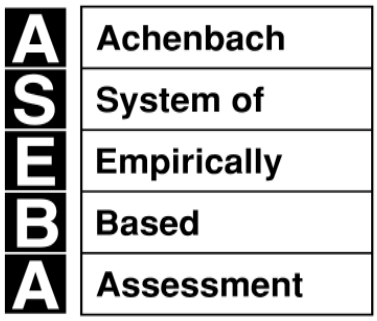

This is an excerpt from:

Manual for the ASEBA Older Adult Forms & Profiles

For Ages 60-90+

- **Older Adult Self-Report**
- **Older Adult Behavior Checklist**

An Integrated System of Multi-informant Assessment

Thomas M. Achenbach, University of Vermont
Paul A. Newhouse, University of Vermont
& Leslie A. Rescorla, Bryn Mawr College

Chapter 8

Reliability, Internal Consistency, and Cross-Informant Agreement

Reliability refers to agreement between repeated assessments of characteristics when the characteristics themselves are expected to remain constant. When instruments such as the OASR and OABCL are completed, it is important to know the degree to which scale scores remain consistent over periods when the adaptive characteristics and problems of the people who are assessed are not likely to change much. In this chapter, we present test-retest reliabilities for the OASR and OABCL scale scores over 8-day intervals.

Another property of scale scores is their *internal consistency*. This refers to the degree to which the items of a scale are correlated with each other. Internal consistency is sometimes called *split-half reliability*. This is because it can be estimated by correlating the sum of scores on half the items of a scale with the sum of scores on the other half. However, internal consistency among the items of a scale scored on a single occasion cannot tell us the reliability with which the scale will produce similar results on different occasions.

Test-retest reliability and internal consistency are typically viewed as psychometric properties of the scales themselves. However, reports of adaptive functioning and problems inevitably depend on the informants' perspectives. When there are doubts about older adults' mental competence, reports by informants are often sought as substitutes for self-reports. Yet, even when older adults' mental competence is in question, it is important to obtain and consider their views of their own functioning, if possible. Furthermore, different informants may contribute different but valuable information about older adults' functioning. For example, a spouse or partner may report strengths and problems that differ from the strengths and problems reported by the older adult's grown children or by caregivers. As illustrated in preceding chapters, the OASR and OABCL are designed to obtain and systematically compare data from self-reports and reports by other people. In this chapter, we present findings for *cross-informant agreement* between scale scores obtained from self-reports and from reports by others.

TEST-RETEST RELIABILITIES OF SCALE SCORES

To assess reliability in both the rank ordering and the magnitude of scale scores, we computed test-retest Pearson correlations (*r*) and *t* tests of differences between OASR ratings and between OABCL ratings on two occasions. The test-retest reliability samples were obtained by requesting older adults and informants to complete their respective forms twice at intervals averaging 8 days.

Test-Retest Correlations

As shown in Table 8-1, reliability was generally very high, with all test-retest *r*s being significant at $p<.01$ and most being in the .80's and .90's. The mean *r* for the adaptive functioning scales was .89 for the OASR and .94 for the OABCL. For the empirically based problem scales, the mean *r*s were also .89 and .94, while the *r* for the Total Problems scores on both forms was .95. For the DSM-oriented scales, the mean *r* was .88 on the OASR and .93 on the OABCL.

There were significant ($p<.05$) Time 1 to Time 2 increases in scores on the scales that are marked with superscript *d* in Table 8-1. Three of the significant increases from Time 1 to Time 2 in each column would be expected by chance, based on the number of analyses that were done, using a $p<.05$ protection level (Sakoda, Cohen, & Beall, 1954). Superscript *e* indicates the differences that were most likely to be sig-

Table 8-1
8-Day Test-Retest Reliabilities, Alphas, and Cross-Informant Correlations[a]

	Test-Retest r		*Alpha Coefficients*		*Cross-Informant r*
Scales	***OASR***	***OABCL***	***OASR***	***OABCL***	***OASR x OABCL***
Adaptive Functioning	*N* = 53	55	440	732	1,142
Friends	.93	.95	.70	.70	.54
Spouse/Partner[b]	.81	.93	.66	.71	.40
Personal Strengths	.91	.95	.83	.90	.48
Mean *r*[c] and mean alpha	.89	.94.	.73	.77	.48
Q Correlation for Personal Strengths items	NA	NA	NA	NA	.37
Anxious/Depressed	.92[d]	.94[d, e]	.92	.92	.54
Worries	.87	.94[d]	.69	.66	.46
Somatic Complaints	.74	.92[d, e]	.82	.79	.55
Functional Impairment	.83	.96	.86	.90	.65
Memory/Cognition	.84	.93	.84	.89	.47
Thought Problems	.94	.95[d]	.81	.85	.41
Irritable/Disinhibited	.91	.94	.82	.90	.44
Total Problems	.95[d, e]	.95[d, e]	.96	.97	.54
Mean *r*[c] and mean alpha	.89	.94	.84	.86	.51
Critical Items	.90[d]	.95[d]	.89	.91	.53
DSM-Oriented					
Depressive Problems	.79	.95[d]	.88	.89	.53
Anxiety Problems	.83	.90	.82	.81	.52
Somatic Problems	.78	.94[de]	.77	.75	.51
Dementia Problems	.91[d, e]	.93	.79	.88	.44
Psychotic Problems	.93[d, e]	.94[d]	.69	.82	.29
Antisocial Personality Problems	.93	.92	.63	.78	.34
Mean *r*[c] and mean alpha	.88	.93	.76	.82	.44
Mean Q correlation for problem items	NA	NA	NA	NA	.38

[a]Test-retest samples comprised respondents who completed forms at mean intervals of 8 days. Cronbach's *alpha*s were computed for referred and nonreferred samples described in Chapter 9. Cross-informant *r*s were computed for all members of our nonreferred and referred samples who had both an OASR and OABCL. All *r*s were $p<.01$. Mean *r*s were computed by Fisher's *z* transformation.

[b]Because the Spouse/Partner scale was not relevant for all informants, *N*s for test-retest *r* were 16 and 23 and for alpha were 185 and 313 for the OASR and OABCL, respectively; the cross-informant *N* was 498.

[c]Mean *r* computed by *z* transformation.

[d]Time 1 < Time 2, at $p<.05$, by *t* tests.

[e]When corrected for the number of comparisons, Time 1 vs. Time 2 difference was not significant (Sakoda et al., 1954).

nificant by chance, because they yielded the smallest *t* values among the *t* tests of Time 1 vs. Time 1 scores.

The tendency of older adults' problem scores to increase from Time 1 to Time 2 is *opposite* to the tendency that has been found for such scores to decrease among children and younger adults (e.g., Achenbach & Rescorla, 2001, 2003; Helzer, Spitznagel, & McEvoy, 1987; Roberts, Solovitz, Chen, & Casat, 1996; Vandiver & Sher, 1991). None of the adaptive functioning scales showed significant changes in mean scores over the test-retest intervals.

INTERNAL CONSISTENCY OF SCALE SCORES

High internal consistency is often considered to be desirable and is sometimes interpreted as indicating that a scale is highly reliable. However, because internal consistency is computed on the basis of item scores that were all obtained at one point in time, it cannot really tell us the test-retest reliability of scale scores from one occasion to another. Furthermore, some scales with very high internal consistency may not be as valid as some scales with lower internal consistency (Cattell & Kline, 1977).

As an example, if a scale consists of 15 versions of the same question, it should have very high internal consistency, because respondents should give similar answers to the 15 versions of the question. However, despite its high internal consistency, a scale consisting of 15 versions of the same question would usually be less valid than a scale that used 15 different questions to assess the same phenomenon. This is because each of the 15 different questions is likely to tap different aspects of the target phenomenon and to be subject to different errors of measurement.

As detailed in Chapter 7, our syndrome scales were derived from factor analyses of the correlations among ASEBA items. The composition of the syndrome scales is therefore based on internal consistencies among certain subsets of items. Nevertheless, because some users may wish to know the degree of internal consistency of our scales, Table 8-1 displays Cronbach's (1951) *alpha* for each scale, computed for the 440 OASRs and 732 OABCLs for the demographically similar samples of referred and nonreferred older adults described in Chapter 9. Alpha represents the mean of the correlations between all possible sets of half the items comprising a scale. Alpha tends to be directly related to the length of the scale, because half the items of a short scale provide a less stable measure than half the items of a long scale.

As Table 8-1 shows, the alphas for the Friends and Spouse/Partner scales ranged from .66 to .71, whereas the alphas for the Personal Strengths scale were .83 and .90. The higher alphas for the Personal Strengths scale probably resulted from its greater length (20 items vs. 4 to 7 items for the Friends and Spouse/Partner scales).

For the empirically based problem scales, the alphas ranged from .66 to .97. The only alphas that were below .70 were for the OASR and OABCL Worries syndrome, which has the fewest items. For the DSM-oriented scales, the alphas ranged from .63 to .89. The only alphas below .70 were on the OASR Psychotic Problems and Antisocial Personality Problems scales.

CROSS-INFORMANT AGREEMENT

Pearson Correlations

To determine whether the overall level of cross-informant agreement differed according to the type of informant, we computed Pearson *r*s between OASR and OABCL Total Problems scores for the following groups: All people in our referred and nonreferred samples who had both an OASR and OABCL, divided according to whether the OABCL was completed ***(a)*** by spouses or partners ($N = 376$); ***(b)*** by grown children ($N = 310$); or ***(c)*** by other informants, which included other family members, friends, and caregivers ($N = 456$). The *r*s were remarkably similar, ranging from .51 for other informants to .53 for children, and .58 for spouses and partners. To determine whether the level of cross-informant agreement differed according to the gender or age of the people who were assessed, we computed *r*s between OASR and OABCL Total Problems scores for people grouped by gender and age as follows: ***(a)*** men ages

60 to 75 ($N = 253$); ***(b)*** men ages >75 ($N = 174$); ***(c)*** women ages 60 to 75 (($N = 426$); and ***(d)*** women ages >75 ($N = 289$). For these groups, the *r*s were also remarkably similar, ranging from .47 for men ages >75 to .50 for men ages 60 to 75, and .57 for women in both age groups. Because the levels of agreement were so similar for the various informants and the different gender/age groups, we computed *r*s between scores for all the OASR and OABCL scales for all individuals who were assessed with both forms, as shown in Table 8-1.

The cross-informant *r*s ranged from .29 for the DSM-oriented Psychotic Problems scale to .65 for the Functional Impairment syndrome, all $p<.01$. The mean cross-informant *r* was .48 for the adaptive functioning scales, .51 for the empirically based problem scales, and .44 for the DSM-oriented scales.

Q Correlations

The bottom line of Table 8-1 displays the mean of the *Q* correlations between the 0-1-2 ratings obtained on the Personal Strengths items and on the problem items of OASRs completed by participants in our matched referred and nonreferred samples (details of the samples are in Chapter 9) and on the counterpart items of OABCLs completed for the participants by people who knew them. Each *Q* correlation reflects the degree of agreement between the pattern of 0-1-2 ratings for OASR items and the corresponding OABCL items rated for the same individual. Like Pearson *r*, *Q* correlations can range from –1.00, which indicates total disagreement, to +1.00, which indicates perfect agreement between two sets of problem item ratings for the same individual. The mean *Q* correlations of .37 between Personal Strengths items and .38 between problem items indicate a modest level of agreement.

If you use the Ages 60-90+ Module to score an OASR and up to seven OABCLs for the same individual, you can have the Module print *Q* correlations between the OASR and each of the OABCLs. If you choose this option, the Module will also print the mean *Q* correlations shown in Table 8-1, plus the 25th percentile *Q* correlation and the 75th percentile *Q* correlation. For Personal Strengths, the 25th percentile *Q* correlation = .19, while the 75th percentile *Q* correlation = .58. For problem items, the 25th percentile Q correlation = .25, while the 75th percentile Q correlation = .50. If a particular *Q* correlation is <25th percentile, the printout will state that it is below average. If the *Q* correlation is >75th percentile, the printout will state that it is above average.

Differences Between OASR and OABCL Scale Scores

To determine whether the mean scale scores on OASR scales differed from those on the OABCL, we performed 2 (gender) x 2 (ages 60-75 vs. 75) x 2 (self vs. informant) ANOVA tests on the OASR vs. OABCL scores for the 1,142 people who had both. OABCL scores were significantly ($p<.01$) higher than OASR scores on the Friends scale, and did not differ significantly for the sum of the six Spouse/Partner items that are the same on both forms. However, OABCL scores were significantly ($p<.01$) lower than OASR scores on the Personal Strengths scale and all problem scales except the DSM-oriented Somatic Problems scale, where the mean scores were identical on both forms. There was thus a consistent tendency for older adults to report more problems than the informants did. However, they also tended to present more favorable pictures of themselves on the Spouse/Partner and Personal Strengths scales, but not on the Friends scale.

SUMMARY

The test-retest reliability of ASEBA older adult forms was supported by 8-day test-retest *r*s that were in the .80s and .90s for most scales. The mean *r*s ranged from .88 for the OASR DSM-oriented scales to .94 for the OABCL adaptive functioning and empirically based scales.

Good internal consistency was found for most scales, with mean alpha coefficients on the OASR and OABCL of .84 and .86 for the empirically based problem scales, .76 and .82 for the DSM-oriented scales, and .73 and .77 for the adaptive functioning scales, respectively.

Cross-informant *r*s between OASR and OABCL

scores averaged .51 for the empirically based problem scales, .44 for the DSM-oriented scales, and .48 for the adaptive functioning scales. The mean *Q* correlation between OASR and OABCL Personal Strengths items = .37 and between problem items = .38. Consistent with findings from other instruments, the modest size of the cross-informant correlations indicates the need to obtain data from multiple informants whenever possible. OASR scores were significantly higher than OABCL scores on the Spouse/Partner and Personal Strengths scales, as well as on all but one problem scale. OASR scores were significantly lower than OABCL scores on the Friends scale.

Chapter 9

Validity

Validity refers to the accuracy with which instruments assess what they are supposed to assess. ASEBA instruments serve many purposes, and their validity can be evaluated in multiple ways. A fundamental purpose of the OASR and OABCL is to aid in identifying needs for help with behavioral, emotional, and social problems and adaptive functioning. The OASR and OABCL provide well-differentiated pictures of people's functioning in terms of specific problems and adaptive characteristics, aggregations of related problems into empirically based and DSM-oriented scales, and broader aggregations of items that encompass diverse aspects of functioning. In this chapter, we present evidence for the *content validity, criterion-related validity,* and *construct validity* of the OASR and OABCL.

CONTENT VALIDITY

The most basic kind of validity is *content validity,* which is the degree to which an instrument's content includes what the instrument is intended to assess.

Problem Items

The OASR and OABCL problem items are products of a long process of development, testing, and refinement on the basis of research and practical experience. The process began with the selection of items for assessing children and youth on the basis of extensive literature searches, consultation with relevant professionals, and repeated pilot testing in a variety of samples (Achenbach, 1965, 1966; Achenbach & Edelbrock, 1983; Achenbach & Lewis, 1971). Applications of the ASEBA approach to assessment of adults began in the 1980s with development of the Young Adult Self-Report (YASR) and Young Adult Behavior Checklist (YABCL) to assess adults who had previously been assessed with ASEBA school-age instruments (Achenbach, 1991a; Achenbach et al., 1995c; Stanger, MacDonald, McConaughy, & Achenbach, 1996).

The *Manual for the Young Adult Self-Report and Young Adult Behavior Checklist* (Achenbach, 1997) provides details of the refinement and testing of the items and scales for the young adult forms and profiles. The items for assessing adults were further refined, tested, and augmented in developing the Adult Self-Report (ASR) and Adult Behavior Checklist (ABCL) for ages 18 to 59, as detailed by Achenbach and Rescorla (2003). Items from these forms that were appropriate for older adults were combined with numerous new items written specifically for older adults to construct pilot versions of the OASR and OABCL.

The pilot versions of the older adult forms were initially administered in interview format to residents of retirement communities and to people attending senior centers. Based on feedback from the interviewees, items were revised and new items were added. Revised pilot editions were administered as interviews and also for independent completion by older adults and informants. Respondents were asked to comment on items and to suggest additional items.

The versions of the forms that resulted from this process were then used in the National Survey and were completed by people in a variety of settings serving older adults, as described in Chapter 6. Problem items were retained for the final versions of the older adult forms if they met at least one of the following criteria: ***(a)*** The items were scored significantly higher on one or both forms for older adults referred for mental health or substance use services than for demographically similar older adults who had not been referred for such services in the preceding 12 months (Chapter 10 presents detailed item analyses); ***(b)*** the items loaded significantly on empirically based syndromes (Chapter 7 presents details of how the syn-

dromes were derived); or ***(c)*** the items were identified by the expert panel as being very consistent with DSM-IV diagnostic categories (Chapter 4 presents details of the DSM-oriented scales).

Adaptive Functioning Items

The adaptive functioning items were hypothesized to reflect aspects of functioning that are important for successful adaptation in various areas. Like the problem items, the older adult adaptive functioning items were developed from the ASEBA adult forms, successive pilot editions of the older adult forms, feedback from older adults in diverse settings, and tests of the ability of the items to discriminate between referred and nonreferred samples of older adults. The items pertaining to friends and to personal strengths are relevant to nearly all older adults who are apt to be assessed with the ASEBA forms. The spouse/partner items are completed only for people who lived with a spouse or partner in the preceding 2 months. As detailed in Chapter 10, all the adaptive functioning items discriminated significantly between referred and nonreferred older adults on at least one of the forms.

In summary, the content validity of the OASR and OABCL items has been supported by a long process of item development, testing, and revision. The content validity of the items has also been supported by findings that all the retained items discriminated significantly between demographically similar referred and nonreferred older adults, loaded significantly on empirically based syndromes, and/or were identified by experts as being very consistent with DSM diagnostic categories.

CRITERION-RELATED VALIDITY OF SCALE SCORES

Criterion-related validity refers to the strength of association between a particular measure, such as a scale scored from an ASEBA form, and an external criterion for characteristics that the scale is intended to assess. In the preceding section, we mentioned that most of the items retained for OASR and OABCL scales discriminated significantly between referred and nonreferred older adults on one or both forms. Here we focus on associations between scales comprising particular sets of ASEBA items and external criterion variables.

Demographically Similar Samples of Referred and Nonreferred Older Adults

To test the ability of each ASEBA scale to discriminate between referred and nonreferred people, we constructed referred and nonreferred samples having similar gender and age distributions. The referred adults came from four mental health and substance abuse outpatient services. The OASRs and OABCLs for nonreferred adults were for individuals in our normative sample who reported that they had not received mental health or substance use services in the preceding 12 months. We controlled for differences in education by treating it as a covariate in ANCOVAs and as an independent variable in multiple regression analyses. As described in Chapter 6, the OASR and OABCL request respondents to indicate the older adult's educational level in terms of nine levels ranging from a score of 1 for no high school diploma and no General Equivalency Diploma (GED) to a score of 9 for doctoral or law degree. Because most of the people assessed in clinical services were whites, non-Latino white people were drawn from the normative sample to ensure that ethnic differences would not be confounded with differences associated with referral status, age, gender, and educational level. Table 9-1 summarizes characteristics of the demographically similar referred and nonreferred samples.

Because there was not enough ethnic variation to test for effects of ethnicity in the demographically similar referred and nonreferred samples, we tested for ethnic differences in scale scores in the OASR and OABCL normative samples as follows: For each adaptive functioning and problem scale of the OASR and OABCL, we performed a multiple regression on age, gender, education, and the binary (dummy) variables of white vs. other ethnicity, African American vs. other ethnicity, and Latino vs. other ethnicity. As was shown in Table 6-1, the *N* was 1,397 for the OASR and 822 for the OABCL. Both samples included 79% non-Latino white and 5% Latino participants. African Americans comprised 14% of the OASR sample and

Table 9-1
Characteristics of Demographically Similar Referred vs. Nonreferred Samples

	OASR		*OABCL*	
Characteristics	***Ref.***	***Nonref.***	***Ref.***	***Nonref.***
	N = 220	*N* = 220	*N* = 366	*N* = 366
Gender				
Men	32%	32%	38%	38%
Women	68%	68%	62%	62%
Age				
Mean	73.9	73.9	77.1	76.9
SD	7.4	7.4	7.8	7.5
Education[a]				
Mean	4.6	3.6	4.0	3.5
SD	2.3	2.2	2.3	2.2
SES[b]				
Mean	2.3	2.1	2.2	2.1
SD	0.7	0.7	0.7	0.7

[a]Education was scored: 1 = No high school diploma or General Equivalency Diploma (GED); 2 = GED; 3 = High school graduate; 4 = Some college; 5 = Associate's degree; 6 = Bachelor's or RN degree; 7 = Some graduate school; 8 = Master's degree; 9 = Doctoral or Law degree.

[b]SES was scored 1 = lower, 2 = middle, 3 = upper, based on an updated version of Hollingshead's (1975) 9-step scale for the occupation of the spouse/partner holding the higher status job: Hollingshead scores 1.0-3.9 = lower; 4.0-6.9 = middle; 7.0-9.0 = upper; we assigned 2-digit codes because occupations that were not clearly scorable were given the mean of their most likely scores.

13% of the OABCL sample. People who identified themselves as being of mixed or other ethnicities comprised 2% of the OASR sample and 3% of the OABCL sample.

Despite the very high statistical power afforded by the large sample sizes, no effects of ethnicity reached the $p<.05$ level of significance in the analyses of the 18 OABCL scales, and only one effect was significant in the analyses of the 18 OASR scales. The one nominally significant effect was less than the number of significant effects expected by chance, which were as follows (Sakoda et al., 1959): 3 effects would be significant by chance in 18 analyses of each binary variable for ethnicity; 6 effects would be significant by chance in the collective total of 54 analyses of the 3 binary variables for ethnicity on the 18 OASR scales (3 binary variables for ethnicity x 18 scales = 54 analyses). When the effects of age, gender, and education were partialed out, there was thus no evidence of nonchance differences between the OASR/OABCL scale scores obtained by people who identified themselves as being of non-Latino white, African American, Latino, or other ethnicity.

Multiple Regression Analyses of Adaptive Functioning Scales

To test associations of referral status and demographic variables with scale scores, we used a structural equation modeling (SEM) approach whereby the raw scores for a scale (the dependent variable) were regressed on the independent variables of referral status, gender, age, and education. Referral status and

gender were binary variables, whereas age and education were continuous variables. We entered all the independent variables simultaneously to test the predictive power of each independent variable with the others partialed out.

Referral Status Effects. Nonreferred people obtained significantly (p <.01) more favorable scores than referred people on all three adaptive functioning scales on both the OASR and OABCL. Table 9-2 displays the percentage of variance uniquely accounted for by the significant (p <.01) effects of referral status, with the effects of age, gender, and education partialed out. Cohen's (1988) criteria for effect sizes (ES) in multiple regressions are as follows: Small ES = 2-13%; medium ES = 13-26%; and large ES >26%. The ES for referral status was large for the OABCL Personal Strengths scale and small for the remaining scales. Figure 9-1 graphically displays the mean scores on each adaptive scale.

Demographic Effects. Scores were significantly higher for women than men on the Friends and Personal Strengths scales of both forms. However, as Table 9-2 shows, these gender effects were very small (ES = 1 to 2%) according to Cohen's (1988) criteria. On the Personal Strengths scale, scores were significantly higher for younger than older adults on the OABCL and were significantly higher for people with higher than lower levels of education on both forms. However, these differences did not exceed chance expectations for the number of analyses (Sakoda et al., 1954).

Multiple Regression Analyses of Problem Scales

As we did for the adaptive functioning scales, we used SEM to regress the raw scores of each problem scale on the independent variables of referral status, age, gender, and education. These analyses were done for the 7 syndromes, Total Problems, Critical Items, and 6 DSM-oriented scales.

Referral Status Effects. As Table 9-3 shows, the effects of referral status greatly outweighed the effects of demographic differences on all scales, with referred people obtaining significantly (p<.01) higher scores on all problem scales. The largest effects of referral status accounted for 44% of the variance in scores on the Memory/Cognition syndrome and 40% on the DSM-oriented Dementia Problems scale. Large effects of referral status were also found on the OABCL Total Problems (29% ES), Critical Items

Table 9-2
Percent of Variance Accounted for by Significant (p <.01) Effects of Referral Status and Gender on Adaptive Functioning Scores in Multiple Regressions

	Ref Status[a]		*Gender*[b]	
Scales	***OASR***	***OABCL***	***OASR***	***OABCL***
	N = 440	732		
Friends	8	11	1	1
Spouse/Partner	9	4	—	—
Personal Strengths	11	30	2	1

Note. For Spouse/Partner, N = 196 OASRs and 320 OABCLs. Analyses were multiple linear regressions of raw scale scores on referral status, gender, age, and education. Effects of age and education did not exceed chance expectations according to Sakoda et al. (1954).

[a]All scale scores were significantly (p <.01) higher for nonreferred than referred people.

[b]All significant effects reflected higher scores for women than men.

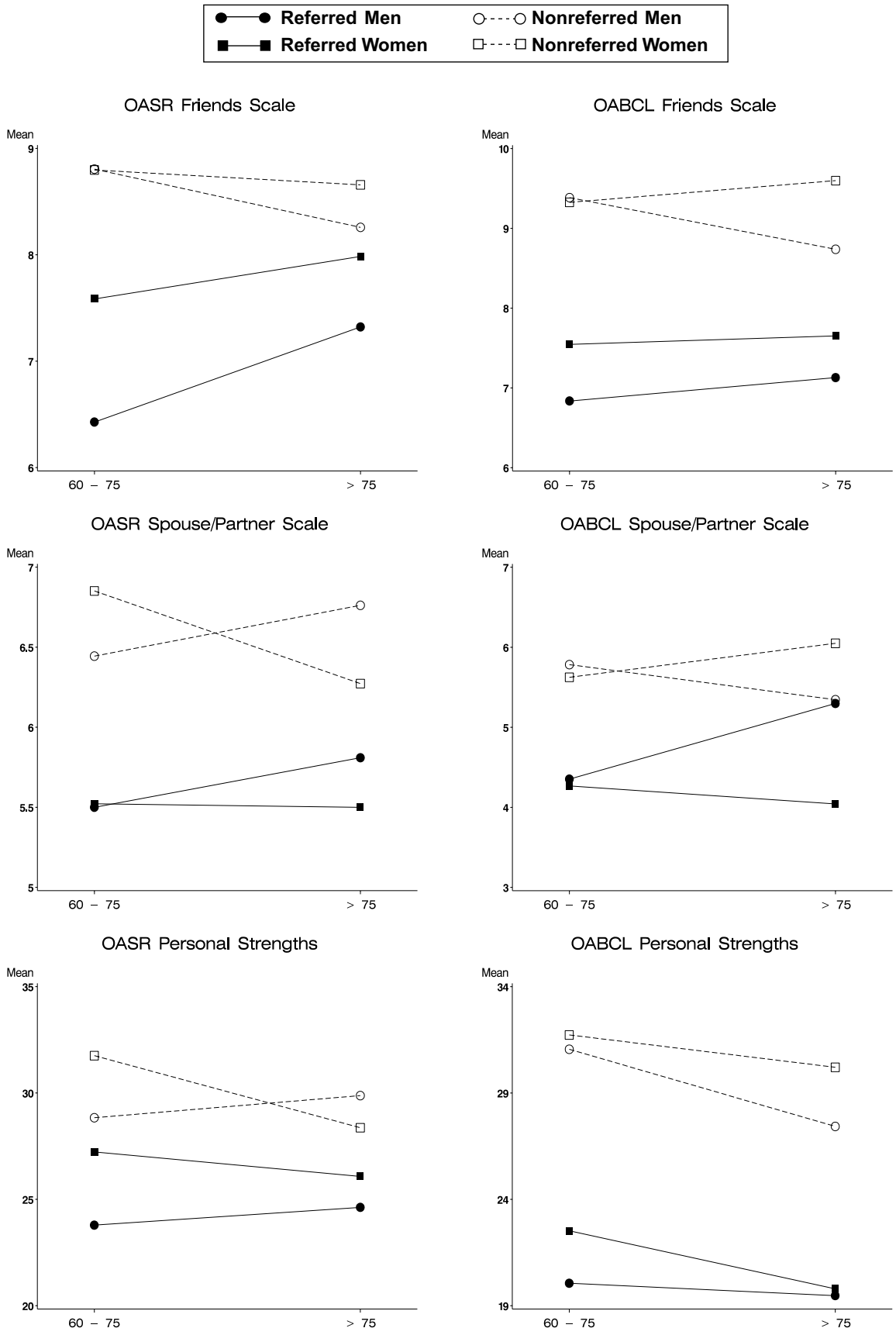

Figure 9-1. Mean scores for adaptive functioning scales.

Table 9-3
Percent of Variance Accounted for by Significant (p <.01) Effects of Referral Status, Age, Gender, and Education on Problem Scale Scores in Multiple Regressions

	Ref Status[a]		*Age*[b]	*Gender*[c]	*Education*[d]	
Scales	***OASR***	***OABCL***	***OABCL***	***OABCL***	***OASR***	***OABCL***
	N = 440	732				
Empirically Based						
Anxious/Depressed	21	25	—	—	1	1
Worries	7	4[e]	—	1[F]	4	3
Somatic Complaints	11	5	—	—	3	1[e]
Functional Impairment	16	26	5	1[Me]	—	—
Memory/Cognition	20	44	1	—	—	—
Thought Problems	11	21	—	—	—	2
Irritable/Disinhibited	2[e]	9	—	—	—	—
Total Problems	20	29	1	—	1[e]	1
Critical Items	24	36	1[e]	—	2	1
DSM-Oriented						
Depressive Problems	23	30	1[e]	—	1[e]	1[e]
Anxiety Problems	20	21	—	—	3	2
Somatic Problems	10	3[e]	—	1[F]	2	—
Dementia Problems	15	40	2	1[Me]	—	—
Psychotic Problems	5	4	—	—	2	—
Antisocial Personality Problems	2[e]	9	—	—	—	—

Note. Analyses were multiple linear regressions of raw scale scores on referral status, age, gender, and education.

[a]All scale scores were significantly (p <.01) lower for nonreferred than referred people.

[b]All significant age effects reflected higher problem scores for older than younger people. The 2 significant age effects on the OASR did not exceed chance expectations.

[c]F = females scored higher; M = males scored higher. There were no significant gender effects on the OASR.

[d]All significant education effects reflected higher problem scores for people with lower educational levels.

[e]Not significant when corrected for number of analyses (Sakoda et al., 1954).

(36% ES), and DSM-oriented Depressive Problems (30% ES) scales. Twelve effects of referral status were medium according to Cohen's criteria, while 13 were small. Figure 9-2 graphically displays the mean scores on each problem scale.

Demographic Effects. As Table 9-3 shows, significant but small effects of age were found on six OABCL scales, all reflecting higher scores for older than younger people. Four 1% ES were found for gender on OABCL scales, with 2 reflecting higher problem scores for females and 2 higher scores for males. Education had significant effects on 9 OASR scales and 8 OABCL scales. All effects were very small, ranging from 1% to 4% ES, and reflected higher problem scores for people with lower levels of education.

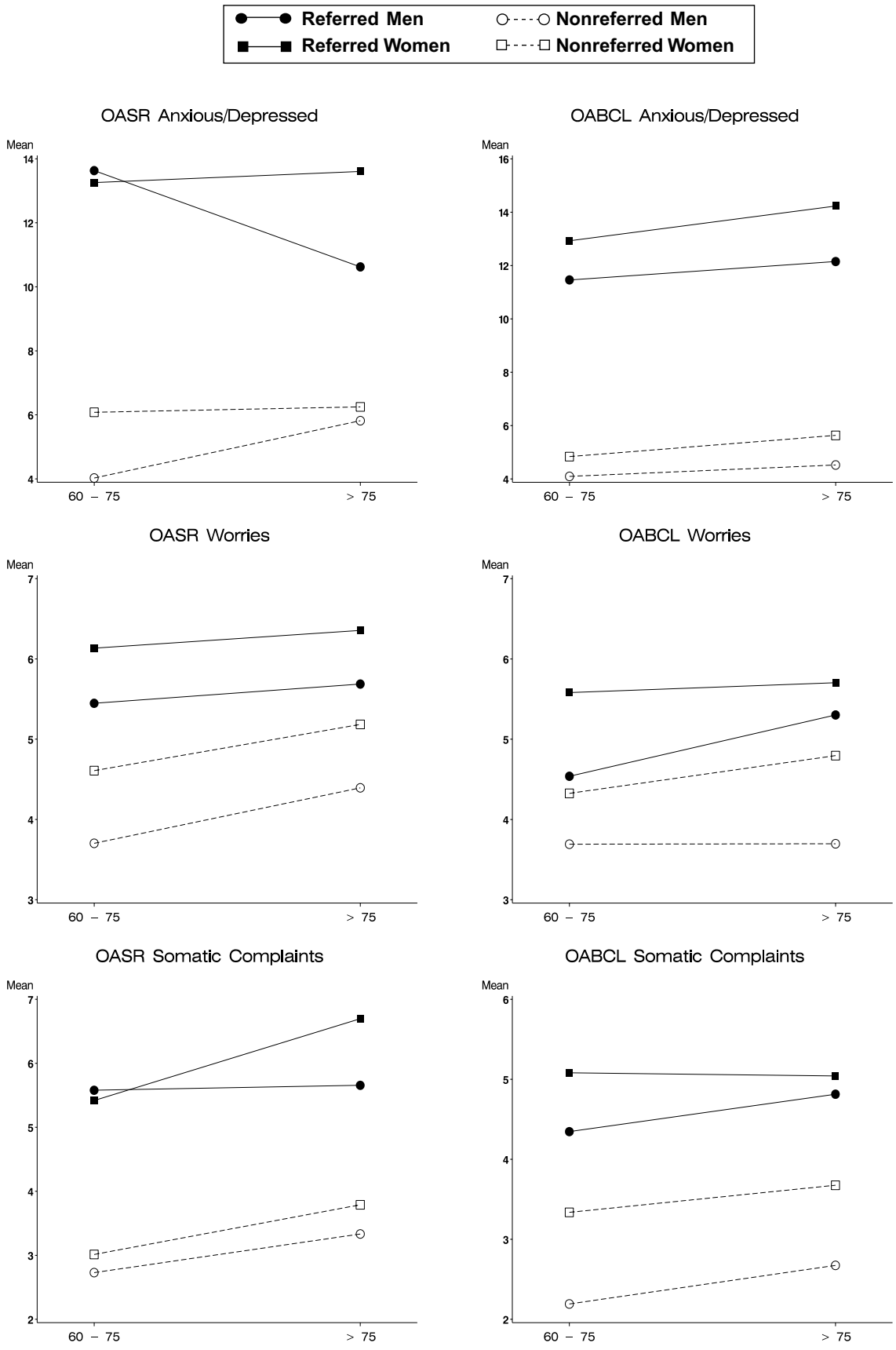

Figure 9-2. Mean scores for problem scales.

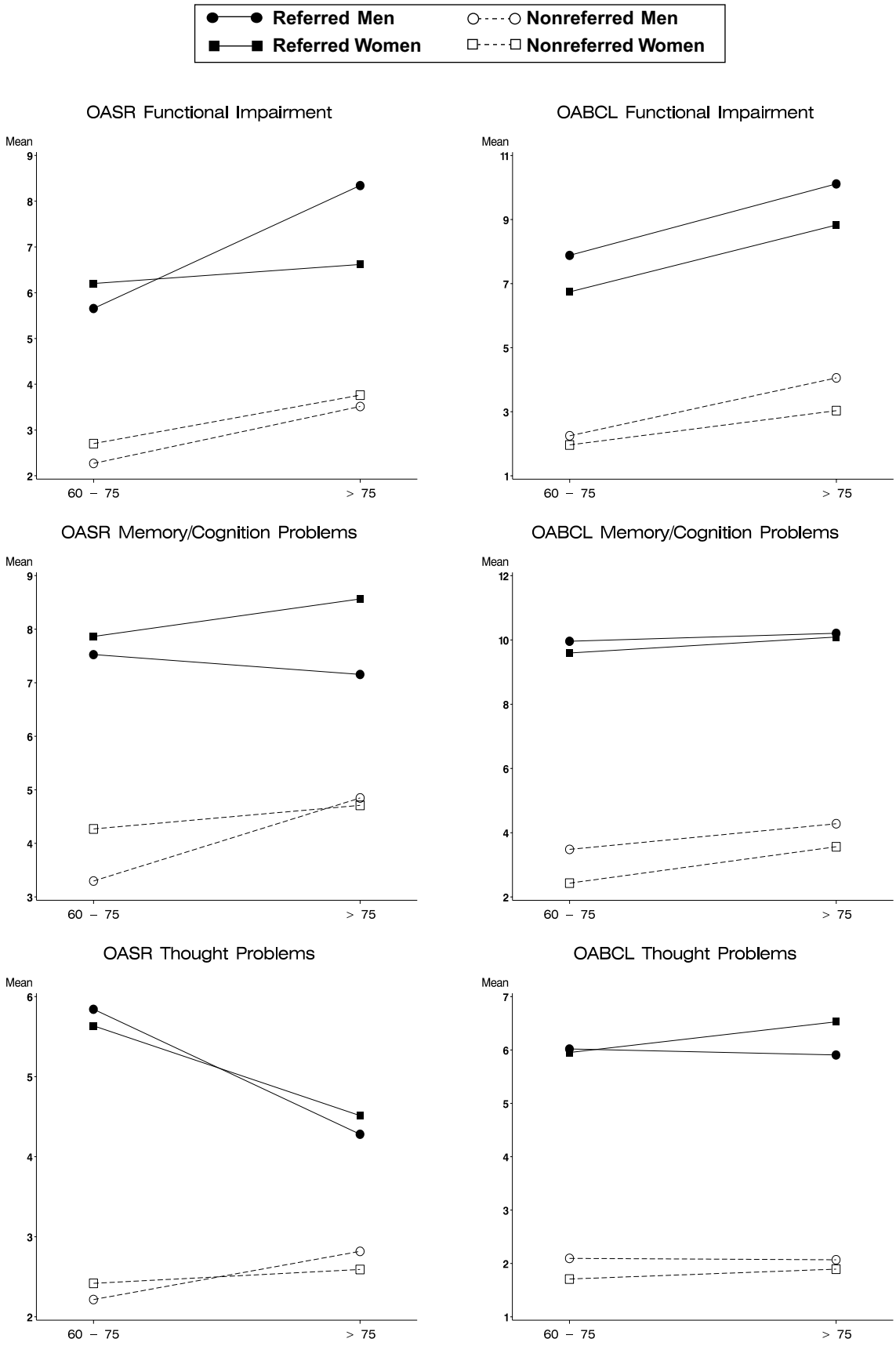

Figure 9-2 (cont.). Mean scores for problem scales.

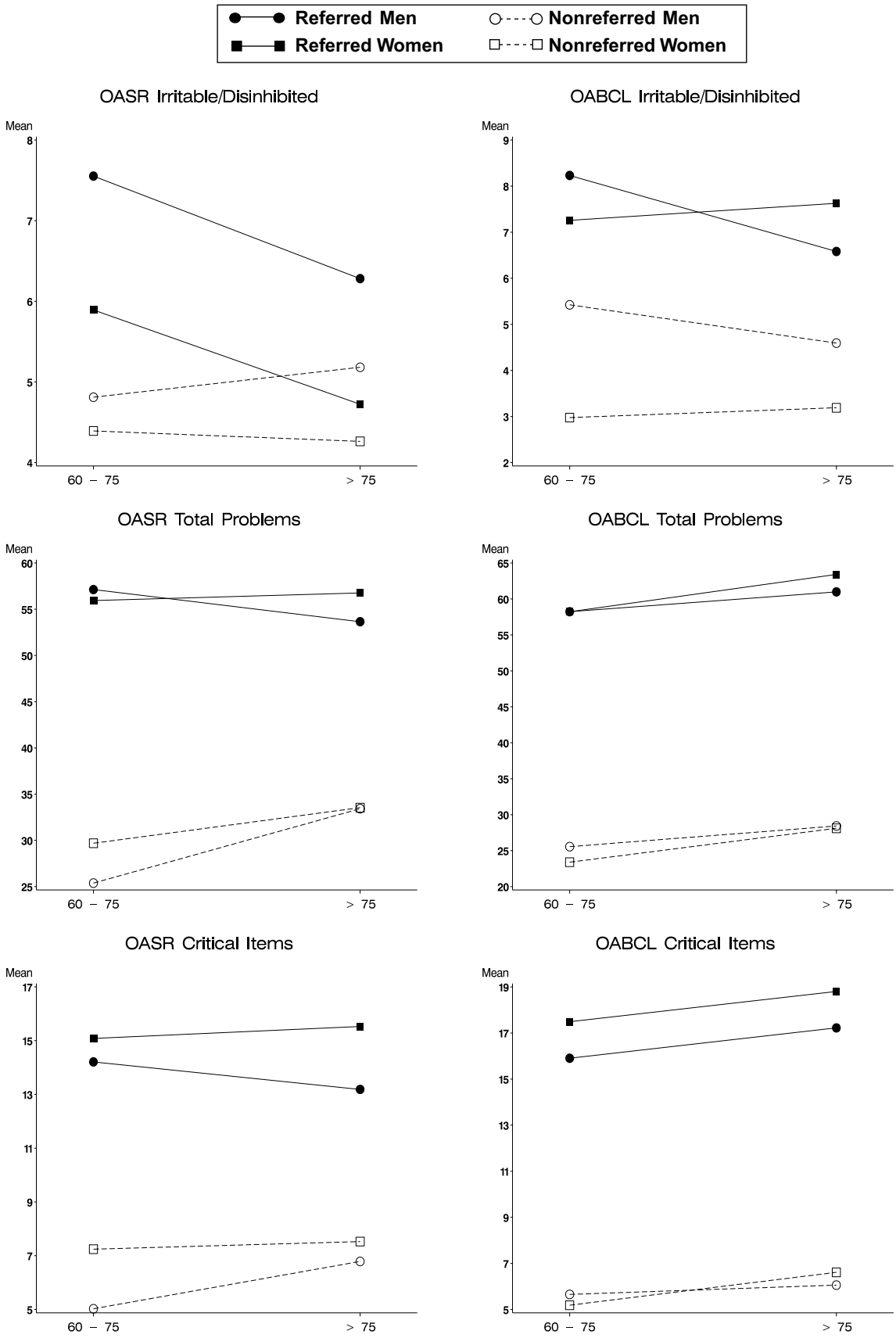

Figure 9-2 (cont.). Mean scores for problem scales.

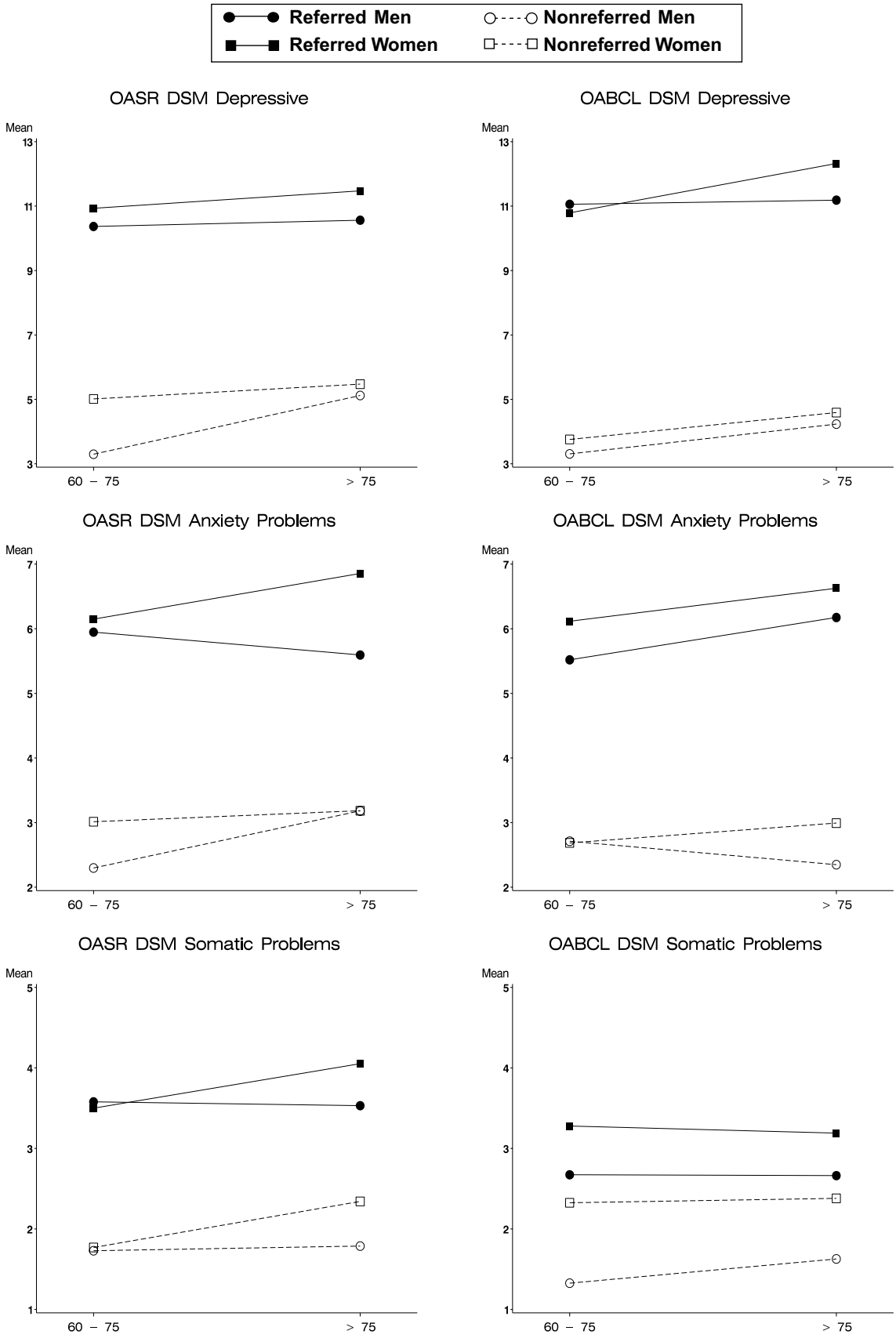

Figure 9-2 (cont.). Mean scores for problem scales.

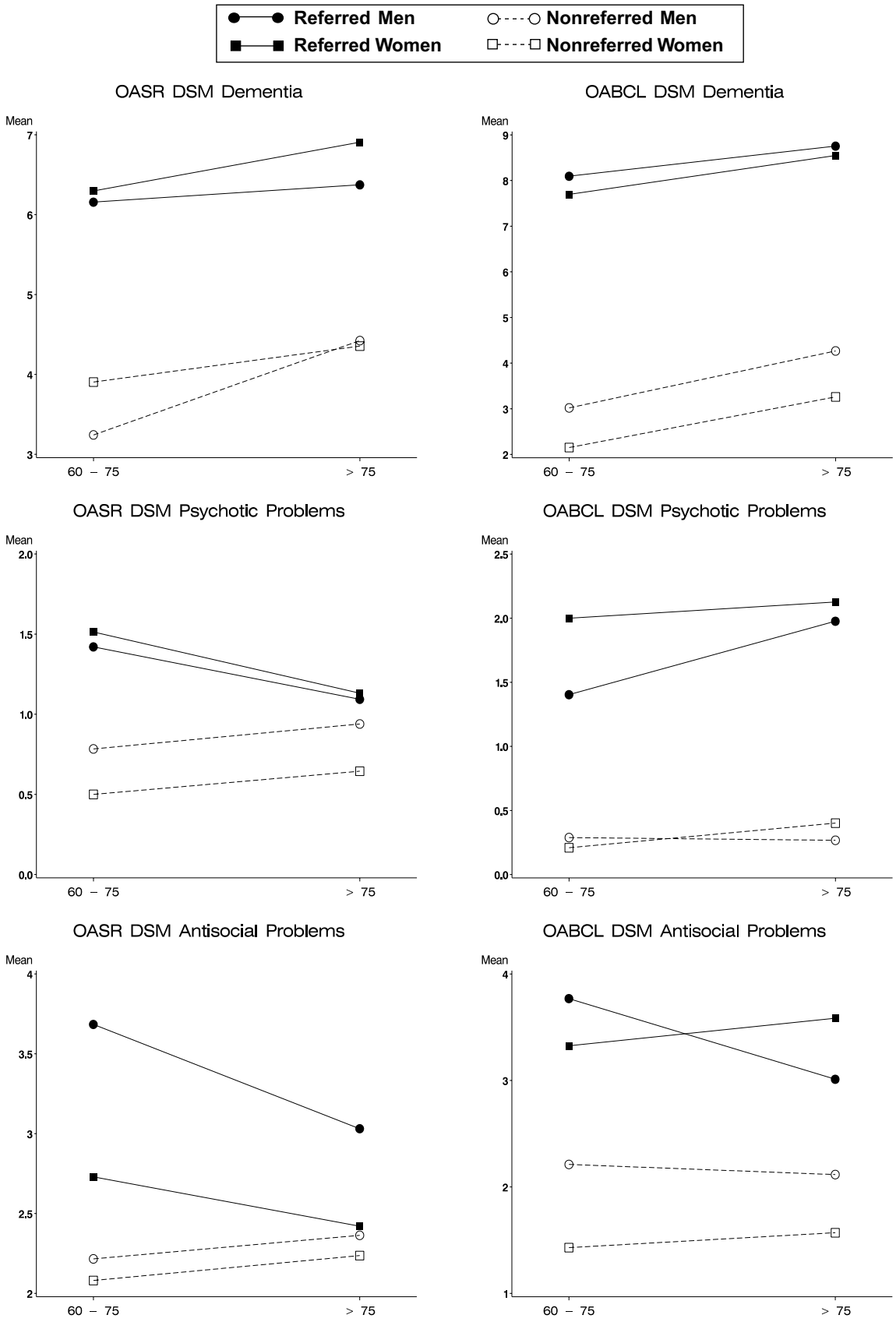

Figure 9-2 (cont.). Mean scores for problem scales.

CLASSIFICATION OF PEOPLE ACCORDING TO CLINICAL CUTPOINTS

The regression analyses reported in the preceding section showed that all quantitative scale scores discriminated significantly ($p<.01$) between referred and nonreferred people. Beside the quantitative scores, each scale has cutpoints for distinguishing categorically between the normal and clinical range. The cutpoints for the different scales were presented in Chapters 6 and 7.

For some clinical and research purposes, users may wish to distinguish between people who are in the normal vs. clinical range according to the cutpoints. Because categorical distinctions are usually least reliable for individuals who score close to the border of a category, we have identified a borderline clinical range for each scale. The inclusion of a borderline category improves the basis for decisions about needs for help.

As an example, a scale score in the borderline range tells us that enough problems have been reported to be of concern but not so many that a person clearly needs professional help. If a person obtains one or more scale scores in the borderline range but none in the clinical range, we should consider options such as the following: ***(a)*** Obtain ratings from more informants to determine whether they view the person as being in the normal, borderline, or clinical range; ***(b)*** have the initial informants rate the person again after 2 to 3 months to see whether the borderline scores move into the normal or clinical range; ***(c)*** use additional assessment procedures to evaluate the kinds of problems comprising the scales that reached the borderline range. In other words, borderline scores can help users make more differentiated decisions than if all scores must be categorized as normal vs. clinical.

Continuous, quantitative scale scores afford greater statistical power than categorization of scores into a couple of levels, such as normal vs. borderline and clinical. Nevertheless, users may wish to distinguish dichotomously between normal and deviant scale scores. In the following sections, we report findings that indicate the degree to which dichotomous classifications of ASEBA scale scores according to the normal range vs. combined borderline and clinical ranges distinguish between demographically similar nonreferred vs. referred people. Because the borderline range encompasses scores that are high enough to be of concern, we have included it with the clinical range for our dichotomous comparisons of deviant scores with scores that are in the normal range.

Odds Ratios (ORs)

One approach to analyzing associations between two dichotomous ways of classifying people is by computing the *relative risk odds ratio* (OR; Fleiss, 1981), which is often used in epidemiological research. The OR indicates the odds that people who have a particular risk factor also have a particular condition (usually a disorder), relative to the odds that people who lack the risk factor have the condition. The comparison between outcome rates for those who do vs. do not have the risk factor is expressed as the ratio of the odds of having the outcome if the risk factor is present, to the odds of having the outcome if the risk factor is absent. For example, a study of relations between smoking (the risk factor) and lung cancer (the outcome) may yield a relative risk OR of 6. This means that people who smoke have 6 times greater odds of developing lung cancer than people who do not smoke.

We applied OR analyses to the relations between ASEBA scale scores and referral status as follows: For each ASEBA scale, we first classified people from our referred and nonreferred samples according to whether their scores were in the normal range or were deviant (i.e., were in the borderline or clinical range). Deviant scores were thus equivalent to a "risk factor" in epidemiological research, whereas referral vs. nonreferral was the outcome. We then computed the odds that people whose scores were deviant on a particular scale were from the referred sample, relative to the odds for people whose scores were not deviant on that scale.

The OR is a nonparametric statistic computed from a 2 x 2 table. For the analysis of each scale scored from each form, we therefore included both genders and all ages to provide a summary OR across all groups

Table 9-4
Significant Odds Ratios and Percent of Older Adults Who Obtained Deviant Scores on Adaptive Functioning Scales

	Odds Ratios		Percent Deviant[a]			
			Referred		Nonreferred	
Scales	OASR	OABCL	OASR	OABCL	OASR	OABCL
	N = 440	732	220	366	220	366
Friends	2	5	15	25	8	6
Spouse/Partner	3	3	18	20	7	8
Personal Strengths	4	12	17	37	5	5

Note. For Spouse/Partner, *N* = 196 OASRs and 320 OABCLs. Odds ratios (OR) indicate the odds that people who obtained scores in the combined borderline and clinical ranges were referred for services. On all scales, the proportion of referred people scoring in the clinical range significantly exceeded the proportion of nonreferred people. All OR and chi squares for 2 x 2 tables were $p < .01$ except the ORs for OASR Friends and Spouse/Partner, which were $p < .05$.

[a]Deviant = combined borderline and clinical range.

for whom the form was scored. The statistical significance of the OR is evaluated by computing confidence intervals.

Adaptive Functioning Scales. Table 9-4 displays the ORs for relations between deviant scores and referral status for the OASR and OABCL adaptive functioning scales. Table 9-4 also shows the percent of referred people whose scores were deviant according to the cutpoints for the normal vs. combined borderline and clinical ranges on each scale. For all OASR adaptive functioning scales, confidence intervals showed that the ORs were significantly ($p<.05$) greater than 1.0, while chi squares for all the scales showed that significantly more referred than nonreferred people obtained deviant scores ($p<.01$). The largest OR was 12 for the OABCL Personal Strengths scale.

Problem Scales. Table 9-5 displays the ORs for relations between deviant scores and referral status for the problem scales. All ORs were significant at $p < .01$. The largest ORs were 27 for the OABCL DSM-oriented Dementia Problems scale and 25 for the OABCL Memory/Cognition Problems syndrome.

Table 9-5 also displays the percent of referred and nonreferred people who had scores in the deviant range on each problem scale of the OASR and OABCL. Chi squares showed that significantly ($p < .01$) more referred people than nonreferred people had scores in the deviant range on every problem scale.

DISCRIMINANT ANALYSES

The foregoing sections dealt with the use of unweighted scale scores to discriminate between referred vs. nonreferred samples of people. Another approach is to use weighted combinations of scores to discriminate between groups. To test this approach, we performed stepwise discriminant analyses in which the criterion groups were the demographically similar samples of referred vs. nonreferred people.

We tested the following six sets of candidate predictors of referral status in the OASR and OABCL samples: ***(a)*** the 7 syndrome scales; ***(b)*** the 6 DSM-oriented scales; ***(c)*** all problem items on a form; ***(d)*** the Friends and Personal Strengths scales; ***(e)*** the Total Problems, Friends, and Personal Strengths scales; and ***(f)*** the problem items, friends items, and personal

Table 9-5
Significant (p<.01) Odds Ratios and Percent of Referred vs. Nonreferred Older Adults Who Obtained Deviant Scores on Problem Scales

			Percent Deviant[a]			
	Odds Ratios		*Referred*		*Nonreferred*	
Scales	*OASR*	*OABCL*	*OASR*	*OABCL*	*OASR*	*OABCL*
Empirically Based	N = 440	732	220	366	220	366
Anxious/Depressed	21	10	41	42	3	7
Worries	4	4	22	19	6	6
Somatic Complaints	4	3[b]	25	17[b]	7	8
Functional Impairment	9	9	38	47	6	9
Memory/Cognition	10	25	41	68	7	8
Thought Problems	5	11	25	46	6	7
Irritable/Disinhibited	2[b]	5	18[b]	25	9	6
Total Problems	6	9	55	66	18	18
Critical Items	13	20	43	58	5	7
DSM-Oriented						
Depressive Problems	14	12	44	47	5	7
Anxiety Problems	9	9	37	44	6	8
Somatic Problems	4	2[b]	27	49[b]	8	9
Dementia Problems	8	27	38	62	7	6
Psychotic Problems	2[b]	10	19[b]	39	9	6
Antisocial Personality Problems	3	4	17	24	7	7

Note. Odds ratios indicate the odds that people who obtained scores in the combined borderline and clinical ranges were referred for services. On all scales, the proportion of adults scoring in the clinical range who were referred significantly exceeded the proportion of people who scored in the normal range. All ORs and chi squares for 2 x 2 tables were significant at p <.01.

[a]Deviant = combined borderline and clinical range.

[b]Not significant when corrected for number of analyses (Sakoda et al., 1954).

strengths items. We did not test the Spouse/Partner scale as a candidate predictor, because it was scored only for people who had lived with a spouse or partner in the preceding 2 months.

Discriminant analyses selectively weight candidate predictors to maximize their collective associations with the particular criterion groups being analyzed, using characteristics of the sample that may differ from other samples. To avoid overestimating the accuracy of the classification obtained by discriminant analyses, it is necessary to correct for the "shrinkage" in associations that may occur when discriminant weights derived in one sample are applied to a new sample.

Cross-Validated Correction for Shrinkage

To correct for shrinkage, we used a "leave-one-out" procedure whereby the discriminant function for each sample was computed multiple times with a different person left out of the sample each time (SPSS, 2003). Each discriminant function was then

cross-validated by testing the accuracy of its classification of each of the "left-out" people as referred vs. nonreferred. Finally, the percentage of correct classifications was computed across all the left-out people. We present the cross-validated percentage of individuals who were correctly classified as referred vs. nonreferred by the discriminant analysis for each set of candidate predictors, i.e., after correction for shrinkage.

Cross-Validated Percent of People Correctly Classified

Table 9-6 displays the cross-validated percent of people who were correctly classified by the discriminant analyses using the six different sets of candidate predictors for each instrument. As you can see in Table 9-6, the percents correctly classified (mean of the percent of True Positives and True Negatives) for the OASR ranged from 63% (Friends and Personal Strengths scales) to 78% (all problem items). For the OABCL, the percents correctly classified ranged from 74% (Friends and Personal Strengths scales) to 86% (all problem items). The largest percent of correct classifications of people as referred vs. nonreferred was thus achieved by using the individual problem items as predictors on both the OASR (78% correctly classified) and the OABCL (86%) correctly classified.

In addition to percents correctly classified, Table 9-6 also displays the percent of true positives, true negatives, false positives, and false negatives, separately for the OASR and OABCL. By looking at Table 9-6, you can see that use of all problem items on the OABCL produced the best results for each of the four classification parameters in addition to total percent correctly classified, as follows: True positives (sensitivity) = 81%; true negatives (specificity) = 90%; false positives = 10%; and false negatives = 19%.

Results for Specific Scales. In the discriminant analyses that used the seven syndrome scales as candidate predictors, the following syndromes survived as significant predictors for both the OASR and OABCL: Anxious/Depressed; Worries; Memory/Cognition Problems; and Irritable/Disinhibited. There was thus considerable consistency in the prediction of referral status from diverse patterns of problems according to self-reports and informants' reports.

For the OASR alone, the Functional Impairment syndrome, and for the OABCL alone, the Thought Problems syndrome also survived as significant predictors.

From the six DSM-oriented scales, the following three survived as significant predictors for both the OASR and OABCL: Depressive Problems, Dementia Problems, and Antisocial Personality Problems. For the OABCL alone, Somatic Problems also survived as a significant predictor. Based on both the syndromes and the DSM-oriented scales, it is thus evident that diverse sets of problems contribute significantly to discriminating between older adults who are deemed to need professional help vs. those who are not.

Results for Specific Problem Items. For the OASR, 9 problem items survived as significant predictors, while for the OABCL, 25 survived. For the OASR, the first item to be entered was *45. I am fearful or anxious*, while for the OABCL it was *110. Has trouble remembering things he/she is told*. There was consistency between the OASR and OABCL analyses with respect to the second item to be entered, which was *93. Unhappy, sad, or depressed.* This item has been repeatedly found to be an especially powerful discriminator between referred vs. nonreferred children, adolescents, and adults, according to different informants in multiple samples (Achenbach, 1991b, d; Achenbach & Edelbrock, 1983, 1986, 1987; Achenbach & Rescorla, 2001, 2003; Verhulst, Akkerhuis, & Althaus, 1985). Even when entered with a very large number of other problem items, this item often obtains considerably larger discriminant function coefficients than nearly all other problem items. As shown in Chapter 10, ANCOVAs of all problem items revealed that item 93 was the second most strongly associated with referral status on the OASR (18% ES) and the sixth most strongly associated with referral status on the OABCL (20% ES).

In addition to item 93, the following four other problem items survived as significant predictors of referral status on both the OASR and OABCL: *39. Does things that others don't like; 69. Trouble making decisions; 72. Worries about his/her family;* and *120. Has trouble remembering things he/she is told.*

Table 9-6
Cross-Validated Percent of Older Adults Classified as Referred vs. Nonreferred by Discriminant Analyses

	OASR					OABCL				
Candidate Predictors	*Correctly Classified*	*TP*	*TN*	*FP*	*FN*	*Correctly Classified*	*TP*	*TN*	*FP*	*FN*
7 syndromes	72	62	83	17	38	82	76	87	13	24
6 DSM-oriented scales	71	62	81	20	38	82	77	87	13	23
All problem items	78	71	85	15	30	86	81	90	10	19
Friends & Personal Strengths scales	63	57	69	31	43	74	69	79	21	32
Friends, Personal Strengths, & Total Problems scales	70	61	78	22	39	76	71	82	18	29
Friends, Personal Strengths, & problem items	75	75	76	25	26	80	74	86	14	26

Note. Numbers in table are percents. TP = true positive rate (sensitivity); TN = true negative rate (specificity); FP = false positive rate; FN = false negative rate.

In summary, the discriminant analyses achieved the best cross-validated accuracy of 86% of people correctly classified as referred vs. nonreferred when selecting from among all the OABCL problem items. The strength of item *93. Unhappy, sad, or depressed* as a predictor on both the OASR and OABCL and its strong discriminative power in analyses of numerous samples of children, adolescents, and adults attest to its association with diverse conditions that warrant professional help across a wide age span.

PROBABILITY OF PARTICULAR TOTAL PROBLEMS SCORES BEING FROM REFERRED VS. NONREFERRED SAMPLES

To provide another perspective on relations between ASEBA scores and referral status, Table 9-7 displays the probabilities that particular Total Problems *T* scores were from our referred samples rather than from our demographically similar nonreferred samples. The probabilities were determined by tabulating the proportion of people from our matched referred vs. nonreferred samples whose Total Problems *T* scores were within each of the *T* score intervals shown in Table 9-7. We used *T* scores in order to provide a uniform metric across both age groups for each gender on the OASR and OABCL.

On the OABCL, the probability that a score was from the referred sample increased fairly consistently as the Total Problem *T* scores increased. However, on the OASR, a relatively large proportion of referred people obtained Total Problems *T* scores <44. It is possible that these people were referred for services because of signs of dementia or other cognitive problems that affected their awareness of their own problems. Culminating with a probability of .95 for *T* scores >67, the highest *T* scores on the OASR were obtained mainly by referred people. However, the finding that low *T* scores were fairly probable among referred people underlines the value of having the OABCL as well as the OASR completed. When a person reports few problems on the OASR but receives high problem scores on one or more OABCLs, this may indicate a lack of awareness of problems that are apparent to others.

Table 9-7
Probability of Total Problems *T* Scores Being from Referred Samples

Total Problems T Scores	*OASR*	*OABCL*
	N = 440	*N* = 732
0-35	.42	.22
36-39	.27	.08
40-43	.41	.15
44-47	.24	.14
48-51	.32	.28
52-55	.34	.38
56-59	.49	.58
60[a]-63	.60	.63
64-67	.68	.79
68-100	.95	.92

Note. Samples were demographically similar referred and nonreferred older adults.
[a]*T* scores ≥ 60 are in the combined borderline and clinical range.

CONSTRUCT VALIDITY OF OASR AND OABCL SCALES

According to a dictionary definition, a *construct* is "an object of thought constituted by the ordering or systematic uniting of experiential elements" (Gove, 1971, p. 489). ASEBA scales can be viewed as representing constructs that have been derived by systematically ordering scores on the items of the ASEBA forms, which tap people's experience pertaining to the individual being assessed.

Construct validity concerns evidence that supports hypothesized variables (*hypothetical constructs*) for which there are no definitive criterion measures. A primary reason for developing ASEBA instruments was to derive syndromal constructs that embody patterns of problems that occur together. Studies of ASEBA child, adolescent, and adult syndromes have revealed numerous correlates and considerable developmental stability for the syndromes (evidence has been reviewed by Achenbach & Rescorla, 2000, 2001, 2003).

The correlates and developmental courses of the different syndromes indicate that they reflect important differences in patterns of child, adolescent, and adult functioning. The validity of constructs such as ASEBA syndromes is supported by the accumulation of evidence for systematic relations between measures of the constructs and other variables. These systematic relations are called *nomological* (i.e., lawful) *networks* (Cronbach & Meehl, 1955).

Each ASEBA syndrome can be viewed in statistical terms as representing a *latent variable* derived by factor analyzing ASEBA items. The versions of a syndrome derived from separate factor analyses of the OASR and OABCL provide ways of operationalizing the construct represented by the syndrome. Furthermore, the versions of a syndrome scored from the OASR and OABCL ratings provide separate quantitative measures of the latent variable represented by that syndrome.

People differ in their knowledge of an individual's functioning, in their roles in relation to the individual being assessed, in what they recall, and in personal characteristics that can affect their ratings. Consequently, the correlations among ratings by different respondents, especially those playing different roles with respect to the individual they rate, may be modest, as shown in Chapter 8. Nevertheless, the test-

retest reliability of OASR and OABCL ratings is good, as documented in Chapter 8. Furthermore, the content and criterion-related validity of ratings by different informants has been documented in the preceding sections of this chapter. The findings thus indicate that different informants can contribute to the assessment process.

Assessment of the syndromal constructs via data from multiple sources is consistent with the way in which psychological constructs are conceptualized and evaluated. Because psychological constructs involve inferences about latent variables that are not directly observable, their validity must be evaluated in terms of various kinds of indirect evidence relevant to their validity. The *Bibliography of Published Studies Using the ASEBA* (Bérubé & Achenbach, 2004) lists some 5,000 published studies of ASEBA instruments. Many of the studies provide evidence for the construct validity of ASEBA scales in terms of significant associations with other variables, prediction and evaluation of outcomes, and consistency with theoretical formulations. In the following sections, we summarize several kinds of support for the construct validity of the OASR and OABCL scales.

Correlations of OABCL Scales with Neuropsychiatric Inventory (NPI) Scales

The Neuropsychiatric Inventory (NPI; Cummings et al., 1994) is widely used to assess psychopathology in patients with dementia on the basis of interviews with informants. Clinical interviewers ask yes/no screening questions for scales that assess delusions, hallucinations, agitation, depression/dysphoria, anxiety, euphoria/elation, apathy, disinhibition, irritability/lability, and aberrant motor behavior. Optional screening questions are also provided for sleep/nighttime behavior disorders and appetite/eating disorders. If the informant responds affirmatively to a screening question, the clinician asks 7 to 8 yes/no questions about specific problems related to the screener. For each category of problem, the clinician then asks questions about the frequency of the problem's occurrence in order to make ratings on a 4-point scale ranging from less than once per week to daily. The clinician also asks about the severity of each category of problem in order to make ratings on a 3-point scale ranging from mild to severe. Scores for frequency are multiplied by scores for severity to yield "domain scores" for each category of problem. The clinician also asks the informant about the disruptiveness or distress that the patient causes for the informant in order to make ratings on 5-point scales from no distress to very severe or extreme distress.

Table 9-8 displays significant ($p<.05$) *r*s between raw scores on OABCL scales and NPI scales (i.e., the number of problem items endorsed on each NPI scale) for 48 patients seen at the University of Vermont/Fletcher Allen Health Care Memory Center in Burlington, Vermont (Brigidi, 2004). The NPI was administered by phone to grown children, spouses, partners, and friends of the patients at an average of 9.5 days ($SD = 2.9$) after they had independently completed paper copies of the OABCL. Table 9-8 omits the NPI scores for euphoria/elation, sleep/nighttime disorders, and appetite/eating disorders because they lack counterparts among the OABCL scales. The NPI total score included in Table 9-8 was the sum of NPI domain scores for the NPI scales, excluding the sleep/nighttime behavior and appetite/eating disorders scales. The NPI distress score included in Table 9-8 was the sum of distress scores for all 12 NPI problem areas.

According to Cohen's (1988) criteria, *r*s <.30 are small ES, .30 to .49 are medium ES, and ≥.50 are large ES. As shown in Table 9-8, 73 *r*s between OABCL and NPI scores met Cohen's criteria for large ES. The highest *r*s were for NPI distress score with the OABCL Total Problems scale ($r = .82$) and with the OABCL Thought Problems syndrome ($r = .81$). The diverse problems encompassed by the Total Problems scale and the particular subset of problems encompassed by the Thought Problems syndrome were thus especially strongly associated with distress experienced by the informants in relation to providing care for the older adults who were assessed. The next highest *r*s were for the NPI total domain score with the OABCL Total Problems scale and Thought Problems syndrome (both *r*s = .79). Thus, whether measured in terms of distress reported by informants or the sum of domain scores across the specific areas assessed by

Table 9-8

Significant ($p<.05$) Correlations of OABCL Scales with Neuropsychiatric Inventory Scores

	Neuropsychiatric Inventory Scores									
OABCL Scales	***Delusions***	***Halluci-nations***	***Agitation***	***Depression***	***Anxiety***	***Apathy***	***Disinhi-bition***	***Irritability***	***Total***	***Distress***
Adaptive										
Friends	—	—	—	—	—	-.53	-.33	-.51	-.37	-.30[a]
Spouse/Partner	-.53	-.47	-.42	-.54	—	-.47	-.46	-.30[a]	-.62	-.66
Personal Strengths	—	-.31[a]	-.31[a]	—	-.36	-.61	-.35	-.44	-.58	-.53
Empirically Based										
Anxious/Depressed	.42	.42	.40	.53	.57	.48	.47	.57	.73	.73
Worries	—	—	.32	.52	.61	.36	—	.46	.58	.58
Somatic Complaints	—	.35	.33	.52	.57	—	—	.51	.50	.49
Functional Impairment	—	.41	.34	.44	.38	.51	.31[a]	.48	.51	.55
Memory/Cognition	.30[a]	.40	.33	.37	.30[a]	.47	.35	.42	.51	.53
Thought Problems	.58	.64	.45	.31[a]	.48	.47	.56	.62	.79	.81
Irritable/Disinhibited	.52	.42	.52	.37	.34	.34	.50	.68	.72	.74
Total Problems	.43	.54	.49	.54	.57	.50	.50	.69	.79	.82
Critical Items	.54	.55	.47	.48	.53	.47	.53	.59	.75	.78
DSM-Oriented										
Depressive Problems	.29[a]	.47	.41	.59	.51	.58	.40	.58	.69	.71
Anxiety Problems	.39	.30[a]	.29[a]	.38	.62	.36	.36	.41	.58	.58
Somatic Problems	—	—	—	.46	.54	—	—	.45	.42	.41
Dementia Problems	.31[a]	.39	.37	.35	.23[a]	.47	.40	.38	.48	.52
Psychotic Problems	.71	.66	.36	—	.38	.30[a]	.45	.43	.64	.66
Antisocial Personality Problems	.47	.46	.29[a]	.38	.62	.36	.36	.67	.71	.74

Note. $N = 48$ except for Spouse/Partner, where $N = 23$. Data were for older adults assessed at the University of Vermont/Fletcher Allen Health Care Memory Center (Brigidi, 2004). The Neuropsychiatric Inventory (Cummings et al., 1994) was administered over the phone to informants by clinical interviewers at an average of 9.5 days ($SD = 2.9$) after the informants had independently completed the OABCL.

[a]Not significant when corrected for the number of analyses (Sakoda et al., 1954).

the NPI, the most global NPI scores were especially strongly associated with the most global OABCL score, plus the Thought Problems syndrome. The OABCL Anxious/Depressed syndrome, Irritable/Disinhibited syndrome, Critical Items scale, and DSM-oriented Antisocial Personality Problems scale also had *r*s in the .70s with both the total and distress scores of the NPI.

Most of the more specific NPI scores had their highest *r*s with the most analogous OABCL syndromes and/or DSM-oriented scales, plus the OABCL Total Problems scale. For example, the NPI delusions and hallucinations scores correlated .71 and .66, respectively, with the OABCL DSM-oriented Psychotic Problems scale, and .58 and .54, respectively, with the OABCL Thought Problems syndrome. The NPI depression score correlated .59 with the OABCL DSM-oriented Depressive Problems scale and .53 with the OABCL Anxious/Depressed syndrome. The NPI anxiety score correlated .62 with the DSM-oriented Anxiety Problems scale, .61 with the OABCL Worries syndrome, and .57 with the OABCL Anxious/Depressed syndrome. The NPI irritability score correlated .68 with the OABCL Irritable/Disinhibited syndrome and .67 with the OABCL DSM-oriented Antisocial Personality Problems scale.

Although the NPI does not provide scores for adaptive functioning, Table 9-8 shows that several NPI problem scores had large negative *r*s with the OABCL adaptive functioning scales. These included *r*s of -.62 and -.66 between the OABCL Spouse/Partner scale and the NPI total and distress scores, respectively; *r*s of -.61, -.58, and -.53 between the OABCL Personal Strengths scale and the NPI apathy, total, and distress scores, respectively; and *r*s of -.53 and -.51 between the OABCL Friends scale and the NPI apathy and irritability scores, respectively.

The many large *r*s between OABCL scales and NPI scores that assess analogous constructs support the construct validity of the OABCL scales. Although the OABCL and NPI data came from the same informants, they were substantially different in format, content, and scoring procedures. Furthermore, the OABCL was self-administered and scored by clerical staff, whereas the NPI was administered by phone at a mean of 9.5 days later by a clinical interviewer who interpreted and scored the responses.

Correlations of OABCL Scales with Measures of Cognitive, Behavioral/Emotional, and Adaptive Functioning

The measures described below were completed by clinicians to assess older adults who were seen at the Geriatric Psychiatry Clinic or the Memory Center of the University of Vermont and Fletcher Allen Health Care (Brigidi, 2004). Two of the measures, the Mini-Mental Status Examination and the Clock Drawing Test, were also administered to older adults who were recruited for the normative sample from senior housing and senior centers, as described in Chapter 6.

Mini-Mental State Examination (MMSE). The MMSE (Folstein et al., 1975) scores global cognitive functioning on a scale of 0 to 30 points, as follows: responses to questions assessing orientation to time and place (10 points); short-term memory for three words (3 points); counting backwards from 100 by 7s (5 points); recall of the three words previously tested for short-term memory (3 points); several brief tests of language (8 points); and copying a figure (1 point).

Clock Drawing Test. In a version developed by Brodaty and Moore (1997), this test measures executive functioning by having the older adult draw a clock face, mark the hours, and draw the hands to indicate 20 minutes to 4 o'clock. The quality of the drawing is scored on a 10-point scale.

The following additional measures were administered to older adults who were seen at the University of Vermont/Fletcher Allen Health Care Memory Center:

Alzheimer's Disease Assessment Scale—Cognitive Subscale (ADAS). In the ADAS (Rosen, Mohs, & Davis, 1984), a trained observer administers 11 items that assess memory, language, and drawing. Performance is scored on scales that are summed to yield a total score of 0 to 70 for severity of impairment.

Geriatric Depression Scale (GDS). In administering the GDS (Yesavage et al., 1983), an interviewer asks 30 yes/no questions that yield total scores of 0 to 30 for depression.

Clinical Dementia Rating (CDR). Developed by Morris (1993), the CDR globally measures the severity of dementia on the basis of interview questions that assess memory; judgment and problem solving; awareness of community affairs; home and hobbies; and personal care. A total severity score of 0 to 3 is obtained by aggregating ratings from the different domains.

Dementia Severity Rating (DSR). The DSR is a modified version of the Global Deterioration Scale (Reisberg, Ferris, de Leon, & Crook, 1982). It assesses stages of dementia ranging from Grade 0 (no significant impairment) to Grade 5 (profound impairment, totally dependent). It was completed by a neuropsychologist on the basis of clinical interviews with patients and informants, plus neuropsychological test data.

Trail Making Test Part A (TRA). The TRA (Partington & Leiter, 1949; Reitan & Wolfson, 1993) is a neuropsychological test of orientation and attention that requires the person to draw lines connecting circles numbered 1 through 15. The score is the number of seconds taken to complete the task. Normed scale scores are available for ages up to 80, but the actual number of seconds was used in the sample reported here, because ages ranged above 80.

Instrumental Activities of Daily Living (IADL). Developed by Lawton and Brody (1969), the IADL is a questionnaire measure of functional ability that is completed by informants. A modified IADL was used that included 31 items pertaining to self-care; household care; employment; recreation; shopping; money; travel; communication; and social relationships. Each item was rated on a scale of 0 for not a problem to 8 for severe impairment. The total impairment score was obtained by summing the scores for the items that were relevant to an individual and determining what percent that score was of the highest possible sum for those items, for a possible range of 0 to 100%.

Table 9-9 displays *r*s between OABCL scales and the foregoing measures. The *r*s in Table 9-9 are generally lower than those that were displayed in Table 9-8 between the OABCL and NPI, which were both based on informants' responses. However, the larger *N*s for the correlations in Table 9-9 warranted use of a $p<.01$ significance criterion, rather than the $p<.05$ significance criterion used for Table 9-8.

Nineteen of the *r*s in Table 9-9 reflected large ES for associations between OABCL scales and the other measures. The largest *r*s were between the OABCL Functional Impairment syndrome, on the one hand, and the IADL ($r = .74$), the CDR ($r = .60$), and the DSR ($r = .60$). The strength of the Functional Impairment syndrome as a measure of diverse aspects of functioning was further supported by correlations of -.53 and -.50 with the MMSE and Clock Drawing Test, respectively. These correlations were negative because high scores on the MMSE and Clock Drawing Test indicate good functioning. The Functional Impairment syndrome also correlated significantly with all the other measures displayed in Table 9-9.

The DSM-oriented Dementia Problems scale had large *r*s with the MMSE, Clock Drawing Test, DSR, and IADL, as well as significant *r*s with all other measures. The Personal Strengths scale had large negative *r*s with the CDR, DSR, and IADL, indicating that this OABCL measure of adaptive functioning tapped in a positive way the construct that these other measures tapped in a negative way. Most OABCL scales had significant *r*s with a majority of the other measures. No OABCL scale had significant *r*s with less than two of the other measures.

Associations Between OABCL Scales and Diagnoses of Alzheimer's Disease (AD)

Associations between OABCL scales and diagnoses of Alzheimer's disease (AD) were tested in a sample that included people assessed at the Memory Center of the University of Vermont and Fletcher Allen Health Care (Brigidi, 2004). The sample also included people who were assessed in senior housing and senior centers as candidates for the normative sample, which was described in Chapter 6. All members of the

Table 9-9

Significant (p<.01) Correlations of OABCL Scales with Cognitive, Behavioral/Emotional, and Adaptive Functioning Measures

OABCL Scales	*MMSE*	*Clock*	*ADAS*	*GDS*	*CDR*	*DSR*	*TRA*	*IADL*
N =	526	528	234	237	189	189	226	204
Adaptive								
Friends	.28	.29	—	—	-.20	—	—	-.29
Spouse/Partner	—	—	—	—	—	—	—	-.32
Personal Strengths	.40	.38	-.24	—	-.54	-.53	-.23	-.51
Empirically Based								
Anxious/Depressed	-.39	-.36	—	.33	.32	.33	.33	.41
Worries	—	—	—	.28	—	—	.27	.32
Somatic Complaints	—	-.11[a]	.30	—	—	—	.31	—
Functional Impairment	-.53	-.50	.34	.26	.60	.60	.29	.74
Memory/Cognition	-.59	-.53	.29	.25	.40	.43	.24	.52
Thought Problems	-.41	-.41	.18[a]	.25	.40	.37	.28	.40
Irritable/Disinhibited	-.27	-.26	—	—	.29	.28	—	.38
Total Problems	-.43	-.42	.18[a]	.33	.42	.42	.30	.56
Critical Items	-.52	-.49	.26	.30	.47	.48	.33	.51
DSM-Oriented								
Depressive Problems	-.41	-.39	.18[a]	.35	.39	.41	.32	.52
Anxiety Problems	-.37	-.33	.16[a]	.30	.32	.31	.31	.36
Somatic Problems	—	—	—	.26	—	—	—	.21
Dementia Problems	-.59	-.55	.34	.21	.46	.51	.24	.58
Psychotic Problems	-.42	-.43	.21	.20	.43	.37	.25	.32
Antisocial Personality Problems	-.25	-.27	—	—	.24	.22	—	.36

Note. MMSE = Mini-Mental State Examination; Clock = Clock Drawing Test; ADAS = Alzheimer's Disease Assessment Scale-Cognitive Subscale; GDS = Geriatric Depression Scale; CDR = Clinical Dementia Rating; DSR = Dementia Severity Rating; TRA = Trail Making Test Part A; IADL = Instrumental Activities of Daily Living. All measures consisted of clinicians' scoring of patients' functioning except the IADL, which was completed by informants. MMSE and Clock data were from members of the normative sample, as well as from Geriatric Psychiatry Clinic and Memory Center patients from whom the other measures were obtained (Brigidi, 2004).

[a]Not significant when corrected for the number of analyses (Sakoda et al., 1954).

sample were assessed with the OABCL and with the MMSE, which is widely used for diagnostic assessment of dementia. The physicians who made the AD diagnoses had access to the MMSE results but not to the OABCL results.

Logistic regressions were used to test OABCL scales in combination with age and education (scored 1 to 9) as predictors of AD diagnoses. In the sample of 489 people, 130 were diagnosed as having AD. Table 9-10 displays the results of logistic regressions of AD diagnoses on OABCL Memory/Cognition Problems syndrome scores and DSM-oriented Dementia Problems scale scores. Separate logistic regressions yielded significant (p<.01) ORs for Memory/Cognitive Problems and Dementia Problems scores (both ORs = 1.7).

It is worth comparing the results in Table 9-10 with results obtained from a study in which physicians made AD diagnoses without knowing the MMSE results. The study was by Tierney, Herrmann, Geslani, and Szalai (2003), who tested the MMSE and the Cambridge Mental Disorders of the Elderly Examination (CAMDEX; Roth et al., 1986) as predictors of AD diagnoses. Like the OABCL, the CAMDEX is scored from information provided by people who know the older adult who is being assessed. Higher correct classification rates were achieved by the OABCL Memory/Cognitive Problems syndrome alone (87% correct) and the DSM-oriented Dementia Problems scale alone (84% correct) than by the CAMDEX (82% correct). These findings suggest that the broad-spectrum OABCL can enhance identification of older adults who have disorders such as AD, even though the OABCL was developed via the bottom-up approach rather than being based on existing diagnostic categories and criteria.

Differential Associations of OABCL Scales with Diagnoses of Alzheimer's Disease (AD) and Affective Disorders

It may be difficult to distinguish between affective disorders and early stages of AD in older adults. To determine whether OABCL scales are differentially associated with affective disorders vs. AD, Brigidi (2004) analyzed scores obtained by people who were diagnosed as having affective disorders or as having AD. OABCL syndrome scores were included with MMSE scores, age, gender, and education (scored 1-9) in multiple discriminant analyses to derive discriminant functions for distinguishing between people with AD diagnoses vs. diagnoses of affective disorders vs. nonreferred people. The sample of 584 older adults included the 130 diagnosed as having AD who were included in the analyses described in the previous section; 342 nonreferred people assessed for the normative sample described in Chapter 6; and 112 diagnosed as having affective disorders at the Geriatric Psychiatry Clinic or the Memory Center. The affective disorders included Bipolar and Major Depressive disorders. Ages ranged from 60 to 97.

Two significant discriminant functions were derived for classifying people according to the AD, nonreferred, and affective disorders categories. The cross-validated classifications (i.e., corrected for shrinkage via the "leave-one-out" procedure) were correct for 83% of the 584 people, including 75% correct for AD, 96% correct for nonreferred people, and 52% correct for affective disorders. Most (83%) of the misclassified people with affective disorder diagnoses were classified in the nonreferred category rather than in the AD category.

In the first discriminant function, by far the strongest predictors for AD were MMSE scores (-.84) and the Memory/Cognitive Problems syndrome (.62). In the second discriminant function, the six remaining OABCL syndromes were significant predictors of affective disorders, with the strongest predictors being the Anxious/Depressed syndrome (.69) and the Functional Impairment syndrome (.64).

In another multiple discriminant function analysis, the OABCL DSM-oriented scales were used in place of the OABCL syndromes. The overall cross-validated classification results were generally similar to the results obtained when the OABCL syndromes were used, as follows: 82% of the 584 people were correctly classified, including 73% correct for AD, 95% correct for nonreferred people, and 53% for affective disorders; 85% of the misclassified people with affective disorders were classified in the

Table 9-10
Associations of OABCL Scales with Diagnoses of Alzheimer's Disease

		Alzheimer's Diagnoses		
Predictor	***OR***[a]	***Sensitivity***	***Specificity***	***% Correct***
Memory/Cognition Problems Syndrome	1.7	83	90	87
DSM-oriented Dementia Problems	1.7	80	88	84

Note. Ages were 60 to 97. $N = 489$.

[a]OR = Odds ratios in logistic regressions that included the OABCL scale, age, and education (scored 1-9). Both OR were significant at $p < .01$.

nonreferred category rather than the AD category. In the first discriminant function, the strongest predictors for AD were MMSE scores (-.86) and the OABCL DSM-oriented Dementia Problems scale (.61). In the second discriminant function, the remaining five DSM-oriented scales were all significant predictors for affective disorders, with the strongest predictors being Depressive Problems (.83) and Anxiety Problems (.65).

Comparisons of the mean scores showed that the following OABCL scales significantly ($p<.01$) discriminated among all three groups: Anxious/Depressed and Memory/Cognition Problems syndromes; DSM-oriented Depressive Problems, Anxiety Problems, and Dementia Problems scales. All other syndromes and DSM-oriented scales significantly ($p<.01$) discriminated at least one group from one or two other groups.

SUMMARY

In this chapter, we presented a variety of evidence for the validity of the OASR and OABCL scores. The *content validity* of the problem items was supported by ***(a)*** their ability to discriminate between referred and nonreferred older adults, and/or ***(b)*** their significant loadings on empirically based syndromes, and/or ***(c)*** their identification by experts as being very consistent with DSM-IV diagnostic categories that are relevant to older adults. All of the adaptive functioning items also discriminated significantly between referred and nonreferred older adults on the OASR and/or OABCL.

The *criterion-related validity* of the older adult scale scores was supported by the significant ($p<.01$) associations of all OASR and OABCL adaptive functioning and problem scale scores with referral status, with demographic effects partialed out. Categorical analyses via odds ratios and chi squares showed that classification of scores as being in the normal range vs. combined borderline and clinical range was significantly associated with referral status for all OASR and OABCL adaptive functioning and problem scales. Discriminant analyses showed that the best cross-validated classification rate (86% correctly classified as referred vs. nonreferred) was achieved by the OABCL problem items. Among the specific problem items that survived as significant predictors, item *93. Unhappy, sad, or depressed* was the second strongest predictor on both the OASR and OABCL. The finding that this item was an exceptionally strong predictor of referral status for older adults is consistent with previous findings for children, adolescents, and younger adults. As shown in Table 9-7, the probability that particular *T* scores for Total Problems were from referred samples can be used to evaluate the likelihood that an individual Total Problems score is high enough to warrant concern. However, the finding that a substantial proportion of referred older adults reported relatively few problems on the OASR indicates the need for having the OABCL completed as well.

No other assessment instruments for older adults span such a broad range of adaptive functioning and problems as the OASR and OABCL. The *construct validity* of ASEBA older adult scales was supported by significant associations with several less broad instruments, including the Neuropsychiatric Inventory (NPI), the Mini-Mental State Examination (MMSE), the Clock Drawing Test, the Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS), the Clinical Dementia Rating Scale (CDR), the Dementia Severity Rating (DSR), the Trail Making Test Part A (TRA), and the Instrumental Activities of Daily Living (IADL). Significant associations were also found with diagnoses of Alzheimer's Disease and with affective disorders. Now that the OASR and OABCL are available for general clinical and research use, their construct validity can be tested in relation to additional measures as well.