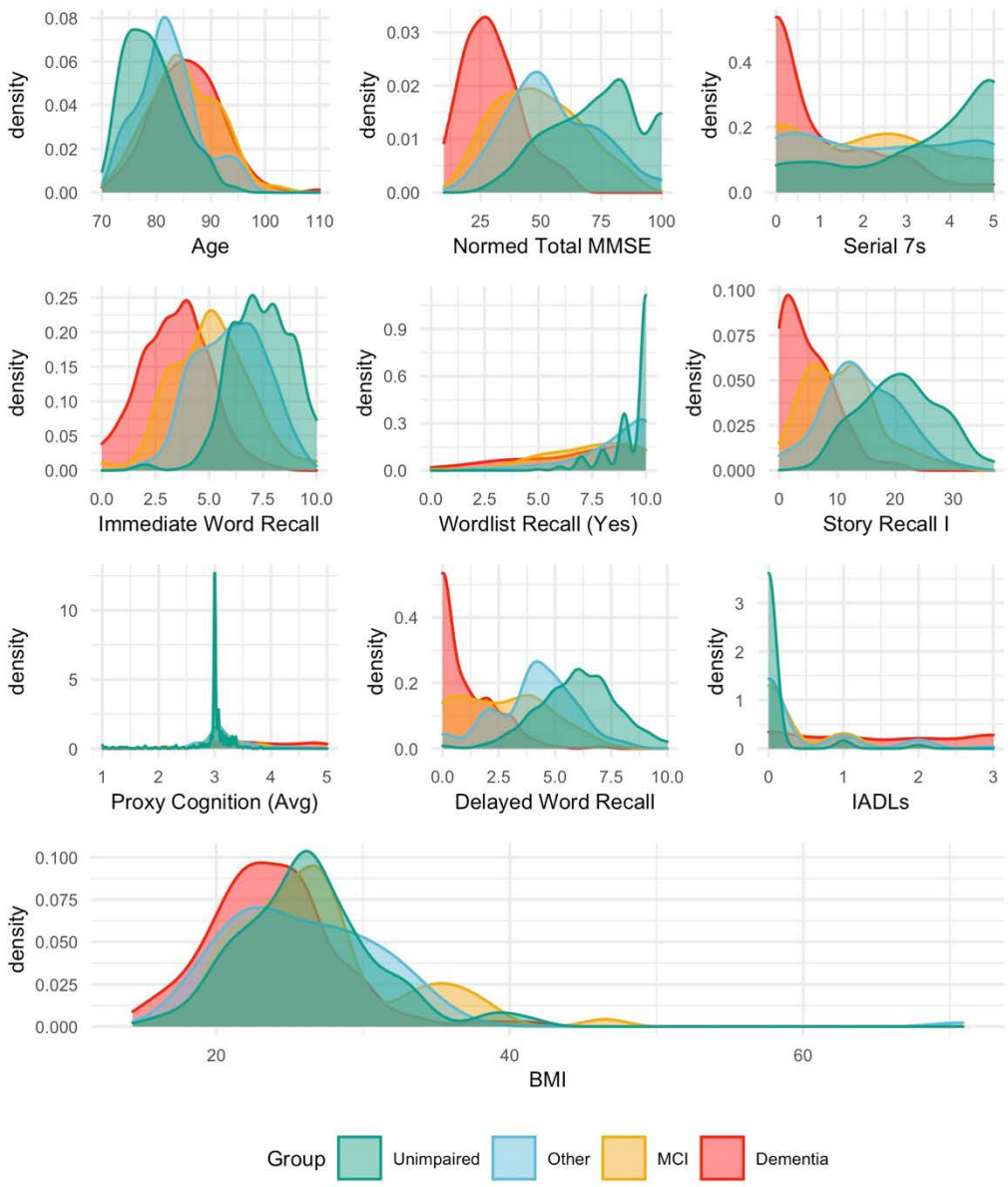

Figure 1: Density plots for continuous variables in **Table 1** stratified by ADAMS impairment class.

impairment classes revealed that a mixture of distributions might do a better job of recovering the overall observed distribution of MMSE scores (**Figure 1**). To motivate modeling Total MMSE scores as a mixture of normal distributions, the Total MMSE score was transformed from its original scale [0, 30] to a normalized scale [0, 100] using a transformation developed and validated by (Philipps et al., 2014). The distribution of normed Total MMSE scores more closely resembled a normal distribution compared to raw Total MMSE scores. The transformation function is available in the NormPsy R Package (Proust-Lima & Philipps, 2018).

Any dataset could be viewed as a mixture of individuals who have no impairment, other impairment, MCI, or dementia. The correct mix of these individuals is the inference goal

of algorithmic dementia classification methods. Latent class mixture models simultaneously model data and infer individual impairment class membership. We embedded the three broad steps of the proposed latent class mixture modeling approach for algorithmic dementia classification in a Bayesian framework to incorporate prior information in the model: (1) Make a synthetic version of a dataset with detailed neuropsychological assessment data but unknown dementia classification. (2) While making the synthetic copy, the model will assign impairment status to each individual. Since the mixture of impairment classes determines what the synthetic data look like, the better the synthetic data quality (i.e., the more it looks like the real data) the more trustworthy the inferred impairment classes. (3) generate many synthetic datasets to measure uncertainty in inference.

2.3.1 The General Location Model

The Bayesian latent class mixture model outlined above is an extension of the general location model which provides a framework for modeling a mix of categorical and continuous variables *(Little & Schluchter, 1985; Olkin & Tate, 1961; Schafer, 1997)*. Using the general location model, continuous variables are modeled using normal distributions with parameters determined by an observation's contingency cell membership, determined by cross-classification of categorical variables. Density plots of continuous variables in **Table 1**, stratified by ADAMS impairment class are shown in **Figure 1**. By inspection, a number of the variables could reasonably be modeled as mixtures of normal distributions, which motivated the use of the general location model in this framework.

Following the setup of (Schafer, 1997), let W_1 and W_2 be the categorical variables in **Table 1**, race/ethnicity (white, Black, Hispanic) and stroke history (ever/never), and let $Z_1, Z_2, \dots Z_{10}$ be the continuous variables in **Table 1**. Let $X = (W, Z)$ be an $n \times 12$ matrix of observed data. Contingency cell membership in the present analysis was determined by cross classification of the two categorical variables race/ethnicity and stroke history, yielding 6 possible cells. Let $C = \{c_d : d = 1, 2, \dots 6\}$ be the vector of observed counts for each contingency cell and let U be an $n \times d$ matrix with rows u_i^T, where u_i is a D-vector with a 1 in position d if observation i falls into cell d and 0s in all other position. All the information about W is contained in C, thus the distribution of X can be characterized as $f(X) = f\big((W, Z)\big) = f\big((C, Z)\big) = f(Z \mid C) f(C)$, where

$$C \sim M(n, \pi) \tag{5}$$

$$Z_i | u_i \sim N(\mu_d, \Sigma). \tag{6}$$

For the distribution in (5), $\pi = \{\pi_d : d = 1, 2, \dots, 6\}$ is a vector of cell probabilities parameterizing the Multinomial distribution corresponding to C. Note that the mean of the normal distribution in (6) is indexed by d, denoting that the means are allowed to vary by contingency cell but with an assumed constant covariance structure Σ across cells. This unrestricted formulation of the general location model includes main effects of race/ethnicity and stroke and all race/ethnicity by stroke interactions effects on values of the continuous variables Z.

Due to small cell counts, however, the model had difficulty estimating all race/ethnicity by stroke interaction effects. on Z. We moved to a restricted general location model where $\mu = \{\mu_d : d = 1, \dots, 6\}$ is restricted to the form

$$\mu = A\beta, \tag{7}$$

and A is an ANOVA-like design matrix that specifies main effects of race/ethnicity and stroke on Z only.

2.3.2 The Synthetic Data Generating Model

The restricted general location model described in **Section 2.3.1** was embedded in a Bayesian latent class mixture model to generate synthetic versions of observed data. Markov Chain Monte-Carlo (MCMC) was used to sample from posterior distributions of the parameters. At iteration b of an MCMC chain, let G^b be the predicted impairment class from the multi-part models described in **Section 2.2**. The restricted general location model was used to model data within each subset defined by G^b. The impairment group-specific Bayesian specification of the restricted general location model at iteration b for was

$$Z_i|u_i \sim N_{10}(\mu_G, \Sigma_G) \tag{7}$$

$$C_G \sim M(n_G, \pi_G) \tag{8}$$

$$\beta_G|\Sigma_G \sim MN_{4\times10}(\beta_0, V_0, \Sigma_G/\kappa_0) \tag{9}$$

$$\Sigma_G \sim W_{\nu_0}^{-1}(\Lambda_0^{-1}) \tag{10}$$

$$\pi_G \sim D(\alpha_G) \tag{11}$$

with hyperparameters κ_0, ν_0, α_G, and Λ_0^{-1}, and where $\mu_G = A\beta_G$

A diagram of the data generating model is provided in **Figure 2**, including the specific posterior distributions from which variables were sampled. Posterior distributions were derived using similar techniques to those outlined in (Schafer, 1999) and (Gelman et al., 2014).

2.3.3 Data-driven Priors via Bootstrap Sampling

Sampling in this framework was fast and convenient due to conjugate distributions, thus computational time was not a challenge in this model. Small contingency cell counts led to difficulty defining priors in this model, however. Non-informative priors led to model convergence issues. Increasing prior cell counts by defining priors based on cell counts from the larger HRS sample improved convergence but led to poor model fit due to the difference in case mix between HRS and ADAMS. Defining data-driven priors based on bootstrap sampling proved to be a valuable technique to overcome these challenges.

Bootstrapping uses resampling with replacement to obtain realistic replicates of the data *(Efron & Tibshirani, 1994)*. Since the goal of this analysis was to create realistic synthetic versions of ADAMS, bootstrapping seemed like a promising way of reproducing the empirical distribution of the data while avoiding the need to make oversimplifying distributional assumptions. Replicate data were obtained through a three-step process: (1) Resample ADAMS data with replacement, drawing a sample of equal size to the original sample. (2) Calculate and store parameter estimates characterizing effects of covariates on impairment class membership, contingency cell counts, and effects of contingency cell membership on continuous covariates. (3) Repeat the process

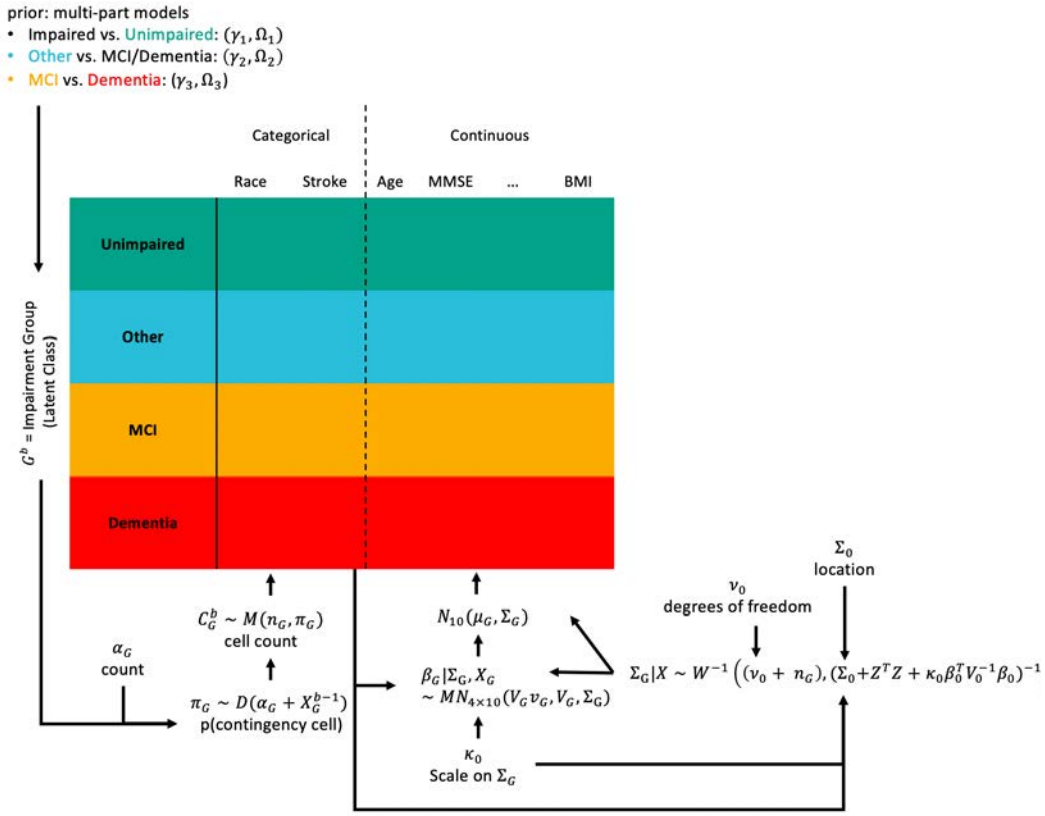

Figure 2: Data-generating model for iteration b of an MCMC chain.
M = Multinomial distribution; D = Dirichlet distribution; N = Normal distribution; MN = Matrix normal distribution; W^{-1} = Inverse Wishart distribution; V_G and ν_g are parameters for the posterior MN distribution that were derived using usual techniques.

10,000 times to represent both sampling variability and estimation uncertainty in model parameters. This process was motivated by Bayesian non-parametric methodology (Rubin, 1981) and empirical Bayes concepts (Casella, 1985).

3. Bayesian Model Diagnostics

Understanding how model assumptions encoded in the prior interact with the likelihood to affect posterior inferences is an important part of Bayesian modeling.(Gabry et al., 2019) discuss the value of visualizations at every stage of Bayesian modeling and outline the workflow for prior predictive checks, model diagnostics, and posterior predictive checks using a real data example. Each of these modeling checks is described below as it relates to the present analysis. All checks were performed for both ADAMS training and ADAMS hold-out samples.

3.1 Prior Predictive Checks

Prior predictive checks are visualizations of synthetic datasets generated from prior distributions only. These checks are meant to answer the question "are the priors compatible with the data?" In other words, do assumptions encoded in the prior lead to realizations of the data that capture the full range of possible values? In developing prior distributions, we embraced the perspective of (Gabry et al., 2019) where prior predictive

checks are envisioned as part of an iterative process of specifying models, fitting models, evaluating model fit, and updating model specifications.

In the present analysis, prior predictive checks were performed for distributions of contingency cell counts stratified by ADAMS impairment class and for continuous predictors in the model. One thousand synthetic ADAMS datasets were generated from a model using only prior distributions. Contingency cell counts and distributions of continuous variables were stored for each synthetic dataset. Ideal prior predictive distributions for 1000 synthetic contingency cell counts would be centered at the true count. Ideal distributions of synthetic continuous variables would be slightly more variable than the observed distributions to ensure that the full range of values were captured features encoded in the priors.

3.2 Model Convergence Diagnostics

Once prior predictive checks are satisfactory, many synthetic datasets should be generated from the full model and convergence assessed across these runs. MCMC chains of all model parameters are monitored for convergence. Ideal MCMC plots look like "fuzzy caterpillars" that stabilize around some value (Gelman et al., 2014). A lack of convergence signals the need for model tuning by adjusting hyperparameters until proper convergence is achieved. In this analysis, MCMC chains of impairment class proportions and continuous variable means and covariances stratified by contingency cells were monitored for convergence.

Model stability can be assessed by monitoring convergence of multiple chains initiated at different locations in the sample space. Ideal diagnostic plots for multiple chain convergence would look like overlapping "fuzzy caterpillars" which would demonstrate that each chain has converged and that the chains mix well (i.e., model convergence is robust to the starting point in the sample space). In the present analysis, MCMC chains were initiated in different parts of the sampling space based on proportions of impairment class membership. Five chains were monitored: (1) a "warm start" chain with impairment class proportions close to observed proportions (40% Unimpaired, 20% Other, 10% MCI, 30% Dementia), (2) a "random" chain with equal proportions for all impairment classes (25% Unimpaired, 25% Other, 25% MCI, 25% Dementia), (3) "mostly dementia" chain where proportion of dementia dominated other impairment classes (10% Unimpaired, 20% Other, 30% MCI, 40% Dementia), (4) "mostly MCI" chain where proportion of MCI dominated other impairment classes (10% Unimpaired, 30% Other, 40% MCI, 20% Dementia), and (5) "mostly impaired" chain where only 5% of participants were initiated in the unimpaired class (5% Unimpaired, 15% Other, 25% MCI, 55% Dementia).

3.3 Posterior Predictive Checks

Once proper model convergence is achieved, posterior predictive checks assess whether salient features of the data are captured by the model. Analogous to prior predictive checks, ideal posterior predictive distributions would be centered around the observed value of the statistic of interest. In the present analysis, posterior distributions of contingency cell counts and median and skew for continuous variables were assessed. All posterior statistics were stratified by ADAMS impairment class.

4. Results

The Bayesian latent class mixture model was fit in ADAMS training and hold-out samples and 1000 synthetic versions of each sample were produced. Results from

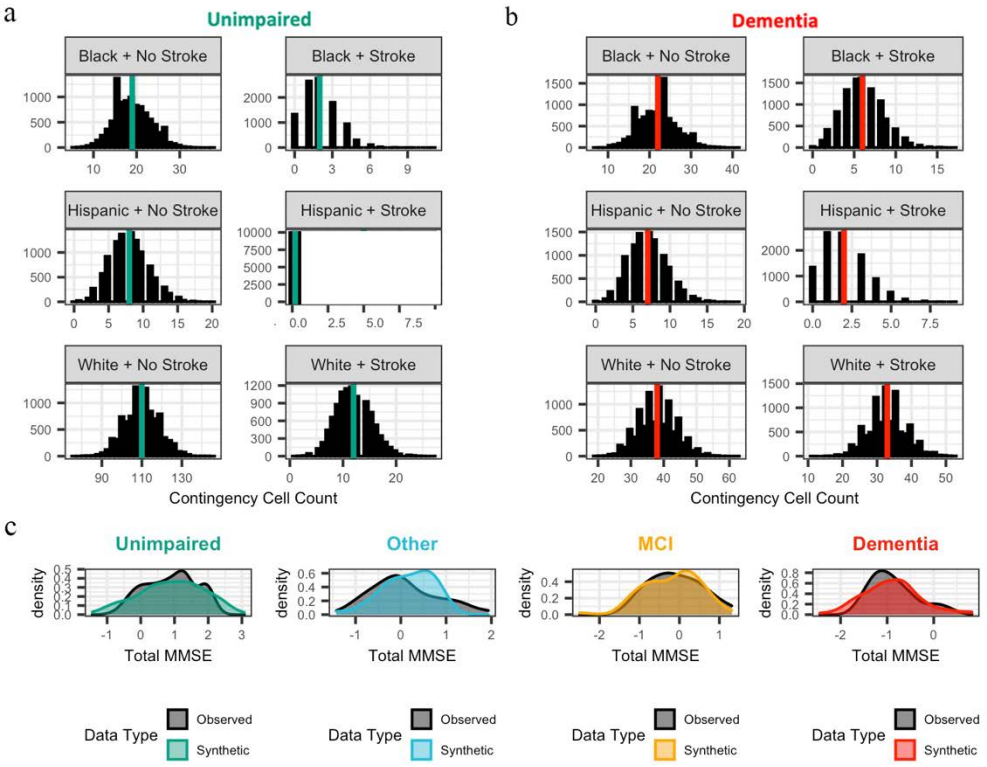

Figure 3: Prior predictive distributions based on 1000 synthetic datasets for (a) contingency cell counts in the ADAMS Unimpaired group, (b) contingency cell counts in the ADAMS Dementia group, and (c) Total MMSE in all ADAMS impairment groups.

selected prior predictive checks in the ADAMS training sample are presented in **Figure 3**. As expected from bootstrap prior distributions, prior predictive contingency cell counts were centered around observed counts (colored lines in **Figures 3a-b**). By nature of resampling the data, any observed 0 cell counts remained 0 in prior predictive distributions (Hispanic + Stroke group, **Figure 3a**). Prior predictive distributions for Total MMSE stratified by ADAMS impairment class were more variable than the observed ADAMS data in most of the 1000 synthetic datasets, which demonstrated that the full range of observed values was captured by prior distributions (**Figure 3c**). **Figure 3c** shows Total MMSE prior predictive distributions overlayed on observed ADAMS training data for one synthetic dataset; an animated gif cycling through all 1000 synthetic datasets was used to determine whether there was enough variability across the 1000 synthetic datasets.

All MCMC chains showed model parameter convergence. Analyses of multiple chains of impairment class proportions showed good convergence and mixing, demonstrating that the model was stable regardless of the starting point (**Figure 4**).

Results from selected posterior predictive checks in the ADAMS training sample are presented in **Figure 5**. Posterior cell counts for Black and Hispanic participants were satisfactory while those for white participants were less so. Distributions of posterior

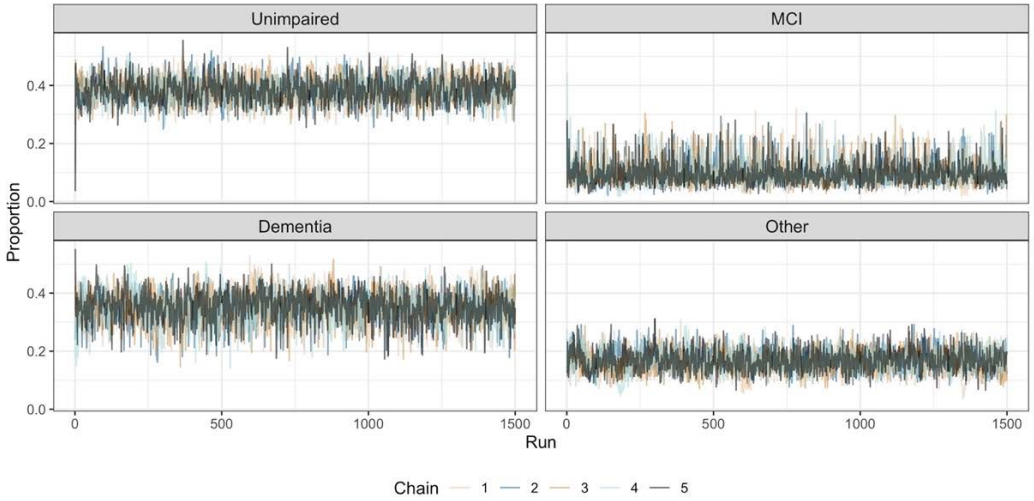

Figure 4: Five MCMC chains of proportion of impairment class membership, stratified by impairment class. Each chain was initiated at different points in the parameter space. Chain 1 (warm start): 40% Unimpaired, 20% Other, 10% MCI, 30% Dementia; Chain 2 (random chain): 25% Unimpaired, 25% Other, 25% MCI, 25% Dementia; Chain 3 (mostly dementia): 10% Unimpaired, 20% Other, 30% MCI, 40% Dementia; Chain 4 (mostly MCI): 10% Unimpaired, 30% Other, 40% MCI, 20% Dementia; Chain 5 (mostly impaired): 5% Unimpaired, 15% Other, 25% MCI, 55% Dementia.

counts for white participants captured true counts, but only in the tails of the distributions (**Figure 5a-b**). Of note, a desirable property of this model is that cell counts that were 0 in the prior did not remain zero in the posterior. Since the 0 cell counts were random, not structural, realistic replicates of the dataset would be expected to have small, non-zero cell counts. Posterior distributions of continuous variable medians were roughly centered around observed medians in the ADAMS data apart from a couple variables (**Figure 5c**).

The complete set of prior predictive checks including animated gifs for continuous variables, MCMC convergence plots, and posterior predictive checks for both ADAMS training and testing samples is accessible at https://github.com/cshawsome/link-transport-integrate.

Figure 6 shows 95% credible intervals of participant counts in each impairment class across 1000 synthetic ADAMS training datasets. Every credible interval captured the observed ADAMS impairment class count; the largest discrepancy in mean count was in the dementia group where the model overestimated the count by just 25 people on average. The modeling process described in **Section 3** was repeated for the hold-out sample.

Figure 7 shows 95% credible intervals of participant counts in each impairment class from 1000 synthetic ADAMS hold-out datasets. Again, every credible interval captured the observed ADAMS impairment class count, and in the hold-out sample the largest discrepancy was in the MCI group where the model underestimated the count by just 5 people on average.

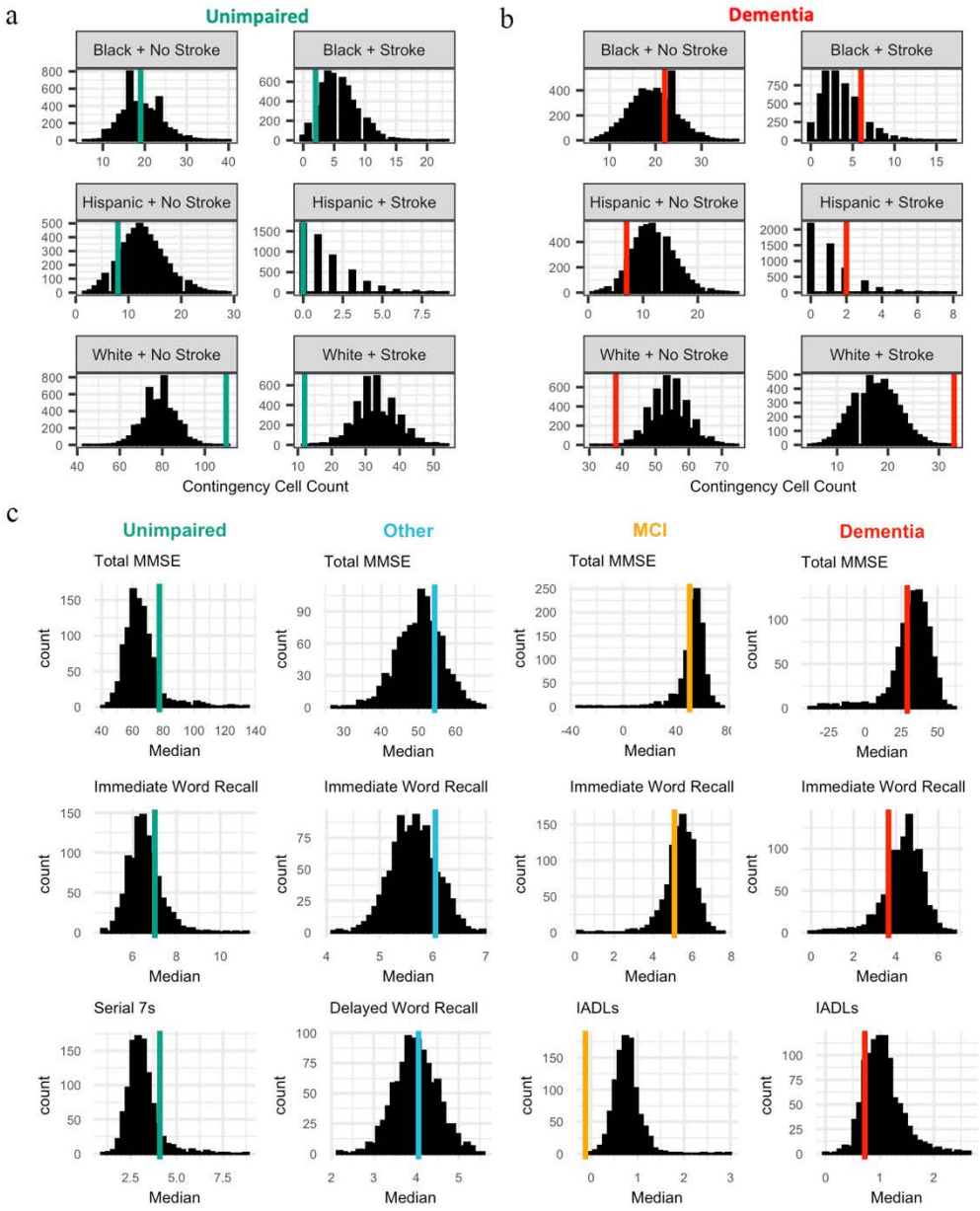

Figure 5: Posterior predictive distributions based on 1000 synthetic datasets for (a) contingency cell counts in the ADAMS Unimpaired group, (b) contingency cell counts in the ADAMS Dementia group, and (c) selected continuous variables important in each multi-part model.

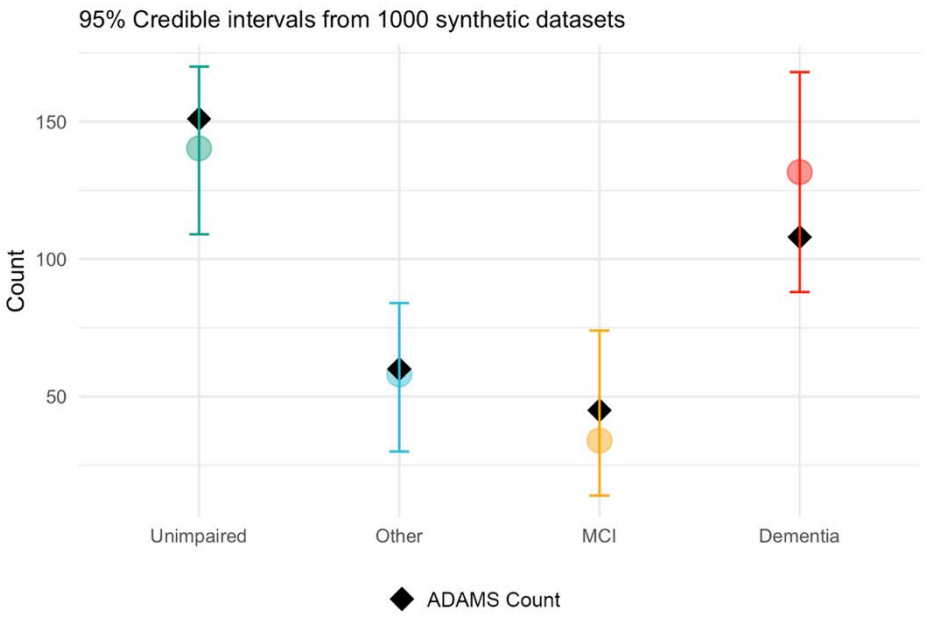

Figure 6: ADAMS training sample results. 95% credible intervals for participant counts within each impairment group across 1000 synthetic datasets compared to observed ADAMS counts.

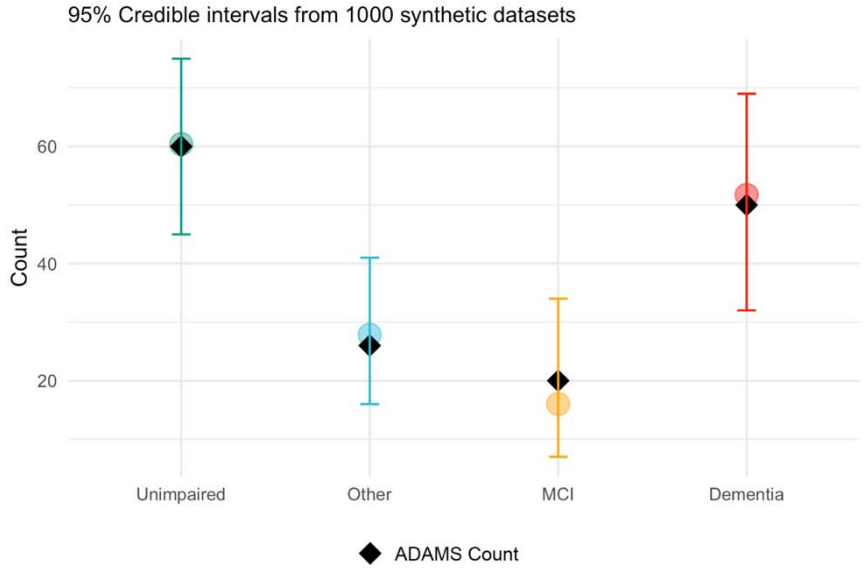

Figure 7: ADAMS hold-out sample results. 95% credible intervals for participant counts within each impairment group across 1000 synthetic datasets compared to observed ADAMS counts.

5. Conclusion

We have illustrated an algorithmic dementia classification framework that incorporates information from sociodemographic characteristics, health and health behaviors, general cognitive assessments, and detailed neuropsychological measures which are crucial in clinical dementia diagnosis; but due to a lack of availability in large studies, are currently not used in dementia classification algorithms. Understanding the relationship between

detailed neuropsychological measures and participants' impairment status is an important first step to translating information from these measures to more general samples so that the information can be used in algorithmic dementia classification methods. By successfully creating synthetic versions of ADAMS using a Bayesian latent class mixture model, we have laid the groundwork for translating key measures to and strengthening algorithmic dementia classification in population-based surveys.

An immediate next step in this project is to expand this framework to handle participants with missing covariate information by performing multiple imputation (Stef van Buuren, 2019) on the dataset prior to fitting the Bayesian latent class mixture model. The variable selection procedure described in **Section 2.2** involved simplifying assumptions that will be relaxed once data are multiply imputed and all candidate variables can be given full consideration.

Posterior predictive checks of contingency cell counts showed some lack of fit for categorical variables. Additional next steps for this project are to remedy this by relaxing some modeling assumptions. One possible strategy is to use a flexible covariance structure across impairment classes and contingency cell categories if necessary (Liu & Rubin, 1998).

Methods in the present work were demonstrated and validated using the ADAMS substudy of HRS. Our next goal is to apply this algorithmic dementia classification framework to HCAP, a newer (initiated in 2016) and larger (n=3,496) HRS substudy with neuropsychological assessments. HCAP, however, did not perform clinical dementia ascertainment for its participants. Models are currently being developed by HCAP investigators to predict probability of impairment for HCAP participants (Langa et al., 2020). In contrast to existing methods that predict probability of dementia for participants, the Bayesian latent class mixture modeling framework presented here would classify HCAP participants into multiple impairment classes by leveraging sociodemographic characteristics, lifestyle and health variables, general cognitive assessments, and detailed neuropsychological assessments. The ultimate goal of this work is to generalize synthetic versions of HCAP to the full HRS sample for algorithmic dementia classification in a population-representative study.

Acknowledgements

Authors would like to thank the MEthods for LOngitudinal studies in DEMentia (MELODEM) initiative for providing an inviting and supportive venue through their bi-monthly working group meetings and annual progress meeting for early presentations of this work that led to productive conversations and useful feedback. We are grateful to the following individuals who provided helpful feedback and guidance during early phases of this project: Dr. Maria Glymour (University of California, San Francisco, Department of Epidemiology and Biostatistics), Dr. Dan Mungas (University of California, Davis, Department of Neurology), and Dr. Rob Weiss (University of California, Los Angeles, Department of Biostatistics).

This work was supported by grant F31AG071191 (PI: Shaw) from the National Institute on Aging.

References

Alzheimer's Association. (2021). 2021 Alzheimer's Disease Facts and Figures. *Alzheimers and Dementia*, *17*(3), 1–108.

Bennett, D. A., Schneider, J. A., Buchman, A. S., Barnes, L. L., Boyle, P. A., & Wilson, R. S. (2012). Overview and findings from the rush Memory and Aging Project. *Current Alzheimer Research*, *9*(6), 663. https://doi.org/10.2174/156720512801322663

Casella, G. (1985). An Introduction to Empirical Bayes Data Analysis. *The American Statistician*, *39*(2), 83–87.

Demirovic, J., Prineas, R., Loewenstein, D., Bean, J., Duara, R., Sevush, S., & Szapocznik, J. (2003). Prevalence of dementia in three ethnic groups: The South Florida program on aging and health. *Annals of Epidemiology*, *13*(6), 472–478. https://doi.org/10.1016/S1047-2797(02)00437-4

Efron, Bradley., & Tibshirani, R. J. (1994). *An introduction to the bootstrap* (1st ed.). Chapman & Hall/CRC.

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization in Bayesian workflow. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, *182*(2), 389–402. https://doi.org/10.1111/rssa.12378

Gelman, A., Carlin, J. B., Stern, H., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian Data Analysis* (3rd ed.). CRC Press.

Gianattasio, K. Z., Wu, Q., Glymour, M. M., & Power, M. C. (2019). Comparison of Methods for Algorithmic Classification of Dementia Status in the Health and Retirement Study. *Epidemiology*, *30*(2), 291–302. https://doi.org/10.1097/EDE.0000000000000945

Gross, A. L., Hassenstab, J. J., Johnson, S. C., Clark, L. R., Resnick, S. M., Kitner-Triolo, M., Masters, C. L., Maruff, P., Morris, J. C., Soldan, A., Pettigrew, C., & Albert, M. S. (2017). A classification algorithm for predicting progression from normal cognition to mild cognitive impairment across five cohorts: The preclinical AD consortium. *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring*, *8*, 147–155. https://doi.org/10.1016/j.dadm.2017.05.003

Heeringa, S. G., Fisher, G. G., Hurd, M., Langa, K. M., Ofstedal, M. B., Plassman, B. L., Rodgers, W. L., & Weir, D. R. (2009). *Aging, Demographics and Memory Study (ADAMS) Sample Design, Weighting and Analysis for ADAMS*.

Kasper, J. D., Freedman, V. A., & Spillman, B. (2013). *Classification of Persons by Dementia Status in the National Health and Aging Trends Study*.

Knopman, D. S., Gottesman, R. F., Sharrett, A. R., Wruck, L. M., Windham, B. G., Coker, L., Schneider, A. L. C. C., Hengrui, S., Alonso, A., Coresh, J., Albert, M. S., Mosley, T. H., Richey Sharrett, A., Wruck, L. M., Windham, B. G., Coker, L., Schneider, A. L. C. C., Hengrui, S., Alonso, A., … Mosley Jr., T. H. (2016). Mild cognitive impairment and dementia prevalence: The Atherosclerosis Risk in Communities Neurocognitive Study. *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring*, *2*, 1–11. https://doi.org/10.1016/j.dadm.2015.12.002

Langa, K. M., Plassman, B. L., Wallace, R. B., Regula Herzog, A., Heeringa, S. G., Ofstedal, M. B., Burke, J. R., Fisher, G. G., Fultz, N. H., Hurd, M. D., Potter, G. G., Rodgers, W. L., Steffens, D. C., Weir, D. R., & Willis, R. J. (2005). The Aging, Demographics, and Memory Study: Study Design and Methods. *Neuroepidemiology*, *25*, 181–191. https://doi.org/10.1159/000087448

Langa, K. M., Ryan, L. H., McCammon, R., Jones, R. N., Manly, J. J., Levine, D. A., Sonnega, A., Farron, M., & Weir, D. R. (2020). The Health and Retirement Study Harmonized Cognitive Assessment Protocol Project: Study Design and Methods. *Neuroepidemiology*, *54*(1), 64–74. https://doi.org/10.1159/000503004

Little, R. J. A., & Schluchter, M. D. (1985). Maximum likelihood estimation for mixed continuous and categorical data with missing values. *Biometrika*, *72*(3), 497–512.

Liu, C., & Rubin, D. (1998). Ellipsoidally symmetric extensions of the general location model for mixed categorical and continuous data. *Biometrika*, *85*(3), 673–688. https://doi.org/10.1093/biomet/85.3.673

Lopez, O. L., Becker, J. T., Chang, Y.-F. F., Sweet, R. A., DeKosky, S. T., Gach, M. H., Carmichael, O. T., McDade, E., & Kuller, L. H. (2012). Incidence of mild cognitive impairment in the Pittsburgh Cardiovascular Health Study-Cognition Study. *Neurology*, *79*(15), 1599–1606. https://doi.org/10.1212/WNL.0b013e31826e25f0

Manly, J. J., Bell-McGinty, S., Tang, M.-X., Schupf, N., Stern, Y., & Mayeux, R. (2005). Implementing diagnostic criteria and estimating frequency of mild cognitive impairment in an urban community. *Archives of Neurology*, *62*(11), 1739–1746. https://doi.org/10.1001/archneur.62.11.1739

Mayeux, R., Reitz, C., Brickman, A. M., Haan, M. N., Manly, J. J., Maria Glymour, M., Weiss, C. C., Yaffe, K., Middleton, L., Hendrie, H. C., Warren, L. H., Hayden, K. M., Welsh-Bohmer, K. A., S Breitner, J. C., Morris, J. C., Bryan Alzheimer, K., & Knight Alzheimer, J. (2011). Operationalizing diagnostic criteria for Alzheimer's disease and other age-related cognitive impairment-Part 1. *Alzheimers Dement*, *7*(1), 15–34. https://doi.org/10.1016/j.jalz.2010.11.005

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., Klunk, W. E., Koroshetz, W. J., Manly, J. J., Mayeux, R., Mohs, R. C., Morris, J. C., Rossor, M. N., Scheltens, P., Carrillo, M. C., Thies, B., Weintraub, S., Phelps, C. H., Jack Jr., C. R., … Phelps, C. H. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, *7*(3), 263–269. https://doi.org/10.1016/j.jalz.2011.03.005

Olkin, I., & Tate, R. F. (1961). Multivariate Correlation Models with Mixed Discrete and Continuous Variables. *The Annals of Mathematical Statistics*, *32*(2), 448–465. https://doi.org/10.1214/aoms/1177705052

Olsen, M. K., & Schafer, J. L. (2001). A two-part random-effects model for semicontinuous longitudinal data. *Journal of the American Statistical Association*, *96*(454), 730–745. https://doi.org/10.1198/016214501753168389

Philipps, V., Amieva, H., Andrieu, S., Dufouil, C., Berr, C., Dartigues, J.-F., Jacqmin-Gadda, H., & Proust-Lima, C. (2014). Normalized Mini-Mental State Examination for Assessing Cognitive Change in Population-Based Brain Aging Studies. *Neuroepidemiology*, *43*, 15–25. https://doi.org/10.1159/000365637

Plassman, B. L., Langa, K. M., Fisher, G. G., Heeringa, S. G., Weir, D. R., Ofstedal, M. B., Burke, J. R., Hurd, M. D., Potter, G. G., Rodgers, W. L., Steffens, D. C., Willis, R. J., & Wallace, R. B. (2007). Prevalence of Dementia in the United States: The Aging, Demographics, and Memory Study. *Neuroepidemiology*, *29*, 125–132. https://doi.org/10.1159/000109998

Prina, A. M., Mayston, R., Wu, Y.-T., & Prince, M. (2019). A review of the 10/66 dementia research group. *Social Psychiatry and Psychiatric Epidemiology*, *54*, 1–10. https://doi.org/10.1007/s00127-018-1626-7

Proust-Lima, C., & Philipps, V. (2018). *NormPsy: Normalisation of Psychometric Tests* (R package version 1.0.8).

R Core Team. (2020). *R: A Language and Environment for Statistical Computing* (4.0.0). R Foundation for Statistical Computing.

Rubin, D. B. (1981). The Bayesian Bootstrap. *The Annals of Statistics*, *9*(1), 130–134. https://doi.org/10.1214/aos/1176345338

Rubin, D. B. (1996). Multiple Imputation after 18+ Years. *Journal of the American Statistical Association*, *91*, 473–489. https://doi.org/10.1080/01621459.1996.10476908

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data* (D. Cox, V. Isham, N. Keiding, N. Reid, & H. Tong, Eds.; 1st ed.). Chapman & Hall.

Schafer, J. L. (1999). Multiple imputation: a primer. *Stat Methods Med Res*, *8*(1), 3–15. https://doi.org/10.1191/096228099671525676

Sonnega, A., Faul, J. D., Ofstedal, M. B., Langa, K. M., Phillips, J. W., & Weir, D. R. (2014). Cohort Profile: the Health and Retirement Study (HRS). *International Journal of Epidemiology*, *43*(2), 576–585. https://doi.org/10.1093/ije/dyu067

Stef van Buuren. (2019). *Flexible Imputation of Missing Data, Second Edition*.

Trittschuh, E. H., Crane, P. K., Larson, E. B., Cholerton, B., Mccormick, W. C., McCurry, S. M., Bowen, J. D., Baker, L. D., & Craft, S. (2011). Effects of Varying Diagnostic Criteria on Prevalance of Mild Cognitive Impairment in a Community Based Sample. *J Alzheimers Dis*, *25*(1), 163–173. https://doi.org/10.1161/CIRCULATIONAHA.110.956839

Weir, D. R., Langa, K. M., & Ryan, L. H. (2016). *Harmonized Cognitive Assessment Protocol (HCAP): Study Protocol Summary*.

Wilson, R. S., Aggarwal, N. T., Barnes, L. L., Mendes De Leon, C. F., Hebert, L. E., & Evans, D. A. (2010). Cognitive decline in incident Alzheimer disease in a community population. *Neurology*, *74*(7), 951–955. https://doi.org/10.1212/WNL.0b013e3181ec684c

Zhang, A., Brown, L. D., & Tony Cai, T. (2019). Semi-supervised inference: General theory and estimation of means. *Annals of Statistics*, *47*(5), 2538–2566. https://doi.org/10.1214/18-AOS1756

Appendix

Table A1: Candidate variables for inclusion in multi-part models of predicted ADAMS impairment classes (unimpaired, other, MCI, dementia).

Sociodemographic Characteristics	*Neuropsychological Exam and Cognition*	*Health and Health Behaviors*
Age	Total MMSE‡ Score	Stroke History (yes/no)
Sex/Gender	Backwards Count (20, 86)	Hypertension (yes/no)
Race/Ethnicity	Serial 7s	Diabetes (yes/no)
Education	Item Naming (scissors, cactus)	CVD (yes/no)
Marital Status	President/VP Naming	BMI
Retirement Status	Animal Naming	IADLs#
	Boston Naming Test	ADLs
	Word Recall (Immediate, Delayed)	Depression
	Word list recall (Yes, No)	Smoking
	Story Recall (Immediate, Delayed)	Alcohol Use
	Constructional Praxis (Immediate, Delayed)	
	Symbol/Digit Substitution	
	Trails (A, B)	
	Subjective Change in Cognition	
	Average Proxy Cognition	

‡MMSE: Mini-mental state exam
#IADL: Instrumental Activities of Daily Living