Recognition Memory Deficits in Alzheimer's Disease: Modeling Clinical Groups and Individual Patients

James P. Pooley (jpooley@uci.edu)
Michael D. Lee (mdlee@uci.edu)
Department of Cognitive Sciences, 3151 Social Sciences Plaza A
University of California, Irvine, CA, 92697-5100

William R. Shankle (rshankle@mccare.com)
Medical Care Corporation, 19782 MacArthur Boulevard
Irvine, CA 92612

Abstract

Recognition memory tests are useful for understanding Alzheimer's disease (AD). In the clinical context, it is important to model performance at both the group level (e.g., for the characterization of clinical subpopulations) and individual level (e.g., for the diagnosis of a patient). Using a clinical data set from AD patients, we show how a signal detection theory model that assumes hierarchical individual differences in discriminability and response bias adequately describes these data at both the group and individual levels, and also present preliminary descriptive and predictive analyses of the data at both levels.

Keywords: Alzheimer's disease; Hierarchical Bayesian modeling; Cognitive psychometrics; Signal detection theory; Recognition memory.

Introduction

Severe degradation of episodic memory is the hallmark behavioral symptom of Alzheimer's disease (AD). Differences in the severity of this mnemonic degradation best distinguish adults who are aging normally from those affected with AD on the basis of behavior alone (Locascio, Growdon, & Corkin, 1995). Given this diagnostic power, it is not surprising that most neuropsychological testing for AD has tended to focus on episodic memory. Episodic memory may be measured using a variety of experimental paradigms, but the simplest approach is to use a yes/no recognition memory test.

In a yes/no recognition memory test, individuals are shown a study list of items to memorize and are then tested on their ability to discriminate these studied *target* items from unstudied *distractor* items on subsequent test lists. On each trial of the experiment, an individual is shown an item and simply indicates whether or not the item was on the study list. These responses fall into four classes: hits, misses, false alarms, and correct rejections. Based on counts of these responses, performance on yes/no recognition memory tests is often measured using signal detection theory (SDT) models.

SDT (Macmillan & Creelman, 2005) provides a general framework for understanding how the variability in the memory representations of target and distractor items interacts with a cognitive decision process to affect performance on recognition memory tests. Use of SDT models to characterize the results of recognition memory experiments is common in cognitive psychology, where the goal is often to model the group performance of cognitively normal individuals. In contrast, SDT models are used less often in AD research, where it is important to model performance at both the group level (e.g., for the characterization of clinical subpopulations) and the individual level (e.g., for the diagnosis of a patient).

The goal of this paper is to use SDT to model recognition memory performance of AD patients at both the group and individual levels. In the next section, we describe the SDT model in more detail. We then describe new clinical data, and evaluate three SDT approaches to modeling the individual patients and clinical groups in these data. These models make different assumptions about individual differences, and we show that only a hierarchical model is satisfactory. Finally, we use this hierarchical model to present some first analyses of the clinical data in terms of both the memory characteristics of groups of patients with different levels of cognitive impairment, and in a predictive test of individual patient diagnosis.

Signal Detection Theory

SDT models are often used in cognitive psychology as simple models of how individuals make decisions. Under this view, the parameters of the SDT model have psychological interpretations that give insight into the underlying memory and decision processes involved in a recognition task.

The basic SDT model shown in Figure 1 assumes that, on each trial, the presented item evokes some latent memory strength. The memory strengths of both target and distractor items are assumed to have Gaussian distributions, with the mean of the distractor distribution separated from the mean of the target distribution by a distance d'. In this way, d' measures the *discriminability* of the target and distractor items, and so represents the acuity of memory for the items.

Due to the assumed overlap of the target and distractor distributions, an individual needs a decision strategy for relating memory strength to responses in the recognition test. SDT models assume this is done using a criterion level of memory strength k below which the individual will respond 'new' and above which the individual will respond 'old'. The area h under the target distribution above the criterion corresponds to the hit rate, and the
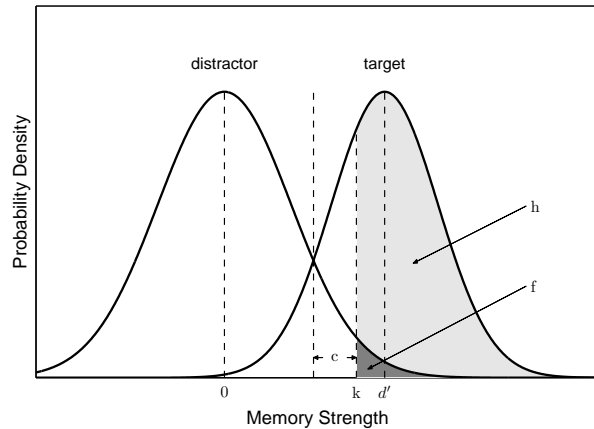
Figure 1: The unequal variance signal detection theory model and parameters.

area f under the distractor distribution above the criterion corresponds to the false alarm rate.

The distance c between this criterion and unbiased responding is commonly used as a measure of *response bias* due to its independece of d' (Snodgrass & Corwin, 1988). The response bias measures the tendency of an individual to give one response rather than another.

Most SDT models in psychology assume that the standard deviations of the target and distractor distributions are equal. Results of recognition memory experiments (e.g., Mickes, Wixted, & Wais, 2007), however, support a version of SDT in which the standard deviation of the target distribution is 25% larger than the standard deviation of the distractor distribution, so that $\tau = \sigma_{\text{distractor}}/\sigma_{\text{target}} = 0.8$.

Clinical Data Set

Our data come from a neurology clinic where 533 patients completed a yes/no recognition memory test consisting of 10 target and 10 distractor words selected from the CERAD word list (e.g., Shankle et al., 2005). The majority of these patients had AD, vascular dementia, lewy body disease, or a mixture of these diagnoses.

Independent of patient performance on the recognition memory test, a trained neurologist used the functional assessment staging test (FAST) to classify each patient. The FAST (Reisberg, 1988) is a well-validated diagnostic tool used by clinicians to classify patients into one of seven stages, each of which corresponds to a level of functional impairment. Specifically, progressing from stage 1 to stage 7 corresponds to 'normal aging', 'possible mild cognitive impairment', 'mild cognitive impairment', 'mild dementia', 'moderate dementia', 'moderately severe dementia' and 'severe dementia'.

Of the 533 patients, 145 were judged to be in FAST stage 1, 93 in stage 2, 96 in stage 3, 131 in stage 4, 46

in stage 5, and 22 in stage 6.[1] In this data set, variability on at least two qualitatively distinct levels is expected. At an individual level, each patient is expected to differ in their mnemonic ability; at a group level, patients with a given FAST stage are expected, on average, to differ in mnemonic ability from patients with a different FAST stage. The first row of Figure 3 summarizes patient performance , with the panels corresponding to the six stages. Each point corresponds to a patient, showing their combination of hits and false-alarms in the task. It can be seen that, generally, performance worsens (i.e., there are fewer hits and/or more false-alarms) with increasing stage, but also that there remains considerable variability across patients within the same stage. This means any attempt to model these data should account for this systematic variation both between groups (i.e., how the funtional stages differ in terms of memory and decision strategy) and within groups (i.e., how individual patients differ in their memory and decision strategy).

Three SDT Models

In this section, we evaluate the ability of three SDT models to account for our clinical data. Although the three models share the basic SDT assumptions, each makes different assumptions regarding individual differences in discriminability and response bias.

In order to perform Bayesian inference, we implemented the SDT models as probabilistic graphical models in WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000), software that uses a variety of Markov chain Monte Carlo (MCMC) algorithms (Gilks, Richardson, & Spiegelhalter, 1996) to obtain samples from the joint posterior distributions of these parameters. Probabilistic graphical models are widely used for Bayesian inference within statistics (e.g., Jordan, 2004) and more recently in cognitive science (Lee, 2008; Shiffrin, Lee, Kim, & Wagenmakers, 2008), including in the current context of recognition memory and SDT (Dennis, Lee, & Kinnell, 2008). All of our analyses used 2000 posterior samples collected following a burn-in (i.e., a set of samples discarded to ensure that the recorded samples are representative of the posterior distribution) of 1000 samples.

No Individual Differences

Our first attempt to account for the clinical data uses an SDT model in which all patients with a given FAST stage have the same discriminability and response bias. This approach is seen in the cognitive modeling literature when SDT is fit to averaged or aggregated data, estimating one set of parameters to describe all the individuals in a group.

Graphical Model The graphical model is shown in the left panel of Figure 2. Consistent with the assumption that all patients with a given FAST stage have the

---

[1]Since patients with a diagnosis of FAST stage 7 can, at best, speak approximately 5 or 6 words per day and, at worst, cannot lift their head, no data from patients with this diagnosis were included in our data set.
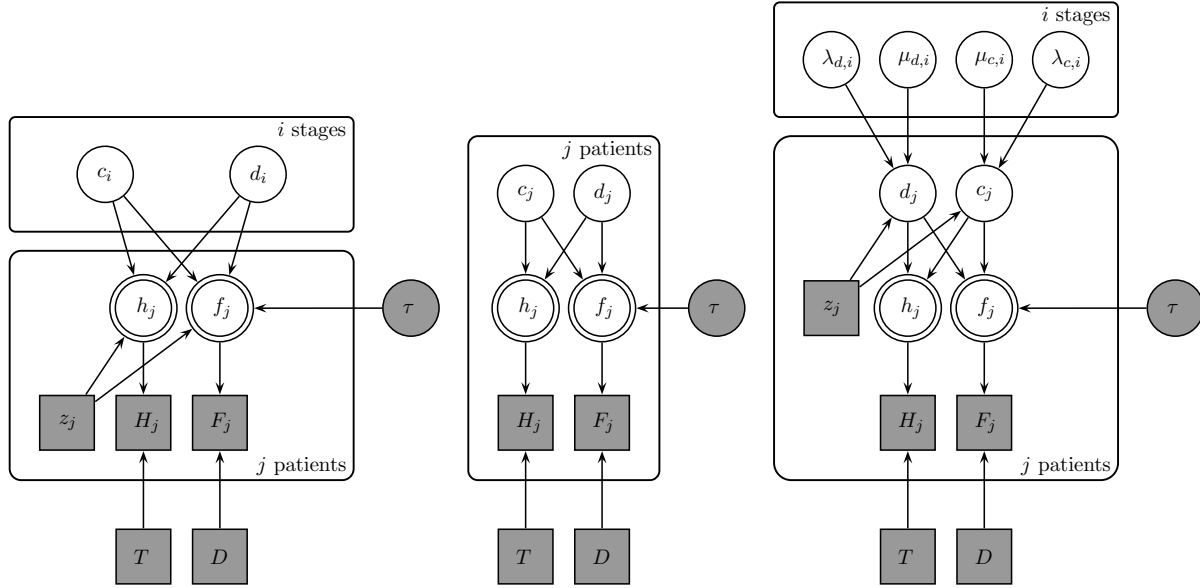
Figure 2: Graphical models for the SDT models that assume no (left), full (middle), and hierarchical (right) individual differences in discriminability and response bias.

same discriminability and response bias, the plate with $i = 1, \ldots, 6$ corresponds to the six FAST stages. Within this plate, the d_i and c_i nodes correspond to the discriminability and response bias, respectively, for the ith FAST stage. The plate with $j = 1, \ldots, 533$ corresponds to the 533 patients. Within this plate, the z_j node indicates the FAST stage of the jth patient, which via SDT determines the hit and false alarm rates h_j and f_j for that patient. Formally, $h_j = \Phi(d_{z_j}/2 - c_{z_j})$ and $f_j = \Phi(-(d_{z_j}/2 + c_{z_j})/\tau)$, where τ was set to the empirical value of 0.8. Based on the hit and false alarm rates and the $T = 10$ target and $D = 10$ distractor words presented to all patients during the recognition tests, the jth patient produces $H_j \sim \text{Binomial}(h_j, T)$ hits and $F_j \sim \text{Binomial}(f_j, D)$ observed false alarms during the test. We used the standard non-informative priors $d_j \sim \text{Gaussian}(0, 2)$ and $c_j \sim \text{Gaussian}(0, 1/2)$.

Results   We use posterior predictive distributions as a standard Bayesian assessment the fit of models to data (e.g., Gelman, Carlin, Stern, & Rubin, 2004). These distributions for hits and false alarm counts are shown in the second row of Figures 3 and 4, at the group and individual level respectively. In each panel, the box sizes are proportional to the mass of the posterior predictive distribution. The second row of Figure 3 shows that the group level model only adequately accounts for the variability of FAST stage 1, perhaps stage 2, but no other stages. The second row of Figure 4 shows the posterior predictive distribution for a randomly selected individual patient from each FAST stage. In all but one case, the mass of the posterior predictive distributions does not contact the patient's data, represented by the cross. From

this, we conclude that the SDT model with no individual differences does not adequately describe the data.

Full Individual Differences

Our second attempt to account for the clinical data uses an SDT model in which each patient is assumed to have a different level of mnemonic ability. This approach is seen in the cognitive modeling literature when SDT is fit to individual participant data, and parameters are estimated for each separately.

Graphical Model   The graphical model is shown in the middle panel of Figure 2. Since each patient is assumed to have a unique discriminability and response bias (and consequently hit and false alarm rates), the plate corresponding to the patients now extends over the corresponding d_j, c_j, h_j, and f_j nodes. These variables have the same definitions and priors as before, but now applied at the level of individuals rather than stages.

Results   Posterior predictive distributions for the group analysis of the SDT model with full individual differences are shown in the third row of Figure 3. Each of these distributions puts roughly equal mass on each possible pair of hit and false alarm counts, which does not match the observed data. The basic problem is that, by fitting at an individual level, the model cannot make inferences about groups of participants. Where the model does fare well is at the individual level, shown in the third row of Figure 4. Here, the posterior predictive fits to the selected patients are excellent, as they should be. Nevertheless, the inability of the SDT model with full individual differences to describe the FAST groups means that it does not satisfy the dual goals we set at the outset.
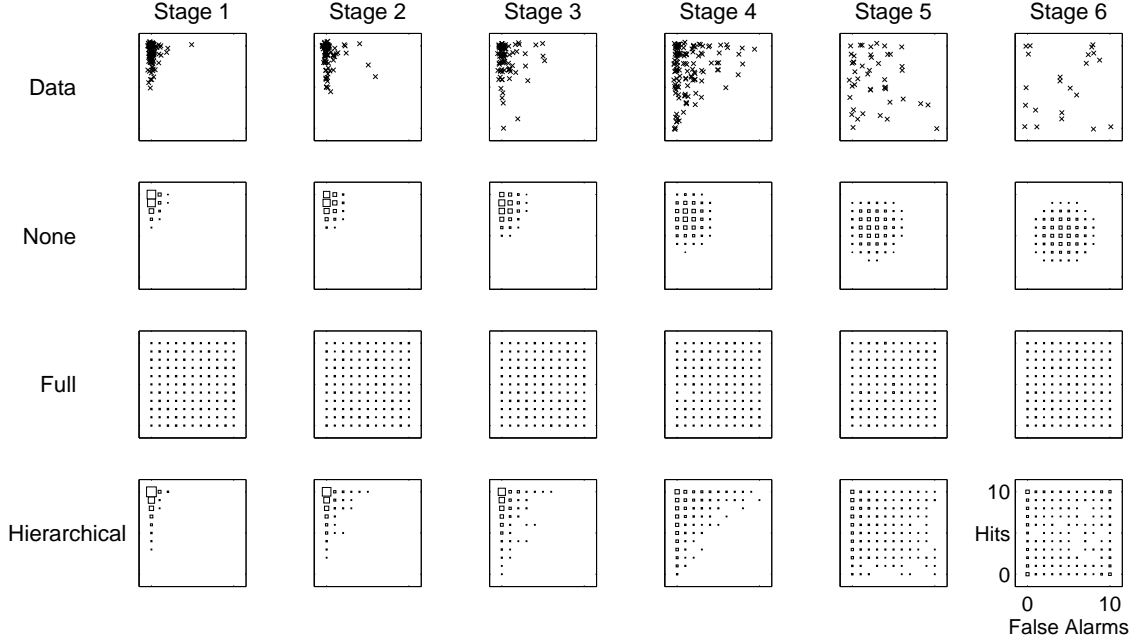
Figure 3: The first row shows the hit and false-alarm counts for each patient, according to their FAST stage. The second, third, and fourth rows show the posterior predictive distributions for hit and false alarm counts for the no difference, full difference, and hierarchical difference models. In the posterior predictive panels, the box sizes are proportional to the mass of the posterior predictive distribution for that combination of hits and false-alarms.

Hierarchical Individual Differences

Our third approach uses an SDT model in which each patient is assumed to have different discriminability and response bias, depending on their FAST stage. We do this by assuming a structure to the individual differences, using a hierarchical model.

Graphical Model The graphical model is shown in the right panel of Figure 2. Since each patient has a unique discriminability and response bias, the plate corresponding to the patients still extends over the corresponding d_j and c_j nodes. However, we now assume that these values are drawn from Gaussian distributions, where the discriminability and response bias distributions for the ith FAST stage have means $\mu_{d,i}$ and $\mu_{c,i}$ and precisions $\lambda_{d,i}$ and $\lambda_{c,i}$. Formally, for example, $d_j \sim \text{Gaussian}(\mu_{d,z_j}, \lambda_{d,z_j})$.

Results Posterior predictive distributions for the group analysis for the SDT model with hierarchical individual differences are shown in the fourth row of Figure 3. In contrast to those produced by the two previous SDT models, these distributions provide a good description of the data. Similarly, the fourth row of Figure 4 shows that the posterior predictive distribution for the randomly selected patients are as good as those obtained using the SDT model with full individual differences. From this, we conclude that the SDT model with hierarchical individual differences does adequately describe the data at both the group and individual level.

This descriptive adequacy makes it sensible to examine the parameter inferences made by the model. We present two analyses of this type, one at the group and one at the individual level, to demonstrate what sort of analyses are possible. At the group level, Figure 5 shows scatterplots of 50 randomly sampled pairs of mean discriminability μ_d and mean response bias μ_c values from the joint posterior distribution over these parameters for each FAST stage. The FAST stage groups are separated with respect to mean discriminability, with progressive stages having lower values. In contrast, the FAST stage groups have roughly equal levels of response bias with only a slight decease evident as the stages progress.

At the individual level, we evaluated the adequacy of the hierarchical model to predict the FAST stage of patients based on their test performance. To accomplish this, we used data from 483 of the 533 patients to obtain a posterior distribution over the discriminability and response bias parameters.[2] We then found the posterior distribution for the FAST stage variables z_j of those patients whose true FAST stage was withheld, and made the predicted staging corresponding to the mode of that distribution (i.e., the standard MAP estimate).

---

[2]Many of the 533 patients participated in the same recognition memory test multiple times. Only data from their first visit was used for the posterior predictive analyses. For the prediction analysis, however, data from subsequent visits on which a patient did not change FAST stage were combined.
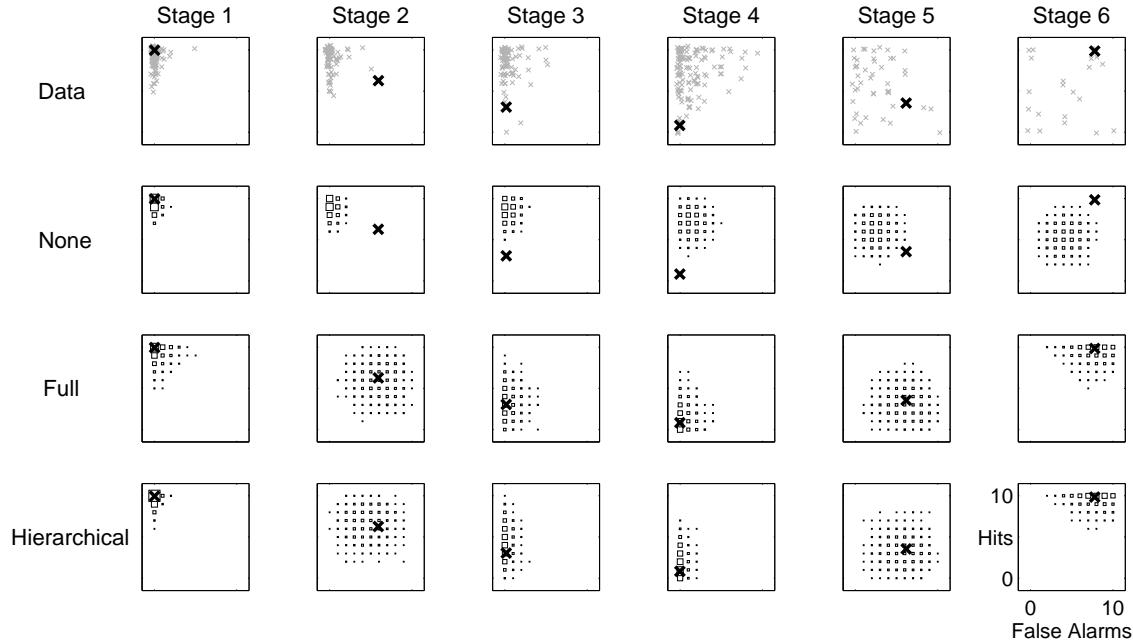
Figure 4: Posterior predictive distributions for hit and false alarm counts for one randomly selected patient from each of the first six FAST stages.

Figure 6 presents the results of this analysis, showing how the model predictions relate to the true diagnoses. Each box corresponds to a (truth,prediction) pair, and the box size is proportional to the frequency of that pairing. Black boxes on the diagonal are correct classifications. The gray outline regions in Figure 6 correspond to the broader classsification dividing FAST stages 1 and 2, which essentially represent normal cognitive functioning, from stages 3–6, which represent cognitive impairment with or without dementia. It can be seen that the predictions of the model are generally good, especially at the broader level, but are certainly not perfect. Our main point is that it is straightforward to make predictions for individuals by assuming hierarchal individual differences, and that these predictions are informed by the different group characteristics observed in Figure 5.

Discussion

We think the hierarchical approach to modeling groups and individuals is an important and useful one. The no individual differences model failed to account for data at both the the group and individual levels, because it failed to provide a mechanism to deal with the variability that existed within a given FAST stage. The full individual differences model failed to account for data at the group level, because it had no representation of this group level. In contrast,the hierarchical individual differences model was able to account for data at both levels successfully. While it may not be surprising that this model was able to account for data at the group level better than the other two models, its ability to account for the data at the indi-

vidual level as well as the SDT model with full individual differences may strike some as a surprising.

Previous work using SDT to model data from patients with dementia (Snodgrass & Corwin, 1988). In this work, a basic SDT model (assuming equal variances) was used to model recognition memory data from dementia patients (diagnosed with either AD or Parkinson's disease). One of the main results of this study was that patients with dementia have abnormally liberal response biases (as measured by c). While Figure 5 appears to show a decreasing trend, there is no evidence of the liberal criterion values found in the previous work.

In our hierarchical SDT model, we assumed differences between *individuals*. One potential limitation is that we did not attempt to account for differences between *items*. Some authors (e.g., Rouder & Lu, 2005) have noted that SDT parameters are systematically underestimated when item variability is not accounted for. Adding such an assumption to the hierarchical model produced here is straightforward, and worth pursuing.

A final issue concerns the performance of the hierarchical SDT model on the prediction task. While the model's performance was imperfect, its predictions (both correct and incorrect) seem sensible. As the first row of Figures 3 and 4 show, a large number of patients with FAST stages 4 (and even some with stages 5 and 6) still perform perfectly on the test. This may tell us more about the utility of recognition memory tests as diagnostic tools for AD, rather than about any fundamental deficits in the hierarchical SDT model. Given these results, modeling data from another recognition or recall paradigm
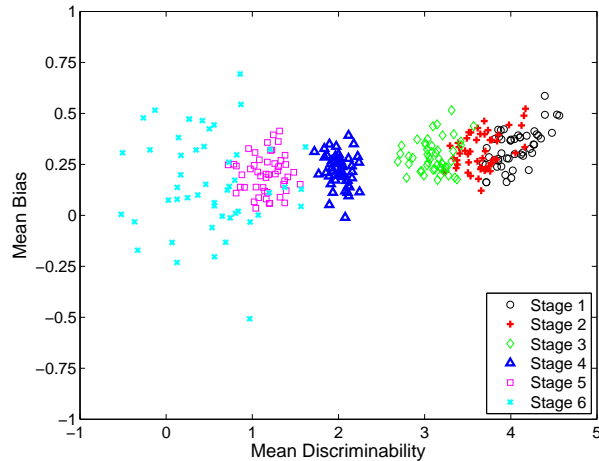
Figure 5: Joint posterior for the means of the discriminability and response bias parameters of the SDT model that assumes hierarchical individual differences.

Figure 6: FAST stage prediction performance.

could prove useful in clinical settings (cf. Locascio et al., 1995).

In conclusion, we have demonstrated, using a large clinical data set, how an SDT model with hierarchical individual differences in discriminability and response bias is able to solve a problem that appears in clinical research: modeling patient performance at both the group and individual levels. Considerable work needs to be done to extend the approach to offer insights into the nature of episodic memory deficits in AD, but we feel that the general approach of applying memory models from cognitive psychology to gain insights into AD holds great promise.

Acknowledgments

References

Dennis, S. J., Lee, M. D., & Kinnell, A. (2008). Bayesian analysis of recognition memory: The case of the list-length effect. *Journal of Memory and Language*, *59*, 361–376.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis*. Boca Raton, FL: Chapman & Hall/CRC.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. E. (1996). *Markov Chain Monte Carlo in Practice*. Boca Raton, FL: Chapman & Hall/CRC.

Jordan, M. I. (2004). Graphical models. *Statistical Science*, *19*, 140–155.

Lee, M. D. (2008). Three case studies in the Bayesian analysis of cognitive models. *Psychonomic Bulletin & Review*, *15*, 1–15.
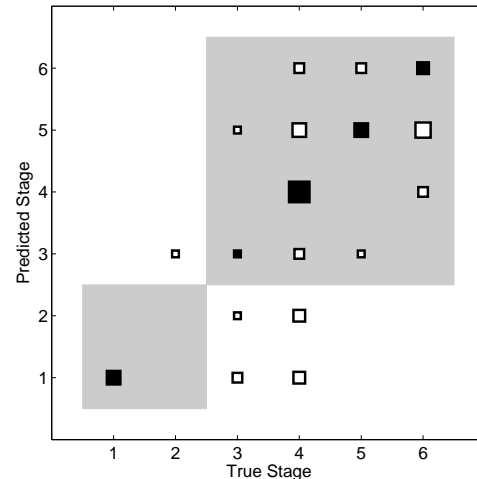
Locascio, J. J., Growdon, J. H., & Corkin, S. (1995). Cognitive test performance in detecting, staging, and tracking Alzheimer's disease. *Archives of Neurology*, *52*, 1087–1099.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS - A Bayesian modeling framework: Concepts, structure, and extensibility. *Statistics and Computing*, *10*, 325–337.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (Second ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Mickes, L., Wixted, J. T., & Wais, P. E. (2007). A direct test of the unequal-variance signal-detection model of recognition memory. *Psychonomic Bulletin & Review*, *14*, 858–865.

Reisberg, B. (1988). Functional assessment staging (FAST). *Psychopharmacology Bulletin*, *24*, 653–659.

Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, *12*, 573–604.

Shankle, W. R., Romney, A. K., Hara, J., Fortier, D., Dick, M. B., Chen, J. M., et al. (2005). Methods to improve the detection of mild cognitive impairment. *Proceedings of the National Academy of Sciences*, *102*(13), 4919–4924.

Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E.-J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science*, *32*, 1248–1284.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*, 34–50.