Published in final edited form as:

Alzheimers Dement. 2006 October ; 2(4): 303–313. doi:10.1016/j.jalz.2006.08.001.

Independent contributions of neural and “higher-order” deficits to symptoms in Alzheimer’s disease: A latent variable modeling approach

Rochelle E. Tractenberg^{a,b,*}, Paul S. Aisen^a, Myron F. Weiner^c, Jeffrey L. Cummings^d, and Gregory R. Hancock^e

^a Department of Neurology; Georgetown University School of Medicine, 291 Building D, 4000 Reservoir Road, NW, Washington, DC 20057, USA

^b Departments of Biostatistics, Bioinformatics & Biomathematics, and Psychiatry, Georgetown University School of Medicine, Washington, DC, USA

^c Department of Psychiatry, University of Texas Southwestern Medical Center, Dallas, TX, USA

^d Department of Neurology, University of California, Los Angeles, Los Angeles, CA, USA

^e Center for Integrated Latent Variable Research and Department of Education, Measurement, Statistics and Evaluation, University of Maryland, College Park, MD, USA

Abstract

Background—Analytic models of Alzheimer’s disease (AD) tend to focus on one type of symptom and assume implicitly that no measurement error is present. These tendencies render changes in symptom domains difficult to model mathematically, although latent variable methods can accommodate both multiple symptom domains and error. This study formulated and compared underlying (latent) factor structures representing previously reported dependence and independence of symptoms of cognitive decline, functional impairment, and behavioral disturbance in AD.

Methods—In confirmatory factor analyses of data from 2 cohorts of AD patients, 2 levels of latent variables were conceptualized. One general neurologic factor represented disease, and symptom factors represented cognition, function, and behavior. Two “null” models had either a single factor or 3 symptom factors. Two 2-level models treated the general factor as underlying both the observed variables and the symptom factors or treated the symptom factors as explaining variability in the observed variables after taking the general factor into account (“residualized”).

Results/Conclusions—The residualized model fit the data in both cohorts significantly better than the other models, and relations in this model between some observed and latent variables were different across cohorts. Neither cohort supported a single factor model; both cohorts independently supported a residualized model that may permit differentiation of symptom- from disease-modifying effects of treatment.

Keywords

Factor analysis; Statistical; Statistics; Data interpretation; Statistical; Activities of daily living; Cognition disorders; Behavior disorders; Geriatric assessment

1. Introduction

Alzheimer's disease (AD) is a prevalent age-associated brain disease characterized by neuropathologic lesions (amyloid plaques and neurofibrillary tangles) that vary in quantity and distribution, loss of neurons and synapses, and decline in neurotransmitter systems. Clinical manifestations include impairment in multiple cognitive domains, functional decline, and behavioral disturbances. The neuropathologic changes cannot be assessed accurately before death, and the variable and diverse symptoms are evaluated with imperfect assessment tools. Characterization of neuropathologic and clinical status is thus problematic, and current methods of tracking progression and gauging the efficacy of treatments are flawed.

Alois Alzheimer's description in 1907 of the patient with the disease later to be named for him included behavioral and psychotic symptoms along with memory loss [1]. However, more recent work on AD, including clinical trials, has emphasized cognitive decline over loss of functional ability and behavioral symptoms such as psychosis, agitation, and depression [2–4], a conceptualization that may limit our ability to derive a satisfactory statistical model of the trajectory of illness in AD [4–10]. In addition to the tendency to focus on just one type of symptom, models of progression in AD also assume implicitly that the instruments used have no measurement error associated with them [11]; these factors together render changes in symptom domains difficult to model mathematically; these difficulties are further compounded by issues concerning missing data and the heterogeneity of both the presentation of AD [12,13] and its progression [14–16].

We have reported previously statistical evidence of partially independent changes over 12 months in cognition, daily function, and behavior in persons with AD [17] using a latent variable method, exploratory factor analysis. This work focused on more than one symptom type, and the assumption of an underlying (latent) factor (or set of factors) accounted for the possibility of measurement error in a way that is not reported generally in AD research. The results suggested that the best model of AD and its progression should incorporate the symptom domains of cognition, behavior, and function. The current study seeks to replicate our earlier findings of relative independence of symptom domains, and extend them by modeling 2 independent samples of AD patients to derive and replicate a cross-sectional conceptualization of the interrelatedness of these 3 types of symptoms. Such a model may improve our understanding of AD, subsequently improving the design of future descriptive and therapeutic studies.

Our earlier work explored relations between changes in symptoms across the 3 domains with correlations [18] and exploratory factor analysis [17]. This study uses methods from structural equation modeling, which emphasizes latent variable modeling of unobservable constructs [19], described below. The Alzheimer's Disease Cooperative Study (ADCS) [20] recently completed 2 multicenter treatment trials for AD. Each study used instruments that specifically monitored cognition, daily function, and behavior, but instruments differed between the 2 studies, so we repeated the modeling for each cohort separately. Because our purpose was to examine the underlying structure of the symptoms and not to generate estimates of strengths of relationships, the difference in instrumentation across cohorts was not a concern in modeling or interpreting our models' respective fits to the data.

2. Methods

This study is a secondary analysis of data collected in 2 National Institutes of Health (NIH)-funded, multicenter, placebo-controlled clinical trials evaluating the effects of anti-

inflammatory drugs on cognitive decline in AD. Both studies were approved by the institutional review boards of all participating centers.

2.1. Prednisone study

Aisen et al. [21] conducted a multicenter, randomized, placebo-controlled clinical trial to determine the effect of steroid anti-inflammatory drug prednisone (PR) on cognitive decline in 138 persons who met National Institute of Neurological and Communicative Diseases and Stroke/Alzheimer Disease and Related Disorders Association (NINCDS-ADRDA) [22] criteria for probable AD at a mild to moderate stage. Use of cholinesterase inhibitors was not allowed. Eligible subjects were medically stable and free of conditions that might respond to glucocorticoid treatment (e.g., rheumatoid arthritis) or increase the risk of such treatment (e.g., diabetes). Study participants had a mean age of 73 years, 50% were women.

2.2. Nonsteroidal study

Aisen et al. [23] conducted a multicenter, randomized, placebo-controlled clinical trial to determine the effect of the nonsteroidal anti-inflammatory drug (NSAID), rofecoxib; a selective cyclooxygenase-2 (COX-2) inhibitor; and naproxen (a nonselective NSAID) on cognitive decline in 351 persons with NINCDS-ADRDA diagnoses of probable AD at a mild to moderate stage. Eligible subjects were medically stable and free of conditions, such as active peptic ulcer disease, that might increase the risk of NSAID treatment. At baseline, 68% of this sample was on a stable dose of a cholinesterase inhibitor. Study participants had a mean age of 74 years; 53% were women.

2.3. Materials

2.3.1. Cognition—Both cohorts had the same 3 cognitive measures administered at baseline. The primary outcome measure in both trials was the Alzheimer's Disease Assessment Scale—Cognitive (ADAS-Cog) [24,25]. The ADAS-Cog, designed specifically for use with persons with AD, measures memory, attention, reasoning, language, orientation, and praxis. Scores range from 0 to 76 with higher scores reflecting higher impairment.

The Clinical Dementia Rating (CDR) [26] was a secondary outcome measure in both studies. The CDR is based on a semistructured interview of both the patient and a knowledgeable informant, assessing/grading overall dementia severity. The instrument rates each of 6 domains: memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care. Each domain ("box") is rated as normal (0) or as involving questionable (0.5), mild (1), moderate (2), or severe (3) impairment. The CDR global rating (0 to 3) describes the severity of dementia, and CDR ≤ 2 was 1 inclusion criterion for both cohorts. Tractenberg et al. [17] found that 2 sums of box scores could be calculated to generate scores in the cognitive (memory, orientation, judgment and problem solving) and functional (community affairs, home and hobbies, personal care) domains. To maximize information and meet identification requirements [19] for the latent variable modeling technique used in this study, the box scores representing these 2 domains at baseline were used as individual indicators. Higher box scores reflect higher impairment in terms of cognition (memory [C1], orientation [C2], judgment and problem solving [C3]) and function (community affairs [F1], home and hobbies [F2], personal care [F3]), respectively.

The Mini Mental State Examination (MMSE) [27] was part of the screening process for both studies. MMSE scores are calculated as the sum of 1/0 correct/incorrect answers to 30 questions, so scores range from 0 to 30 (worst to best).

2.3.2. Function—Both cohorts had 2 functional outcome measures administered at baseline; the CDR box scores representing function (F1–F3, see above) is common to both

studies. Although the second functional measure differed in the studies, both focused on the level of dependency and contained common items. These were scored similarly, as follows.

In the prednisone (PR) study, function was also assessed with the Dependency Scale (DS) [28]. The DS assesses various physical activities of daily living including eating, dressing, and toileting. The total number of “points” (higher = worse function) was converted to the percentage of domains performed independently, so that higher percentages reflected better performance. To put the DS-derived function score on the same scale as the CDR box scores, 100% minus this percentage was calculated so that for both measures of function in the PR cohort, higher scores reflected poorer function.

In the nonsteroidal (NS) study, function was also assessed with an Activities of Daily Living Inventory (ADL) [29]. This instrument measures the level of assistance required to carry out 23 ADLs and instrumental ADLs (IADLs). This instrument includes all the activities that the DS includes, plus additional activities. Similar to our approach with the DS, these ADL scores were converted to the same percentage of independence, which was subtracted subsequently from 100% to match both the CDR-derived and DS scoring; thus, in all cases, higher functional “scores” reflected worse performance, and both cohorts had 4 functional indicators.

2.3.3. Behavior—The behavioral instruments were different for both studies. Both cohorts had one indicator of general behavioral disturbance. One study had a formal assessment of depression, whereas the other had item-level ratings of depressive and apathetic symptoms that were matched to the formal depression instrument to derive a specific depressive-symptoms indicator for that cohort; because the total scores on the general behavior assessments were not the simple sum of the component symptoms, 2 new specific indicator variables, reflecting psychosis and agitation, were also derived, as described below. Thus, a total of 4 indicator variables reflecting general disturbance plus 3 key behavioral symptom types in AD were derived for each cohort.

In the PR study, 2 measures of behavioral symptoms were used. General behavioral symptoms were assessed with the Brief Psychiatric Rating Scale (BPRS) [30]. The BPRS contains 22 items that assesses symptoms of anxiety, motor restlessness/agitation, unnatural motor behavior, grandiosity, depression, hostility, suspiciousness and delusions, hallucinations, uncooperativeness, blunted affect, hypomania and lability, and elation. The BPRS ratings are based either on the rater’s perception during the interview (e.g., uncooperativeness) or on the patient’s report of feelings in the past week.

In the PR study, depression was assessed with the Hamilton Rating Scale for Depression (HRSD) [31]. This scale is administered by clinically trained staff who rates each of 17 symptoms of depression including insomnia, loss of interest in work/activities, motor agitation, anxiety, loss of appetite, and loss of weight. The total score for the HRSD is a simple sum of symptom frequency ratings, whereas the BPRS total score is derived as a function of observed symptoms and their frequency ratings (i.e., not a simple sum of ratings). The total BPRS score was used as an indicator of “general behavioral disturbance” in the PR cohort, whereas indicators of agitation and psychosis were derived as simple sums of frequency ratings for uncooperativeness, lability, motor restlessness/agitation, unnatural motor behavior and hostility (agitation), grandiosity, suspiciousness and delusions, and hallucinations (psychosis).

In the NS study, behavioral symptoms were assessed using the Neuropsychiatric Inventory (NPI) [32], in which clinically trained staff members ask caregivers to rate the frequency and severity of symptoms in 12 psychiatric domains anxiety, motor restlessness/agitation,

aberrant motor behavior, depression, delusions, hallucinations, euphoria, disinhibition, and irritability. Like the BPRS, the NPI scoring involves a function (and not a simple sum) of severity and frequency ratings in each of the domains, respectively, and the total score is the sum of domain-specific frequency and severity ratings products. However, in the administration of the NPI, many subitems representing individual symptoms are rated, but not included in the total score. These subitems, in addition to the domain scores and total score, enable the clinician to identify specific behaviors of importance within each domain.

To generate specific depression scores for NS, the NPI subitems from across domains that most closely match the HRSD items were identified; these subitem ratings were summed to create a depression score for the NS cohort (0 to 31); an agitation indicator was created by summing the 1/0 (present/absent) ratings on each of the items in the agitation/aggression, irritability/lability, and aberrant motor behavior symptoms lists (0 to 22), whereas a psychosis indicator was similarly created from the delusions and hallucinations items (0 to 16).

2.4. Procedure

Modeling was carried out using EQS V. 6.1 [33]. All models (Figs. 1–4) were run separately for the NS and PR baseline responses, and fit indices (described below) were recorded for each model run. Models were fit using standard and robust methods (i.e., methods that are appropriate when modeling assumptions are not met); agreement of both standard and robust results was required for model and fit interpretation. Two “null” models, described below, were developed to reflect current practice and recent findings. In all models, the same indicator (observed) variables were entered in the same order. Models 1, 3, and 4 contain a latent “general neurological status” factor; models 2 through 4 contain latent factors reflecting cognition, function and behavior.

Model 1 (Fig. 1) hypothesizes that there is a single latent (unobservable) cause for all the variability observed in the cognitive (COG), functional (FUN), and behavioral (BEH) scores. It reflects the current clinical conceptualization of AD: all symptoms derive from a single unobserved source, namely, the AD dementing process(es).

Model 2 (Fig. 2) reflects 5 cognitive indicators hypothesized to load on a cognitive factor, 4 functional indicators on a function factor, and 4 behavioral indicators on a behavior factor.

Model 2 differs from model 1 in terms of the sources of variability in the observed instrument responses; the 3 factors are hypothesized to be the sole direct effectors on their respective indicators, with weak and unspecified indirect effects from each of the other 2 factors (i.e., these factors correlate [curved arrows] but do not directly affect the observed variables “outside” their respective domains). Model 2 reflects our earlier results using exploratory factor analysis [17] in independent cohort, which suggested causal independence among latent variables reflecting cognition, function, and behavior. Model 2 hypothesizes different sources for the observed variability in domain-specific indicators; this model embodies the independence in symptomatology we have reported previously, but through the correlations of the latent factors, model 2 also explains the dependence among symptoms that has been reported by other investigators.

Two second-order models were also developed and tested. Both models posit a second-order, “general neurological” (GN) latent variable, which could be conceptualized as neurophysiologic susceptibility or pathology. The first such model (Fig. 3) postulates that GN causes some variability in latent variables representing cognition, function, and behavior, that is, GN has a direct causal effect on these first-order latent variables, resulting in an indirect effect on the observed variables. Additionally, GN has a direct effect on the

observed variables. This “Direct + Indirect” model is model 3. In this model, the variability observed in the indicator variables is caused by the combination of the direct and indirect effects of both the general (second-order) and the symptom specific (first-order) factors. The GN factor also is causally related to each of the latent symptom factors. The association of COG, FUN, and BEH is owing to a common underlying second-order (causal) factor (GN), so that only changes in GN can cause (directly, indirectly) variability in the observed variables.

This model differs from the 3-factor null model (#2) by articulating a disease-specific component in GN. With this conceptualization, it is feasible that AD could become more severe with differential effects on the latent factors; disease severity (level of pathology) could also exert direct effects on the observed symptoms; like model 2, model 3 can account for both dependence and independence of symptoms that have been reported.

The second higher-order model (Fig. 4) postulates that GN has a direct causal effect on the observed variables but does not have a direct effect on the latent symptom factors. In model 4, the 3 latent symptom factors (COG, FUN, BEH) exert direct effects on the observed variables above and beyond, i.e., residualized after quantifying, the effects of GN. These relationships are shown as the “residualized” model (model 4), in which GN has direct (causal) effects on the observed symptoms; above and beyond this, the first-order latent symptom factors have additional direct effects on the observed variables. Thus, the variability in observed variables is caused by the combination of the direct effects of the general factor plus those of the symptom factors after “controlling for” the general factor.

In Model 4, disease severity (GN) does not affect the latent domain-specific factors, but it does affect the observed symptoms. Because the domain-specific (first-order) factors are correlated, both the dependence and independence in symptomatology that have been reported are explained (similar to model 2), but these factors have their influence on the observed variables only after the general factor is taken into account. The main difference between models 3 and 4 is that model 4 reflects the potential for disease severity changing (and resulting influence on observed variables) without influencing the symptom factors; similarly, any one of the symptom factors could change independent of specific changes in the GN factor.

2.5. Model Fit

The models were fit to the data from each cohort’s baseline visits separately using EQS software V. 6.1 (2005). Five different aspects of fit were assessed for each run in each cohort, reflecting general data-model fit (model), assessment of the fit of the model to data in hypothetical replications (Akaike’s Information Criterion [AIC]), incremental model fit relative to an independence model (comparative fit index [CFI]), error in approximation of the data by the model (root mean squared error of approximation [RMSEA]), and the mean absolute value of the covariance residuals (standardized root mean square residual [SRMR]) [19,34]. These indices describe different aspects of the fit of the model; the best fitting model should be superior in all indices. Further, robust model fit statistics [33] were calculated for all available indices, and model fit was only interpreted if robust and standard fit statistics led to the same conclusion.

The data from the NS study cohort were fit first, because with a large sample size ($N = 352$) it more than satisfies the rule of 5 cases per estimated parameter [19], even in the more complex models. The PR study cohort, with $N = 138$, only satisfied this rule for the 1 factor null model. The fit of the models to the PR data was considered supportive; as an independent cohort, these data were modeled to facilitate interpretability of results from the NS cohort.

3. Results

The 2 cohorts have been described elsewhere [21,23]. Both studies had similar inclusion and exclusion criteria. Table 1 presents descriptive statistics of the cohorts. Because the purpose of this modeling enterprise is to establish a model that fits generally, we did not compare the 2 clinical samples; our findings would be most generalizable if the samples differed but are also interpretable if the samples are similar.

Table 2 presents the fit statistics from the 4 models across the 2 cohorts. The fit to the data is described using 5 common fit indices. Robust and standard fit statistics gave similar results for all models; standard values are presented.

Table 2 shows that both samples were best fit with the residualized model, a second-order model with a general neurologic (GN) latent variable as a common causal factor in all observed scores; above and beyond this contribution, cognition, function, and behavior represent separate, but correlated, latent factors for their respective indicators. The lowest AIC and for model fit indicate that the residualized model fits the data best. Because the 2 second-order models have the same degrees of freedom (df), their X^2 for model fit cannot be compared; the AIC values suggest that the residualized model fits the data better with the same number of parameters. All fit indices are superior for the residualized model over the Direct + Indirect model in each study cohort. Standardized root mean square residuals (SRMR) were well below 0.10 [19] in both second-order models but smallest for the residualized models in each cohort. Standardized residuals matrices (data not shown) reflected no particular badness of fit, although the behavior indicators did tend to be the worst fit (i.e., 13 of 20 largest residuals involved behavior variables in NS; 15 of 20 in PR).

The 1-factor null model did not fit the data in either cohort, with all fit indices out of the acceptable range; the 3-factor null model fit the data better by all indices. Thus, the 3-factor null model is the appropriate one for comparisons.

The Direct + Indirect model was a significant improvement over the 3-factor null model for both cohorts (NS: X^2 difference: 86.8 on 13 df, $p < 0.01$; PR: X^2 difference: 48.8 on 13 df, $p < 0.01$). The residualized model also fit the data significantly better than this null model (NS: X^2 difference: 122.5 on 13 df, $p < 0.001$; PR: X^2 difference: 60.1 on 13 df, $p < 0.001$). Although the 2 second-order models cannot be compared with the X^2 difference test, the difference between the values is larger than a critical value for 5 degrees of freedom (differences are NS: 35.7 and PR: 11.3; critical value is 11.07 [with 5 df] for α of 0.05), suggesting the residualized model fits better than the other second-order model. As Table 2 shows, it also has the best fit indices in terms of AIC (both values negative, i.e., smallest), CFI (both values at or very close to 1.0, i.e., highest fit index), SRMR (closest residuals to 0.00, i.e., best prediction of observed covariances) and RMSEA (both values closest to 0.00, i.e., smallest amount of error in the approximation of the model to the data).

Because the residualized model fits both data sets best, Table 3 presents the path weights that describe the relationships between the first- (domain-specific) and second- (general neurologic) order latent variables and the observed indicators from this model. Additionally, the proportion of variance in each observed variable accounted for by the 2 latent factors is estimated by the R^2 that was calculated for each cohort. Table 3 contains only the standardized estimates of the contributions of the latent variables in explaining the variability in the observed variables (i.e., estimates of the degree to which the sum of the 2 path weights does *not* explain the variability in the observed variables are not given but can be simply derived as $1 - R^2$).

There are several important aspects shown in Table 3. It must be noted first, however, that structural equation modeling and confirmatory factor analysis permits the simultaneous fitting of the data from 2 or more groups to a single factor structure. The result would be path weights that can be compared directly across the groups or estimated for the 2 groups combined [19]. That is not an option in the current case because the data collected from the 2 cohorts were not the same. Although the path weights for both cohorts were standardized, this does not support the direct comparison of the path weights themselves. Instead, we must focus on the patterns of loadings exhibited across the 2 cohorts, treating the results from the NS cohort as primary with results from the PR cohort as reinforcement.

In the NS cohort, ADAS path weights are positive, whereas those for MMSE are negative; this reflects the fact that high scores on ADAS, but low scores on MMSE, suggest greater impairment. It is unclear why these results were not observed in the PR cohort; because the cognitive factor path weight for MMSE in the PR cohort was not significant, it is difficult to interpret these results for the PR cohort. The associations between MMSE and ADAS and the general and cognitive factors differed for the cohorts and also were different from what was observed for the other observed cognitive and functional scores. Conversely, for the CDR box scores and all behavioral indicators except behavior, loading patterns were the same for the 2 cohorts. As was observed for the MMSE/ADAS path weights for the 2 cohorts, psychosis and the derived dependency indicator showed different loading patterns (dependency: higher loading on first-order and lower on second-order factor in NS but lower loading on first-order and higher loading on second-order in PR; psychosis: nearly equal loadings in NS but higher loading on first-order and lower loading on second-order in PR) in the 2 cohorts.

In both cohorts, the maximum-likelihood estimation procedure for fitting the residualized model to the data required that the error variance associated with the observed behavior scores to be fixed at zero. The practical result of this pertains to the first- and second-order causal influences on these observed scores: the R^2 values reflect 100% of variance in the observed behavior scores to be explained by the 2 factors (first- and second-order). It is also the case that the models generally fit the observed behavioral data worst (although not badly; see below), and in general the path weights for the symptom-specific factor (BEH) were stronger than for general neurologic status factor.

The general behavior and symptom type (depression, psychosis, agitation) scores were derived from different instruments and symptom lists, and apart from the behavior indicator, the model fits and path weights tended to be similar in terms of loadings being higher for the symptom-specific than for the general factors (this was also found when all models were rerun without the behavior indicator; data not shown). This suggests that the general and specific behavioral indicators were functioning as intended.

In both cohorts the worst explanatory power was achieved for behavioral scores (NS: psychosis, 13.0% of variance explained; PR: agitation, 31.6% of variance explained). Although the 2 worst-explained scores were on the same factor in the 2 cohorts, the 2 best (This ignores the “perfect” explanatory power achieved by fixing the error variance for the general behavioral score to zero.) were not: for the NS sample, the functional indicator, home and hobbies, had the greatest level of variability explained (73.2%), with the next highest level explained by the MMSE (69.7%). However, in the PR sample, the best explanatory power was observed for ADAS (85.9%) with the next highest level explained by the CDR orientation box score (71.9%).

A final set of comparisons across cohorts can be seen in Table 4, where the estimated correlations between the 3 latent domain-specific factors appear. Although the relationships

between the latent factors and the observed variables (i.e., the path weights) are difficult to compare directly, theoretically, the factors representing cognition, function, and behavior are the same for the 2 cohorts.

Cognitive, functional, and behavioral symptoms tend to be associated in persons with AD, but not uniformly across the disease (in terms of severity and study cohort). The fact that the residualized model fit both cohorts best suggests that any observed associations between the 3 domain-specific latent factors may not simply reflect observed symptom comorbidity as a function of disease severity. Table 4 shows that the correlations among these unobservable factors differed fundamentally across the cohorts, although the association between behavior and function is positive and significant in both cohorts (NS: $r = .342$; PR: $r = .308$, both $p < 0.05$). The PR sample exhibited a nonsignificant correlation between cognition and function ($r = .239$), whereas the NS sample had a stronger and significant correlation between these latent factors ($r = .472$, $p < 0.05$). Also discrepant across cohorts was the association between cognition and behavior (NS: $r = .063$, $p > 0.50$; PR: $r = -.328$, $p < 0.05$).

4. Discussion

The purpose of this secondary analysis of baseline responses from 2 cohorts was to clarify the underlying structure of symptoms in AD. We did not set out to demonstrate factor invariance or to estimate path weights but only to explore formulations of the latent variable structure that could account for the wide range of “co-morbidity” or co-occurrence of the 3 main symptom domains in persons with AD.

Both samples, with different instrumentation for 2 of 3 latent symptom factors (function, behavior) supported the same second-order model, namely, one in which the symptom factors had residualized causal influences after taking into account the influence of a general factor on the observed scores. It is entirely possible that other models could fit the data as well or better than those explored here. Our purpose was not to derive the “correct” model but rather to explore the underlying structure of the symptoms that have been observed in persons with AD. Although we do not conclude that the residualized model is “true,” its validity is evidenced by independent support in separate cohorts, the functional and behavioral symptoms of which were assessed with different instrumentation. Because of this independent support, we believe it is a model worth pursuing. The fact that all models fit the data better than the 1-factor model (which did not fit in either cohort) suggests that this is an incorrect conceptualization of AD. The fit of the residualized model provides statistical evidence that there are domain-specific deficits that may or may not appear, above and beyond a generally compromised neurologic state.

The 3 other models involved separate factors for the 3 key symptom domains in AD, cognition, function, and behavior. All 3 of these models replicate earlier findings of *independence* across these 3 symptom domains as well as explain how and why *dependence* of symptoms may have been observed. This explanatory power is an improvement over a 1-factor model, which cannot account for observed independence.

A general neurological (GN) factor was identified in both cohorts and significantly improved the fit of the data in both relative to a first-order 3-factor model, suggesting that some of the variability in common outcome measures is derived from one “general” factor. Degree of synapse loss could be such a factor [35]. The variability across the samples in terms of the estimated degree of correlation among these domain-specific factors may reflect the difficulties that have been encountered in other modeling exercises in AD. Differences in the model specifics (i.e., path weights and patterns) must be replicated in larger samples, i.e., in which the case to parameter ratios are appropriate.

Although the residualized model fit both data sets, the estimated path weights were very different across study cohorts. The path weights for ADAS and MMSE on the cognitive latent variable had the opposite signs across the cohorts, and a possible explanation for these differences is the impact of cholinesterase inhibitor use. In the PR study, use of cholinesterase inhibitors was not allowed, whereas 68% of enrolled subjects in the NS study used one of these drugs at the baseline visit. Our results may suggest that cholinesterase inhibitors improve cognitive test scores (i.e., symptoms) without altering the underlying disease process. The coefficients of variation (CV, standard deviation divided by mean, derivable from Table 1) were similar for MMSE across the cohorts (NS: 17.8%; PR: 20.0%), probably because a specific range of MMSE scores was an inclusion criterion for both studies. However, for ADAS the CVs were different (NS: 24.2%; PR: 44.7%), suggesting less variability in ADAS scores for the NS cohort in which cholinesterase inhibitors were permitted. We might expect that causal effect of the latent symptom factor (COG) on symptoms would be attenuated in individuals undergoing treatment with a symptom-altering therapy such as cholinesterase inhibitors. By this argument, we would also expect to observe weaker causal effects (path weights closer to zero) of the second-order latent variable (GN) on symptoms in the treated group (NS) than in the untreated group (PR); this was observed. The presence of cholinesterase inhibitor treatment in one cohort but not the other may also explain the differences in correlations among the latent domain-specific factors (Table 4). These possibilities must be pursued with 2 cohorts of adequate sizes. Additionally, any effects of setting the scale of the latent factor must be explored in future models or cohorts.

A single underlying factor does not fit the data in these independent cohorts, and a second-order model with symptom-specific and general factors may be a useful model for AD in future studies. It is critical to keep in mind that the significance of the path weights in any given model is not our emphasis in terms of the utility of the modeling we describe here, but rather, our conclusions focus on the overall fit of each of the conceptually driven models to the 2 cohorts independently. We do not advocate strict inferential interpretation of the path weights in either model, but the observation of differential path weights across the cohorts suggests that an immediate application of these results might be the ability to explore where therapeutic interventions exert their effects. For example, acetylcholine-boosting therapies would be hypothesized to affect primarily the symptom factors, whereas anti-inflammatory agents would be expected to affect the general (second-order) factor to a greater extent; because both cohorts were untreated at baseline, this could not be tested.

A latent variable analytic approach to diseases with heterogeneous presentation and course can be very helpful in terms of increasing the precision of estimates as well as our understanding of the disease processes. When replicated and extended to model longitudinal data, our residualized model of symptoms in AD could explore and identify disease-versus-symptom-modifying interventions as well as to quantify their impacts on this and other neurodegenerative disease populations.

Acknowledgments

Supported by a Research Starters Grant in Health Outcomes from the Pharmaceutical Researchers and Manufacturers Association of America (PhRMA) Foundation to RET and a Mentored Research Scientist Award from NIA to RET, K01AG027172. The data were collected by the Alzheimer's Disease Cooperative Study under NIA Grant U01 AG10483.

References

1. Berchtold NC, Cotman CW. Evolution in the conceptualization of dementia and Alzheimer's disease: Greco-Roman period to the 1960s. *Neurobiol Aging*. 1998; 19(3):173–89. [PubMed: 9661992]

2. Devanand DP, Jacobs DM, Tang M-X, Del Castillo-Castaneda C, Sano M, et al. The course of psychopathologic features in mild to moderate Alzheimer disease. *Arch Gen Psychiatry*. 1997; 54:257–63. [PubMed: 9075466]
3. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR. 4. Arlington, VA: American Psychiatric Publishing; 2000.
4. Mohs RC, Schmeidler J, Aryan M. Longitudinal studies of cognitive, functional and behavioural change in patients with Alzheimer's disease. *Stat Med*. 2000; 19:1401–09. [PubMed: 10844705]
5. Haxby JV, Raffaele K, Gillette J, Schaprio MB, Rapoport SI. Individual trajectories of cognitive decline in patients with dementia of the Alzheimer type. *J Clin Exp Neuropsychol*. 1992; 14(4): 575–92. [PubMed: 1400920]
6. Ashford JW, Shan M, Butler S, Rajasekar A, Schmitt FA. Temporal quantification of Alzheimer's disease severity: 'Time Index' model. *Dementia*. 1995; 6:269–80. [PubMed: 8528374]
7. Rasmusson DX, Carson KA, Brookmeyer R, Kawas C, Brandt J. Predicting rate of cognitive decline in probable Alzheimer's disease. *Brain Cognition*. 1996; 31:133–47.
8. Stern Y, Liu X, Albert M, Brandt J, Jacobs DM, et al. Application of a growth curve approach to modeling the progression of Alzheimer's disease. *J Gerontol Med Sci*. 1996; 51A(4):M179–M184.
9. Hasselmo ME. A computational model of the progression of Alzheimer's disease. *MD Computing*. 1997; 14(3):181–91. [PubMed: 9151508]
10. Francis DJ, Fletcher JM, Steubing KK, Davidson KC, Thompson NM. Analysis of change: modeling individual growth. *J Consult Clin Psychol*. 1991; 59(1):27–37. [PubMed: 2002140]
11. Cheng, C-L.; Van Ness, JW. Statistical regression with measurement error. New York: Oxford University Press/Arnold; 1999.
12. Cummings JL. Cognitive and behavioral heterogeneity in Alzheimer's disease: seeking the neurobiological basis. *Neurobiol Aging*. 2000; 21:845–61. [PubMed: 11124429]
13. Buckner, RL. Three principles for cognitive aging research: multiple causes and sequelae, variance in expression and response, and the need for integrative theory. In: Cabeza, R.; Nyberg, L.; Park, LD., editors. *Cognitive neuroscience of aging: linking cognitive and cerebral aging*. New York: Oxford University Press; 2001. p. 267–85.
14. Howieson DB, Camicioli R, Quinn J, Silbert LC, Care B, Moore MM, et al. Natural history of cognitive decline in the old. *Neurology*. 2003; 60:1489–94. [PubMed: 12743237]
15. Stern RG, Mohs RC, Davidson M, Scheidler J, et al. A longitudinal study of Alzheimer's disease: measurement, rate and predictors of cognitive deterioration. *Am J Psychiatry*. 1994; 151:390–96. [PubMed: 8109647]
16. Yesavage JA, Brooks JO. On the importance of longitudinal research in Alzheimer's disease. *J Am Geriatr Soc*. 1991; 39:942–44. [PubMed: 1885869]
17. Tractenberg RE, Weiner MF, Cummings JL, Thal LJ. Independence of changes in behavior from changes in cognition and function in persons with Alzheimer's disease: a factor-analytic approach. *J Neuropsychiatry Clin Neurosci*. 2005; 17(1):51–60. [PubMed: 15746483]
18. Tractenberg RE, Weiner MF, Patterson MB, Schneider L, Thal LJ. Emergent psychopathology in Alzheimer's disease patients over 12 months associated with functional, not cognitive, changes. *J Geriatric Psychiatry Neurol*. 2002; 15(2):110–7.
19. Kline, RB. Principles and practice of structural equation modeling. 2. New York: Guilford Press; 2005.
20. Thal LJ. Development of the Alzheimer's Disease Cooperative Study. *Int J Geriatr Psychopharmacol*. 1997; 1:6–9.
21. Aisen PS, Davis KL, Berg JD, Schafer K, Campbell K, Thomas RG, et al. A randomized controlled trial of prednisone in Alzheimer's disease. Alzheimer's Disease Cooperative Study. *Neurology*. 2000; 54(3):88–93.
22. McKhann G, Drachman D, Folstein M, et al. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of the Department of Health and Human Services Task Force on Alzheimer's disease. *Neurology*. 1984; 34:939–944. [PubMed: 6610841]
23. Aisen PS, Schafer KA, Grundman M, Pfeiffer E, Sano M, Davis KL, et al. Alzheimer's Disease Cooperative Study. Effects of rofecoxib or naproxen vs placebo on Alzheimer disease progression: a randomized controlled trial. *JAMA*. 2003; 289(21):2819–26. [PubMed: 12783912]

24. Mohs RC, Rosen WG, Davis KL. The Alzheimer's Disease Assessment Scale (ADAS): An instrument for assessing treatment efficacy. *Psychopharmacol Bull.* 1983; 18:448–50. [PubMed: 6635122]
25. Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. *Am J Psychiatry.* 1984; 41:1356–64. [PubMed: 6496779]
26. Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical scale for the staging of dementia. *Br J Psychiatry.* 1982; 140:566–72. [PubMed: 7104545]
27. Folstein M, Folstein S, McHugh P. "Mini-Mental State": a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res.* 1975; 12:189–98. [PubMed: 1202204]
28. Stern Y, Albert SM, Sano M, et al. Assessing patient dependence in Alzheimer's disease. *J Gerontol.* 1994; 49(5):M216–22. [PubMed: 8056940]
29. Galasko D, Bennett D, Sano M, Ernesto C, Thomas R, Grundman M, et al. An inventory to assess activities of daily living for clinical trials in Alzheimer's disease. The Alzheimer's Disease Cooperative Study. *Alzheimer Dis Assoc Disord.* 1997; 11(Suppl 2):S33–39. [PubMed: 9236950]
30. Overall JE, Gorham DR. Brief Psychiatric Rating Scale. *Psychol Rep.* 1962; 10:799–812.
31. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry.* 1960; 23:56–62. [PubMed: 14399272]
32. Cummings JL, Mega M, Gray K, Rosenberg-Thompson S, Carusi DA, Gornbein J. The Neuropsychiatric Inventory: comprehensive assessment of psychopathology in Dementia. *Neurology.* 1994; 44:2308–14. [PubMed: 7991117]
33. Bentler, PM.; Wu, E. EQS structural equations program manual. Encino, CA: Multivariate Software Inc; 1995.
34. Loehlin, JC. Latent variable models: An introduction to factor, path, and structural equation analysis, 4E. Mahwah, NJ: Lawrence Earlbaum Associates; 2004.
35. DeKosky ST, Scheff SW. Synapse loss in frontal cortex biopsies in Alzheimer's disease: correlation with cognitive severity. *Ann Neurol.* 1990; 27(5):457–64. [PubMed: 2360787]

Fig. 1.

Null model #1: A first-order (measurement) model with one factor. All scores were from the pretreatment visit (baseline [BL]); boxes are observed variables and circles are latent factors. C1-C3 = CDR box scores for cognitive function; DEPEND = % dependence, derived from either the Dependency Scale or the ADL Inventory; F1-F3 = CDR box scores for function; BEH = General behavioral measure, derived as either the NPI or BPRS total score; DEPR = depression score, derived as either the Hamilton total score or the sum of frequency ratings on NPI items most closely matching the Hamilton; PSYC = psychosis score, derived from BPRS or NPI items; AGIT = agitation score, derived from BPRS or NPI items. The latent variable derives its scale from standardization of the factor variance (set = 1.0). E1-E13

terms represent the error associated with the observed variables' reflection of the underlying "true" state.

Fig. 2.

Null Model #2: A first-order (measurement) model with 3 factors. All scores were from the pretreatment visit (baseline [BL]); boxes are observed variables and circles are latent factors. C1-C3 = CDR box scores for cognitive function; DEPEND = % dependence, derived from either the Dependency Scale or the ADL Inventory; F1-F3 = CDR box scores for function; BEH = General behavioral measure, derived as either the NPI or BPRS total score; DEPR = depression score, derived as either the Hamilton total score or the sum of frequency ratings on NPI items most closely matching the Hamilton; PSYC = psychosis score, derived from BPRS or NPI items; AGIT = agitation score, derived from BPRS or NPI items. In this model, the latent variables derive their scales from the first indicator in the list. E1-E13 terms represent the error associated with the observed variables' reflection of the underlying "true" state. Curved arrows represent correlation between latent factors.

Fig. 3.

A second-order model representing disease-specific (second-order) and symptom-specific (first-order) latent variables. All scores were from the pretreatment visit (baseline [BL]); boxes are observed variables and circles are latent factors. C1-C3 = CDR box scores for cognitive function; DEPEND = % dependence, derived from either the Dependency Scale or the ADL Inventory; F1-F3= CDR box scores for function; BEH = General behavioral measure, derived as either the NPI or BPRS total score; DEPR = depression score, derived as either the Hamilton total score or the sum of frequency ratings on NPI items most closely matching the Hamilton; PSYCHOSIS = psychosis score, derived from BPRS or NPI items; AGIT = agitation score, derived from BPRS or NPI items. In this model, the latent variables derive their scales from standardization of the factors' variances (all = 1). E1-E13 terms represent the error associated with the observed variables' reflection of the underlying "true" state; similarly, D2-D4 terms represent error associated with the latent symptom-level variable's reflection of the "true" general neurologic state. Arrows from Gen.Neuro to the 3 factors are curved solely because that fit better; these are causal arrows and not correlations.

Fig. 4.

A second-order residualized model representing disease-specific (second-order) and symptom-specific (first-order) latent variables. All scores were from the pretreatment visit (baseline [BL]); boxes are observed variables and circles are latent factors. C1-C3 = CDR box scores for cognitive function; DEPEND = % dependence, derived from either the Dependency Scale or the ADL Inventory; F1-F3 = CDR box scores for function; BEH = General behavioral measure, derived as either the NPI or BPRS total score; DEPR = depression score, derived as either the Hamilton total score or the sum of frequency ratings on NPI items most closely matching the Hamilton; PSYC = psychosis score, derived from BPRS or NPI items; AGIT = agitation score, derived from BPRS or NPI items. In this

model, all latent variables derive their scales from standardization of the factors' variances (all = 1). E1-E13 terms represent the error associated with the observed variables' reflection of the underlying "true" state. Curved arrows between symptom-level factors represent correlations.

Table 1

Descriptive statistics for scores reflecting cognition, function, and behavior at baseline in 2 clinical trial samples of persons with Alzheimer's disease

Domain (latent factor)	Test/indicator (Observed variable)	Mean $\pm$ SD, Prednisone (N = 138)	Mean $\pm$ SD, Nonsteroidal (N = 351)
Cognition	ADAS	21.74 $\pm$ 9.71	29.94 $\pm$ 7.25
	MMSE	21.64 $\pm$ 4.33	20.76 $\pm$ 3.70
	Memory (C1)	1.11 $\pm$ 0.46	1.23 $\pm$ 0.52
	Orientation (C2)	1.00 $\pm$ 0.55	1.05 $\pm$ 0.51
	Judgment (C3)	0.98 $\pm$ 0.47	1.12 $\pm$ 0.61
Function	Dependency (DEPEND)	39.4% $\pm$ 16.5%	25.8% $\pm$ 12.8%
	Community (F3)	0.94 $\pm$ 0.47	0.98 $\pm$ 0.55
	Home & hobbies (F2)	0.95 $\pm$ 0.55	1.01 $\pm$ 0.60
	Personal care (F3)	0.42 $\pm$ 0.62	0.36 $\pm$ 0.58
Behavior	Behavior (BEH)	28.58 $\pm$ 5.92	9.12 $\pm$ 9.88
	DEPR	3.09 $\pm$ 3.42	2.75 $\pm$ 3.22
	PSYCH	1.22 $\pm$ 2.16	3.17 $\pm$ 7.07
	AGIT	0.86 $\pm$ 1.39	10.55 $\pm$ 11.14

NOTE. All scores were from a pretreatment visit.

C1-C3, CDR box scores for cognitive function; DEPEND, % dependence, derived from either the Dependency Scale or the ADL Inventory; F1-F3, CDR box scores for function; BEH, General behavioral measure, derived as either the NPI or BPRS total score; DEPR, depression score, derived as either the Hamilton total score or the sum of frequency ratings on NPI items most closely matching the Hamilton; PSYCHOSIS, psychosis score, derived from BPRS or NPI items; AGIT, agitation score, derived from BPRS or NPI items.

Table 2

Model-data fit for 2 null (first-order) and 2 second-order models by study sample

Model	Group	χ^2 (df, p value)	AIC	CFI	SRMR	RMSEA (90% CI)
Null 1 (1 factor)	NS	633.4 (65df, <0.0001)	503.4	.733	.106	.158 (.147, .169)
	PR	390.4 (65df, <0.0001)	260.4	.655	.129	.192 (.173, .210)
Null 2 (3 factor)	NS	228.6 (62df, <0.0001)	104.6	.922	.066	.088 (.075, .100)
	PR	106.0 (62df, <0.0001)	-18.0	.953	.069	.072 (.048, .095)
Direct + indirect	NS	141.8 (49df, <0.0001)	43.8	.956	.044	.074 (.059, .088)
	PR	57.2 (49df, =.19)	-40.8	.991	.052	.035 (.000, .068)
Residualized	NS	106.1 (49df, <0.0001)	8.1	.973	.035	.058 (.043, .072)
	PR	45.9 (49df, =.60)	-52.1	1.0	.032	.000 (.000, .050)

NOTE. Models were fit in NS sample first (N = 351) and then PR (N = 138).

AIC: Akaike's Information Criterion; the lower, the better. CFI: Comparative fit index; the closer to zero (and positive) the better, acceptable models have CFI>.96. SRMR: standardized root mean square residuals, the smaller (and <0.10) the better. RMSEA: root mean square error of approximation; the closer to zero (and positive) the better, acceptable models have an upper bound on the 90% CI <0.1. Null Model 1: a single first-order latent factor representing unobservable "Alzheimer's disease plus aging effects" is the only causal effect on observed variables. Null Model 2: 3 first-order latent factors representing cognition, function, and behavior are the only causal effects on observed variables; each has direct effects on its respective indicators and correlates with the other latent variables. Direct/Indirect Model: General Neurological Status is a second-order latent variable with both direct and indirect causal effects on observed variables. This second-order variable also has causal effects on the 3 domain-specific first-order latent factors (cognition, function, and behavior), which have direct causal effects on observed variables. Residualized Model: General Neurological Status is a second-order latent variable with direct effects on observed variables; 3 domain-specific, first-order latent factors (cognition, function, and behavior), have direct causal effects on observed variables above and beyond the effects of the General Neurological Status factor.

Table 3
Standardized structural equations (factor loadings only) for observed variables under residualized model by study cohort

Observed Variable (indicator)	Nonsteroidal path weights *			Prednisone path weights		
	GEN. NEURO (second-order latent variable)	COG, FUN or BEH (first-order latent variable)	R ² , proportion of variance in indicator explained by first- and second-order latent variables	GEN. NEURO (second-order latent variable)	COG, FUN or BEH (first-order latent variable)	R ² , proportion of variance in indicator explained by first- and second-order latent variables
ADAS	.353	.624	.514	.773	-.502	.849
MMSE	-.411	-.727	.697	-.760	.357	.704
Memory (C1)	.792	.185	.661	.684	.125	.483
Orientation (C2)	.742	.333	.661	.846	.055	.719
Judgment (C3)	.608	.351	.493	.674	.021	.455
Dependence	.527	.628	.671	.596	.478	.583
Community (F1)	.773	.262	.667	.705	.438	.689
Home/hobbies (F2)	.765	.383	.732	.636	.553	.711
Personal care (F3)	.427	.531	.465	.535	.520	.557
Behavior	.265	.964	1.0 [†]	.327	.945	1.0 [†]
Depression	.143	.665	.463	.002	.736	.542
Psychosis	.224	.283	.130	.261	.675	.523
Agitation	.300	.544	.386	.177	.534	.316

* All pathweights significant ($p < 0.05$). Bold indicates significant pathweight.

[†]Variance fixed to zero to facilitate model fitting.

Table 4

Correlations between latent domain-specific factors for 2 independent cohorts (PR, NS)

	Cognition	Function	Behavior
Cognition	—	.472 [*] (NS)	.063 (NS)
Function	.239 (PR)	—	.342 [*] (NS)
Behavior	-.328 [*] (PR)	.308 [*] (PR)	—

Abbreviations: PR, data derived from Prednisone cohort; NS, data derived from Nonsteroidal cohort.

*
 $p < 0.05$.