Testing the Assumption of Measurement Invariance in the SAMHSA Mental Health and Alcohol Abuse Stigma Assessment in Older Adults

Bellinda L. King-Kallimanis · Frans J. Oort · Nancy Lynn · Lawrence Schonfeld

Abstract This study examined the assumption of measurement invariance of the SAMSHA Mental Health and Alcohol Abuse Stigma Assessment. This is necessary to make valid comparisons across time and groups. The data come from the Primary Care Research in Substance Abuse and Mental Health for Elderly trial, a longitudinal multisite, randomized trial examining two modes of care (Referral and Integrated). A sample of 1,198 adults over the age of 65 who screened positive for depression, anxiety, and/or at-risk drinking was used. Structural equation modeling was used to assess measurement invariance in a two-factor measurement model (Perceived Stigma, Comfort Level). Irrespective of their stigma level, one bias indicated that with time, respondents find it easier to acknowledge that it is difficult to start treatment if others know they are in treatment. Other biases indicated that sex, mental quality of life and the subject of stigma had undue influence on respondents' feeling people would think differently of them if they received treatment and on respondents' comfort in talking to a mental health provider. Still, in the present study, these biases in response behavior had little effect on the evaluation of group differences and changes in stigma. Stigma decreased for patients of both the Referral and Integrated care groups.

B. L. King-Kallimanis (✉) · F. J. Oort
Department of Medical Psychology, Academic Medical Centre, University of Amsterdam,
Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands
e-mail: B.L.Kallimanis-King@amc.uva.nl

F. J. Oort
Department of Education, Faculty of Social and Behavioural Sciences, University of Amsterdam,
Amsterdam, The Netherlands

N. Lynn
School of Aging Studies, University of South Florida, Tampa, FL, USA

L. Schonfeld
Department of Aging and Mental Health Disparities, Louis de la Parte Florida Mental Health Institute,
University of South Florida, Tampa, FL, USA

🖄 Springer

Introduction

Perceived stigma can be a problem for individuals of all ages with behavioral health problems. Research focused on older adults has found that older patients are often reluctant to discuss their behavioral health problems with a health care provider. As with individuals in other age groups, some older adults may reject the diagnosis of depression or another behavioral health disorder due to the stigma attached to it. A psychiatric diagnosis may spark concerns of a potential loss of independence and fears of being institutionalized (Wagenaar et al. 2002). The problem of stigma not only affects the identification of behavioral problems in older adults, but treatment adherence as well. Greater perceived stigma toward individuals with behavioral health problems is associated with a greater likelihood of treatment discontinuation (Sirey et al. 2001).

In 1963, Goffman defined stigma as an "attribute that is deeply discrediting" and reduces the bearer "from a whole and usual person to a tainted, discounted one" (as cited in Bambauer and Prigerson 2006). Perceived or self stigma is the belief that most people will devalue and discriminate against individuals who use behavioral health services or who have a behavioral health disorder (Link et al. 1989). Factors that have been found to be associated with stigma include, sex, with males being more prone to the negative aspects of stigma (Reynolds and Kupfer 1999), though it is possible that females are more prone to feelings of stigma in relation to alcohol use/abuse (Gomberg 1988). Ethnicity and race are associated with stigma though understudied. There is evidence to suggest in relation to stigma associated with alcohol use disorder that perceived stigma is higher for non-White compared with non-Hispanic white race/ethnicity (Keyes, et al. 2010) and in another study compared to Whites, higher for Blacks, and highest for Asians and Hispanics (Smith et al. 2010). Aside from background characteristics, there exists a relationship between stigma and quality of life, where respondents with lower quality of life experience greater perceived stigma (Mechanic et al. 1994).

Specifically, the stigma surrounding behavioral health care has been widely studied in younger populations. Less, however, is known about the impact of stigma on the health care utilization of older adults. The general topic of stigma and the negative outcomes associated with it has received more attention in the past decade following the 1999 Surgeon General's Report on mental health. This report states that stigma is the most formidable obstacle to further progress in the field of mental illness (Substance Abuse and Mental Health Services Administration, Center for Mental Health Services, National Institutes of Health, & National Institute of Mental Health 1999). In 2003, the World Health Organization (WHO) and the Old Age Psychiatry section of the World Psychiatric Association (WPA) jointly distributed a technical consensus paper on the problem of stigma and discrimination against older adults with mental disorders. This report called for further research in the area (Graham et al. 2003).

Despite the call for further research, few instruments to measure stigma, specifically with older adults, have been developed. One questionnaire does exists, however it focuses on bereaved older adults (Bambauer and Prigerson 2006). Currently, there is no general stigma questionnaire for older adults. In light of findings that older adults in need of behavioral health services are acutely affected by perceived stigma, measures developed specifically for this population would be of great use in understanding the problem of perceived stigma. In response to this shortcoming, the research group associated with the Primary Care Research in Substance Abuse and Mental Health for Elderly (PRISM-E) study with funding from substance abuse and mental health services administration (SAMHSA) developed the SAMHSA Mental Health and Alcohol Abuse Stigma Assessment (see Table 1). With older adults in mind, this measure was designed to assess two attributes of stigma: stigma towards mental illness and stigma toward related services, though it is general in nature and could be used with any age sample with appropriate validation. This new SAMHSA Stigma Assessment has been assessed for reliability and validity, but has not undergone any extensive psychometric testing.

To make comparisons of stigma between groups or change in stigma over time, it is import to test whether the questionnaire and items meet the assumption of measurement invariance. When testing this assumption we can use time or group membership as possible violators of this assumption. The assumption of invariance requires that the relationships between the observed items and the latent attribute (theoretical construct) remain constant across measurement occasions and/or groups. In other words, two people with the same level of, say stigma, have the same probability of responding a certain way to the items of the questionnaire, and knowing what group or measurement occasion they are from has no affect on this probability. If this assumption is violated and measurement bias present, compar-

Table 1 SAMHSA mental health and alcohol use stigma assessment items

---

"Some people consider a mental health or an alcohol problem a mark of shame, and others do not. We are trying to find out what older persons feel about this issue. The next few questions ask how you would react <u>if</u> you had to deal with such a problem."

Item 1.   Would you be embarrassed or ashamed if you had a mental health problem?

Item 2.   Would you be embarrassed or ashamed if you had an alcohol abuse problem?

NOTE: If stigma exists ("Not Very" or above) for only mental health or only alcohol, ask remaining questions accordingly.

Item 3.   Do you think people around you would think differently of you if you received mental health or alcohol abuse treatment?

Item 4.   Would it be difficult for you to start mental health or alcohol abuse treatment if other people knew that you were going to be in treatment?

Item 5.   How comfortable would you be talking about your mental health or alcohol abuse problems with your primary care doctor?

Item 6.   How comfortable would you be talking about your mental health or alcohol abuse problems with a counselor or mental health professional?

Item 7.   Would it be difficult for you to obtain treatment for a mental health or alcohol abuse problem in a setting that was clearly identified as a mental health clinic or alcohol treatment center?

---

isons may be invalid regardless of whether the scale has a high Chronbach's alpha or reliability (Mellenbergh 1989; Meredith 1993).

Different statistical methods are available for testing the assumption of invariance with respect to group membership, time, and other violating variables. The method we chose is confirmatory factor analysis (CFA). With CFA we can test the structure of the items in relation to the latent attributes and also test if the items are invariant across groups and/or time. If the results suggest that the assumption of measurement invariance holds, that is, the items are only measuring the attribute of stigma and are not influenced by group specific attributes or changing respondents' response patterns, then the latent means can be confidently compared for true differences or change. There is a large body of literature that has investigated invariance with respect to multiple-groups (Vandenberg and Lance 2000; Schmitt and Kuljanin 2008). Longitudinal invariance, on the other hand has received less attention.

Response shift is one definition for describing a violation of invariance in longitudinal data (Howard and Dailey 1979; Sprangers and Schwartz 1999). In general, response shift can be thought of as a special case of measurement bias (non-invariance) that can occur when investigating change within an attribute of interest (Oort 2005; Oort et al. 2005). A researcher investigating response shift might expect some items of the measurement instrument to be biased, especially when a catalyst between measurement occasions occurs for all respondents (for example, treatment or planned intervention) that is expect to cause change.

In the current study we compared the latent means of stigma across three measurement occasions and two treatment groups. Such comparisons are only valid and meaningful if the assumption of measurement invariance is met. To investigate this assumption, and assess whether the psychometric properties of the Stigma Assessment are invariant over time and across treatment groups we follow the procedure presented in King-Kallimanis et al. (King-Kallimanis et al. 2010). This procedure has previously only been applied to longitudinal data, in the current study we extend the methodology by testing for measurement invariance across both time and groups. We do not expect to meet the assumption of invariance given that respondents are exposed to a catalyst in the form of treatment. The implication of this expectation is that results associated with mean differences could be incorrect. To circumvent incorrect conclusions, we will illustrate how a violation of this assumption can be overcome and meaningful comparisons of stigma can be made and highlight any differences in conclusions had invariance not been tested.

Method

The data used in this study are from the larger Primary Care Research in Substance Abuse and Mental Health for Elderly (PRISM-E) study. PRISM-E was a multi-site, randomized, comparative trial examining two models of care for persons aged 65 and older with symptoms of depression, anxiety, and at-risk drinking. Participants randomized into the integrated behavioral health care group received behavioral health services located in the same practice as their primary care physician. Participants randomized into the enhanced referral care group received behavioral health services at a specialty behavioral health clinic that was geographically separate from their primary care provider (Levkoff et al. 2004).

Measures

SAMHSA Mental Health and Alcohol Abuse Stigma Assessment The questionnaire contains seven items with a five-point response scale with options ranging from "Not at all" to "Extremely"; an additional response option is included to allow patients to respond "Don't know/Refused" (see Table 1). The scale requires administration by an interviewer as the scale utilizes gateway items. That is, the first two items determine the wording of the following five items, and these items focus on the subject of stigma, namely mental health stigma only, alcohol abuse stigma only or both. Two items are reversed scores, so that higher scores are indicative of greater stigma. Limited psychometric information on the Stigma Assessment exists. This is the only published article based on PRISM-E that uses the full Stigma Assessment.

Diagnosis Participants were assessed at all measurement occasions for symptoms of depression, anxiety, and at-risk drinking using a variety of measures (for further details see Levkoff et al. 2004). Anxiety symptoms were assessed using the MINI Panic Disorder Scale, MINI Generalized Anxiety Scale (MINI; Sheehan et al. 1998), and the Beck Anxiety Inventory (BAI) (Beck et al. 1988); for depression symptoms the MINI Major Depression Scale, MINI Dysthymia Scale, MINI Depression History Scale (MINI; Sheehan et al. 1998), and the Center for Epidemiological Studies Depression Scale (CES-D) (Radloff 1977) were used; and for at-risk drinking The Short Michigan Alcoholism Screening Test – Geriatric Version (Blow et al. 1998) were used. In our analysis diagnosis was dichotomized into two variables: mental health symptoms (yes/no) and at-risk drinking (yes/no) and only baseline information was included.

Health Related Quality of Life – The Medical Outcomes Study SF-36 This questionnaire is comprised 36 items that can be summed into the mental health quality of life (MENT HRQoL) and the physical health quality of life (PHYS HRQoL) component scores. These scores range from 0–100, with higher scores indicating higher levels of quality of life. The SF-36 is a well validated scale with good psychometric properties (Ware et al. 1993).

Background Characteristics Age, sex, education and race are included in the analysis. Education was categorized into three groups, 'less than high school', 'high school' and 'more than high school'. Race was also categorized into three groups, 'White', 'Black' and 'Other'. Other is not tested in relation to measurement invariance due to the diversity of this category, it is included so as not to lose power.

Sample

The final PRISM-E study group comprised 2,022 participants aged 65 or older, with a baseline measurement and three follow-up measurements (3, 6 and 12 months from baseline). In the current analysis we do not investigate the measurement occasion at 12 months as the Stigma Assessment was not

included in the research protocol for that assessment period. We excluded participants who missed an interview ($n=539$) or had a missing mental health symptoms or at-risk drinking information ($n=58$), age, or gender. Patients' with missing data on items 1 and 2 of the Stigma Assessment were also excluded as these items determined the wording of items 3–7 ($n=527$). When partial data were present on the SF-36 and/or items 3–7 of the Stigma Assessment, or a "Don't know/refused" response was present, the values were imputed using the expected maximization procedure available in SPSS. These exclusions resulted in a final sample of 1,198 patients.

The final sample was predominately male (75.39%) due to increased participation from Veteran Affairs facilities, and had a mean age of 73.16 (±6.07). Over three quarters (77.22%) of the sample reported mental health symptoms, a third with at-risk drinking (30.65%). Only, 7.87% of the sample experienced both symptoms of mental health and at-risk drinking (see Table 2).

Table 2 Descriptive statistics for respondent characteristics

	Integrated Care Model ($n=594$)	Referral Care Model ($n=604$)	Total ($n=1,198$)
Age (M (SD))	73.20 (±6.16)	73.13 (±5.99)	73.16 (±6.07)
Gender (female)	146 (24.58%)	148 (24.50%)	294 (24.54%)
Race			
White	327 (55.05%)	334 (55.30%)	661 (55.18%)
Black	140 (23.57%)	145 (24.01%)	285 (23.79%)
Other	127 (21.38%)	125 (20.70%)	252 (21.04%)
Education			
Less than High School	248 (41.75%)	247 (40.89%)	495 (14.32%)
High School	135 (22.73%)	124 (20.53%)	259 (21.62%)
More than High School	211 (35.52%)	233 (38.58%)	233 (38.58%)
MH Symptoms (yes)	460 (77.44%)	464 (76.82%)	932 (77.22%)
At-Risk Drinking (yes)	182 (30.64%)	185 (30.63%)	370 (30.65%)
SF-36 - Physical Component Score (M (SD))	39.63 (±10.92)	39.12 (±10.22)	39.37 (±10.57)
SF-36 - Mental Component Score (M (SD))	42.95 (±13.04)	41.56 (±13.15)	42.25 (±13.11)
Embarrassed / ashamed if you had a mental health problem (BL)	332 (55.89%)	330 (54.64%)	667 (55.26%)
Embarrassed / ashamed if you had a alcohol abuse problem (BL)	391 (65.82%)	384 (63.58%)	782 (64.79%)
Embarrassed / ashamed if you had a mental health problem (3mth)	295 (49.66%)	284 (47.02%)	579 (48.33%)
Embarrassed / ashamed if you had a alcohol abuse problem (3mth)	369 (62.12%)	340 (56.29%)	709 (59.18%)
Embarrassed / ashamed if you had a mental health problem (6mth)	269 (44.44%)	266 (44.04%)	530 (44.24%)
Embarrassed / ashamed if you had a alcohol abuse problem (6mth)	327 (55.05%)	306 (50.66%)	633 (52.84%)

Statistical Analysis

The analysis is an application of methods presented in King-Kallimanis et al (2010). However, we extend these methods to include not only the longitudinal data structure, which includes three measurement occasions, but also multiple groups, the integrated behavioral health care group and the enhanced referral care group.

This procedure follows three main steps that require fitting a series of structural equation models (SEM) and assessing their fit. To achieve this, the maximum likelihood estimation method is used, while the items are not continuous, previous work suggests that with a 5 point Likert scale, estimates will not be too affected (Dolan 1994). Three goodness-of-fit statistics are used to assess the model, the chi-square test of exact fit, the root mean square error of approximation (RMSEA) and the expected cross-validation index (ECVI). A non-significant value of the chi-square test indicates good model fit; however the test is sensitive to small deviations between the data and the model. Therefore, we also consider the RMSEA, where as a rule of thumb, values of less than .05 suggest close fit and values less than .08 suggest satisfactory fit (Browne and Cudeck 1992). And also the ECVI, which is not a stand alone index, but can be used when comparing nested models, as we do in our paper. A smaller value suggests improved overall model fit (Browne and Cudeck 1992).

To assess the appropriateness and significance of changes made to the model we use the chi-square difference test, with a Bonferroni correction, which tests the difference in the fit of the null and alternative models. When using the chi-square difference test to assess the invariance of the items, we use global tests, where all factor loadings and intercepts of the items are simultaneously tested (King-Kallimanis et al. 2010). In the presence of a high level of power an additional method is required to assess the significance of model modifications. We therefore consider standardized observed parameter changes (OPC) as a means to guard against chance findings in Step 2 and Step 3 of the procedure. The OPC is the difference between the standardized parameter of interest (factor loading or intercept) in the null model versus the same standardized parameter in the alternative model. These are scaled for comparison, when a factor loading OPC is greater than 0.1 or an intercept OPC is greater than 0.2 (Cohen 1988) and the chi-square difference test is significant we consider the parameter possibly biased. All analyses are conducted using Mx32 (Neale 2004).

Procedure

Step 1:   Establishing a measurement model.
          SEM was used to fit an appropriate confirmatory factor model using two (integrated behavioral health care and enhanced referral care) variance-covariance matrices for the five items of the Stigma Assessment scale at three measurement occasions. For both groups and all measurement occasions the same factors and pattern of factor loadings were modeled and all parameters were free to be estimated. To assess areas of misfit the standardized residuals were investigated. Any model changes were assessed using the chi-square difference test. Before moving to Step 2, we required that the final model had satisfactory fit as determined by the goodness of fit statistics.

Step 2:   Testing measurement invariance across measurement occasions and groups.

To test measurement invariance, the final model from Step 1 was used and the factor loadings and intercepts were simultaneously constrained to be equal at all time points and across both groups. A strong indicator of measurement bias is a significant deterioration in the fit of this fully constrained model when compared to the model where all parameters are fully estimated. However, if the model does not significantly deteriorate, we still investigate the possible presence of measurement bias. This is because one biased parameter at one time point, in one group may not cause significant deterioration in the overall model fit, yet may be substantially important. In order to investigate whether any observed items(s) were biased the equality constraints imposed were removed and assessed using a series of global tests. In the first iteration there were five global tests, one for each item. In the first test, the equality constraints of the factor loadings and intercepts of the first item are removed and the model compared with that of fully constrained model using the chi-square difference test and the OPCs. Once all five tests are conducted, if there is an item with a significant chi-square difference test and at least one OPC meeting our criteria, this item's equality constraints are permanently removed. In the presence of more than one item meeting these criteria, the item associated with the largest improvement in overall model fit will be considered biased. This procedure is continued iteratively, until no large OPCs remain that are associated with a significant chi-square difference test. Any biases identified will be further explored using the partitioning formula of Oort (2005) and the size of the effects will be assessed using Cohen's *d* effect sizes (Cohen 1988).

Step 3:   Testing measurement invariance with respect to exogenous variables.

Invariance, in this step, was tested by extending the final model in Step 2 to include additional exogenous variables which may induce bias on the observed items. In theory the relationship between the exogenous and observed items should be fully explained via their relationships with the latent variable, if this does not hold true, bias has been found. We included the subject of stigma (Items 1 and 2) of the Stigma Assessment at all measurement occasions, age, gender, race, education, mental health symptoms and at-risk drinking and the Physical and Mental HRQoL component scores of the SF-36. In this model these exogenous variables were correlated with the latent constructs, at all three measurement occasions, and their relationship with the observed items, should be explained via these correlations.

Like in Step 2, we used global tests to investigate the presence of measurement bias between the exogenous variables and observed items. For each exogenous variable we conducted five global tests. This involved adding additional parameters to the model both longitudinally and for both groups. When all direct effects were tested, any effects producing a significant chi-square difference test and large OPCs were included in the model. In the presence of more than one effect meeting these criteria, the effect that produced the largest improvement to the overall model fit was

included in the model. This process continued in an iterative manner until no more large OPCs with a significant chi-square difference test were found.

Results

Step 1:   Establishing a measurement model.

As the structure of the Stigma Assessment had not previously been investigated, we hypothesized two possible measurement models. A uni-dimensional measure of stigma and, based on the content of the items, a two-dimensional measure. We first tested the uni-dimensional model with all five items loading on the single latent variable of stigma (Model 1, Table 3). The chi-square test was significant and the RMSEA also suggested poor fit. Next, we tested the two-dimensional model. The chi-square test was still significant but the RMSEA suggested satisfactory fit (Model 1.F).

The two dimensions were defined as: 1) Perceived Stigma: which included items measuring level of embarrassment and the perception of a negative response from others regarding seeking behavioral health treatment and 2) Comfort Level: which included items measuring the respondents' level of comfort in talking with a service provider about their own mental health or alcohol abuse problems. This final measurement model is depicted in Fig. 1.

Step 2:   Testing measurement invariance across measurement occasions and groups.

In this step all factor loadings and intercepts were constrained to be equal across the three measurement occasions and the two groups. The fit of this model significantly deteriorated when compared to the unconstrained measurement model (χ^2diff(30)=83.61, $p<0.001$).

When investigating bias using global tests and OPCs, we found that the removal of the equality constraints for Item 4 (referring to difficulty in starting treatment if other people knew) led to large OPCs and a significant global test (improvement in model fit) (χ^2diff(10)=36.21, $p<0.001$). No more significant improvements were found.

We conducted a post hoc analysis to see whether the bias was associated with time, group membership or both. The bias is considered to be response shift as the parameters are invariant across groups, but not across measurement occasions. The bias seems predominately associated with the intercepts, with the intercept at each measurement occasion increasing relative to Perceived Stigma. This suggests that when Perceived Stigma decreases over time, Item 4 did not decrease as much as we would have expected, this is regardless of treatment groups.

Step 3:   Testing measurement invariance with respect to exogenous variables.

In this step we used the final model from Step 2 and included the additional exogenous variables. In this model (Model 3) these twelve exogenous variables were correlated with the latent constructs, Perceived Stigma and Comfort Level at all three measurement occasions. The largest

Table 3 Overall goodness-of-fit and chi-square difference test (n=1198)

Model		CHISQ (df)	RMSEA (90% conf. int.)	ECVI (90% conf. int.)	Comparison models	CHISQ DIFF (df)	p	ECVI DIFF (90% conf. int.)
1	Uni-dimensional measurement model	1451.1 (144)	0.123 (0.117 ; 0.129)	1.430 (1.329 ; 1.536)				
1.F	2 factor measurement model	380.0 (120)	0.060 (0.053 ; 0.067)	0.575 (0.529 ; 0.628)				
2	Across occasion constraints on factor loadings and intercepts	463.6 (150)	0.059 (0.053 ; 0.065)	0.594 (0.542 ; 0.652)	2 vs. 1.F	83.6 (30)	<0.001	0.017 (−0.003 ; 0.044)
2.1	Item 4 factor loadings and intercepts free	427.4 (140)	0.058 (0.052 ; 0.065)	0.581 (0.531 ; 0.636)	2.1 vs. 2	36.2 (10)	<0.001	0.012 (−0.001 ; 0.032)
2.F	Item 4 treatment equal across groups, time varies	430.2 (146)	0.057 (0.051 ; 0.063)	0.573 (0.523 ; 0.638)	2.F vs 2.1	2.8 (6)	0.833	−0.008 (−0.005 ; −0.002)
3	As Model 2.F but with addition of exogenous variables	1270.6 (426)	0.058 (0.054 ; 0.061)	2.171 (2.083 ; 2.266)				
3.1	Item 2 on Item 3 at all time points	1151.9 (420)	0.054 (0.050 ; 0.058)	2.082 (1.999 ; 2.172)	3.1 vs. 3	118.7 (6)	<0.001	0.087 (0.059 ; 0.123)
3.2	MENT HRQoL – Item 3	1067.4 (414)	0.051 (0.048 ; 0.055)	2.022 (1.943 ; 2.108)	3.2 vs. 3.1	84.5 (6)	<0.001	0.058 (0.034 ; 0.090)
3.3	MENT HRQoL – Item 6	1010.9 (408)	0.050 (0.046 ; 0.054)	1.985 (1.909 ; 2.069)	3.3 vs. 3.2	56.5 (6)	<0.001	0.034 (0.015 ; 0.061)
3.4	Item 1 on Item 3 at all time points	967.9 (402)	0.048 (0.045 ; 0.052)	1.960 (1.886 ; 2.042)	3.4 vs. 3.3	43.0 (6)	<0.001	0.025 (0.010 ; 0.047)
3.5	Gender – Item 3	926.9 (396)	0.047 (0.043 ; 0.051)	1.936 (1.864 ; 2.016)	3.5 vs. 3.4	41.0 (6)	<0.001	0.022 (0.008 ; 0.045)

df degrees of freedom; *conf. int.* confidence interval; *RMSEA* root mean square error of approximation; *ECVI* expected cross-validation index; *CHISQ* chi-square test; *CHISQ DIFF* chi-square difference test
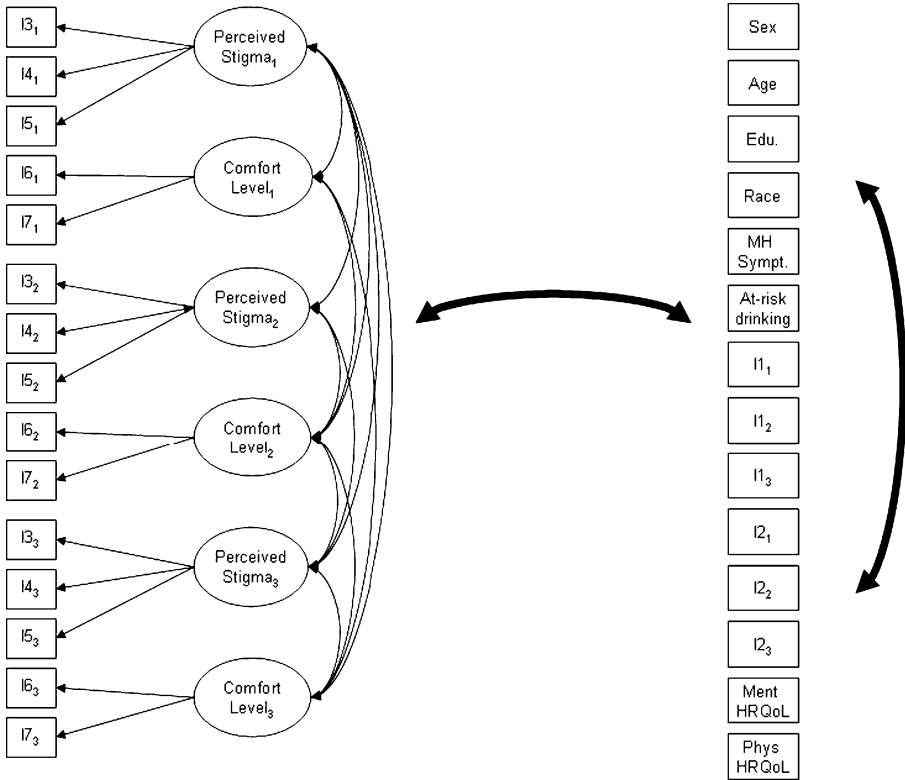
Fig. 1 SAMHSA Mental Health and Alcohol Abuse Stigma Assessment measurement model for three measurement occasions and one group. Note - Abbreviations – Edu. – Education; MH Sympt. – Mental health symptoms; Ment HRQoL – Mental health-related quality-of-life; Phys HRQoL – Physical health-related quality-of-life

correlations were between both subjects of stigma (Item 1 and 2) and Perceived Stigma (range: 0.36–0.20). There were also small negative correlations between MENT HRQoL and Perceived Stigma and Comfort Level. A similar correlation pattern was seen for PHYS HRQoL. The correlations between age, sex, education, mental health symptoms and at-risk drinking and the latent constructs were very small (<0.01). Finally, the correlations between race and both Perceived Stigma and Comfort Level were negative and small (range: −0.23– −0.08).

Using Model 3 as the reference model a total of five biases were found, four of these biases were related to Item 3 of the Stigma Assessment, the other bias was related to Item 6. In the sixth iteration no additional significant effects were found.

Ad hoc analyses indicated that four of the five biases detected were constant across time and over treatment groups. The remaining bias was considered response shift as it was biased over time, but not across groups. The biases of Item 3 with respect to the two subjects of stigma (Item 1 and 2), gender and Mental HRQoL were in the same direction as the correlations and suggest that: people embarrassed by mental health

reported greater stigma, people embarrassed by alcohol abuse reported greater stigma, females reported greater stigma and people with worse Mental HRQoL reported greater stigma, than would be expected given their "true" stigma scores (i.e., common factor scores). Mental HRQoL and Comfort Level, were negatively correlated, as higher scores on Comfort Level are indicative of greater un-comfort. The fifth bias on Item 6 with respect to Mental HRQoL was in the same direction as this correlation, and suggested that people with better Mental HRQoL reported less un-comfort speaking with a mental health profession than would be expected given their "true" Comfort Level score (see Table 4 for selected final parameter estimates).

Measuring True Change in the Stigma Assessment

After accounting for the bias found in Step 2 and 3 we assessed the impact of these biases on Perceived Stigma and Comfort Level, both between treatment groups and longitudinally. The effect sizes (Cohen 1988) of true change for the biased Item 4 is

Table 4 Selected parameter estimates for the final model after all biases were accounted for

Stigma Assessment	Intercepts for three measurement occasions	Factor loadings Perceived Stigma for three measurement occasions	Factor loadings Comfort Level for three measurement occasions
Item 3			
Baseline – 6mth	1.49 (1.41 ; 1.57)	0.41 (0.35 ; 0.48)	0
Item 4			
Baseline	0.89 (0.81 ; 0.98)	0.92 (0.82 ; 1.02)	0
3mth	0.92 (0.81 ; 1.05	0.97 (0.84 ; 1.12)	0
6mth	1.04 (0.92 ; 1.16)	0.90(0.78 ; 1.03)	0
Item 5			
Baseline – 6mth	1.16 (1.08 ; 1.23)	0	0.81 (0.75 ; 0.89)
Item 6			
Baseline – 6mth	1.23 (1.16 ; 1.31)	0	0.79 (0.71 ; 0.86)
Item 7			
Baseline – 6mth		1.18 (1.10 ; 1.26)	0.73 (0.64 ; 0.82)
Exogenous Variables		Integrated Care	Referral Care
Item 2 on Item 3 at all time points		0.12 (0.08 ; 0.17)	0.12 (0.08 ; 0.17)
MENT HRQoL – Item 3			
Baseline		−0.20 (−0.27 ;−0.13)	−0.20 (−0.24 ;−0.09)
3 month		−0.14 (−0.20 ;−0.07)	−0.14 (−0.20 ;−0.07)
6 month		−0.24 (−0.30 ;−0.18)	−0.24 (−0.30 ;−0.18)
MENT HRQoL – Item 6		0.12 (0.08 ; 0.15)	0.12 (0.08 ; 0.15)
Item 1 on Item 3 at all time points		0.15 (0.11 ; 0.19)	0.15 (0.11 ; 0.19)
Gender – Item 3		0.16 (0.11 ; 0.21)	0.16 (0.11 ; 0.21)

Fig. 2 a Perceived Stigma
mean change by group and
between models. **b**. Comfort
level mean change by group
and between models
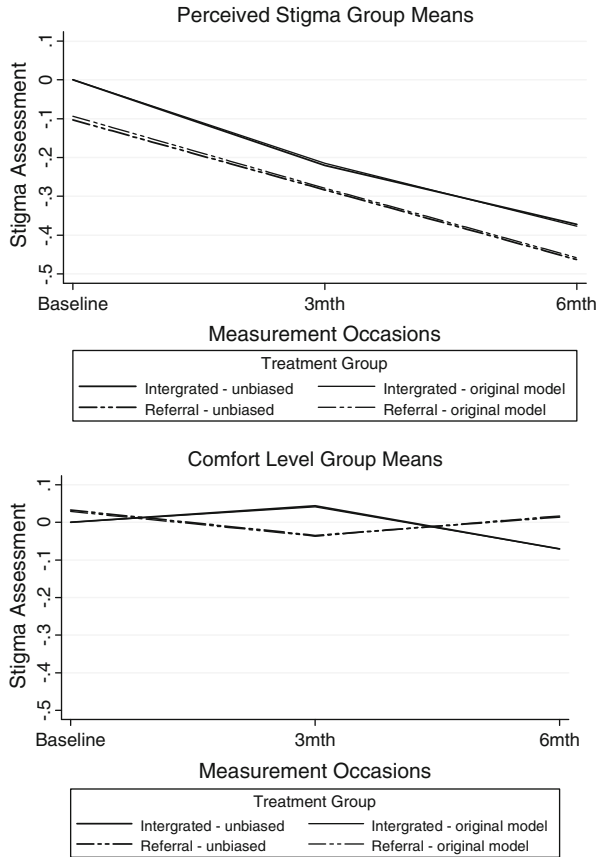
small/medium, −0.29 and −0.28 for the integrated and referral groups. Not taking into
consideration the effect of bias, we would have concluded that the effect size for change
was of small magnitude -.18 for the integrated group and -.15 for the control group,
however a large amount of this change is attributable to bias (0.11 in the integrated group
and 0.12 in the referral group). The biases associated with the exogenous variables had
little impact on their correlations with Perceived Stigma and Comfort Level.

When we consider the latent means of Perceived Stigma and Comfort Level we
see in Fig. 2a and b that there is little difference between the means of the original
model, before bias has been accounted for, and the means after all biases have been
included at the end of Step 3.

Discussion

In this study we found a measurement model that included two latent constructs,
Perceived Stigma and Comfort Level. This suggests that the Stigma Assessment
Scale is not uni-dimensional and therefore when considering change in stigma, the
five items should not be summed together to make a total Stigma score. In addition,
six biases were found when investigating measurement invariance in the longitudinal/

multi-group data of the Stigma Assessment Scale. However, despite these findings, the substantive impact of accounting for these biases on stigma over time and across treatment groups in our analysis was small.

In Step 2, when assessing bias with respect to measurement occasions and groups, we found response shift. This was found in the item assessing whether it would be difficult for the respondent to start treatment if others knew they were to have treatment. In general, overall Perceived Stigma decreased over the course of the study, however Item 4 did not decrease as much as we would have expected given the overall decrease in Perceived Stigma. Apparently, in comparison with their overall Perceived Stigma, respondents initially are reserved when answering this item. However, after treatment they feel more open to report their true level of stigma on this item, in other words, they have recalibrated the response scale anchors. The impact of this bias on substantive conclusions was small. Without considering this bias, we would have erroneously concluded that there was small change in this item over time. After the bias was accounted for in this item, we actually see that there is very little true change and most of the observed change seen is attributable to the bias. That is, respondents' stigma towards starting treatment if others knew, did not change, regardless of treatment group.

When testing the assumption of measurement invariance, a limitation in using SEM is that we are only able to identify biased items when they are in the minority. Perceived Stigma is represented by three observed items; with one biased item, this constitutes a third of the items. Even though this is a minority it is still difficult to be confident that the bias observed is truly associated with Item 4 (difficulty in starting treatment if other people knew). If a bias had been associated with an item measuring Comfort Level, with just two observed items, we would not be able to tell which of the two items were truly biased.

In Step 3, when testing measurement invariance with respect to exogenous variables, we found five examples of bias. Four of these biases were related to Item 3 that assessed whether the respondents thought people would think differently of them if they received treatment. This item, which should primarily measure Perceived Stigma, is particularly vulnerable to bias as it is also significantly indicative of other respondent characteristics. This may be due to the very general wording of the item. The biases associated with this item are, both subjects of stigma (mental illness and alcohol abuse), gender and Mental HRQoL. The effect of Mental HRQoL was in the same direction for both groups, but was significantly larger in the integrated care group. The other bias was associated with the comfort the respondent feels in talking with a mental health counselor. Respondents with better Mental HRQoL reported greater comfort speaking with a mental health professional than those with low Mental HRQoL, even if their Comfort Levels are similar.

As discussed above the biases found were interpretable; however, accounting for these biases had little impact on the interpretation of the final results of the Stigma Assessment in this example. Comfort Level did not differ significantly between treatment groups or significantly change across measurement occasions. Perceived Stigma declined over the course of treatment in both groups. The rate of change was not significantly different between treatment groups; however, there was a significant decrease over the course of treatment. As it turns out stigma may be more difficult to change than has been previously thought. After the Surgeon General's report on

stigma (Substance Abuse and Mental Health Services Administration et al. 1999) there was a strong public relations effort to inform the public about the underlying neurobiology of mental illnesses. In a study that assessed the effect of this campaign on stigma the authors found that while there was a greater understanding of the neurobiology of mental illness, stigma towards mental illness and substance dependence went unchanged (Pescosolido et al. 2010). Stigma towards mental illness is resistant to change, in the current study, the integration of care into primary care practice had little impact on the stigma the participants felt.

While a satisfactory measurement model was found, it required two factors, Perceived Stigma and Comfort Level. With just five items assessing stigma we may be missing other import conceptualizations of stigma. In a review of general stigma measurements, Link and colleagues (2004), point out the numerous components of stigma and the instruments that have already been designed to assess these. In the absence of an elder specific measure of stigma it may be advantageous to adapt one of the more comprehensive measures than to further pursue the Stigma Assessment as a specific elder measurement of stigma.

In addition to problem of missing constructs within the Stigma Assessment, there is also the problem of the use of Items 1 and 2 to determine the wording of the remaining items. Questionnaires employing such devices always require the questionnaire to be administered by a research assistant and require a strict administration protocol so as not to introduce interviewer bias. In this analysis we found bias associated with both these items and one of the items associated with Perceived Stigma. The items seem to induce bias, in future elder specific stigma scales, items could be included that pertain to stigma related to alcohol abuse and stigma associated with mental health rather than using items to determine the wording of the remaining items.

The development of valid and reliable measures of behavioral health stigma for use in an older adult population is important for many reasons. Early identification and treatment is key as older males seem particularly affected by the stigma associated with a behavioral health diagnosis, which is cause for concern because older males have the highest rate of completed suicide (Reynolds and Kupfer 1999). Health care practitioners and social service providers need to have the ability to identify those older adults in need of behavioral health services who might be averse to seeking treatment in order to provide targeted outreach.

As more attention focuses on the stigma associated with behavioral health, being able to accurately measure the impact of research studies such as PRISM-E is paramount. As research funding gets more competitive, the ability to employ a psychometrically sound instrument developed specifically for use in a specific population, such as older adults, may be viewed as a way to better inform the evidence base than study protocols using more generalized measures of stigma.

With proper identification techniques, interventions designed to lessen the feelings of perceived stigma in an individual with behavioral health challenges will play an important part in ensuring older adults get treatment they need. Improving self- esteem and self-efficacy has been linked to treatment access and adherence (Link et al. 2002).

One important limitation of the current study is related to the respondents who agreed to receive behavioral health treatment and participate in the PRISM-E study. This group of participants may differ in a number of ways from those individuals who declined to participate in the study. Namely, those who declined may have greater feelings of stigma which have a greater influence over their behavioral health treatment choices than those who agreed to participate in the study. If this occurred, the sample would be biased toward those whose feelings of stigma have less of an impact on their behavioral health care utilization. Stigma may be more of a barrier to treatment for those who declined to participate in the study than for those who agreed to participate. Another limitation is lack of a formal clinical diagnosis of depression, anxiety and alcohol use/abuse, while the measures used to screen for these conditions, in general, have good agreement with the DSM-IV, it is possible that some people were included who would not have met the criteria for a diagnosis and vice versa. However, this is just the first step in further understanding, the psychometric characteristics of the Stigma Assessment and the respondent characteristics that are related to stigma over time.

Future research endeavors to understand the complex relationship between feelings of stigma and behavioral health outcomes will add to the research base and help to reduce the impact of perceived stigma as a barrier to seeking and receiving behavioral health treatment. By developing well validated and psychometrically sounds instruments to assess stigma we will be more able to break through the barrier of stigma and ensure that all older adults receive the care that they need.

References

Bambauer, K. Z., & Prigerson, H. G. (2006). The stigma receptivity scale and its association with mental health service use among bereaved older adults. *The Journal of Nervous and Mental Disease, 194,* 139–141.

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring anxiety: psychometric properties. *Journal of Consulting and Clinical Psychology, 56,* 893–897.

Blow, F. C., Gillespie, B. W., Barry, K. L., Mudd, S. A., & Hill, E. M. (1998). Brief screening for alcohol problems in elderly populations using the Short Michigan Alcohol Screening Test – Geriatric Version (SMAST-G). *Alcoholism, Clinical and Experimental Research, 22*(Suppl), 131A.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research, 21,* 230–258.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale: Lawerence Erlbaum associates.

Dolan, C. V. (1994). Factor-analysis of variables with 2-response, 3-response, 5-response and 7-response response categories - a comparison of categorical variable estimators using simulated data. *British Journal of Mathematical and Statistical Psychology, 47,* 309–326.

Gomberg, E. S. L. (1988). Alcoholic women in treatment - the question of stigma and age. *Alcohol and Alcoholism, 23,* 507–514.

Graham, N., Lindesay, J., Katona, C., Bertolote, J. M., Camus, V., Copeland, J. R. M., et al. (2003). Reducing stigma and discrimination against older people with mental disorders: a technical consensus statement. *International Journal of Geriatric Psychiatry, 18,* 670–678.

Howard, G. S., & Dailey, P. R. (1979). Response-shift bias - source of contamination of self-report measures. *Journal of Applied Psychology, 64,* 144–150.

Keyes, K. M., Hatzenbuehler, M. L., McLaughlin, K. A., Link, B., Olfson, M., Grant, B. F., et al. (2010). Stigma and treatment for alcohol disorders in the United States. *American Journal of Epidemiology, 172*, 1364–1372.

King-Kallimanis, B., Oort, F., & Garst, G. (2010). Using structural equation modelling to detect measurement bias and response shift in longitudinal data. *ASTA-Advances in Statistical Analysis, 94*, 139–156.

Levkoff, S. E., Chen, H. T., Coakley, E., Herr, E. C. M., Oslin, D. W., Katz, I., et al. (2004). Design and sample characteristics of the PRISM-E multisite randomized trial to improve behavioral health care for the elderly. *Journal of Aging and Health, 16*, 3–27.

Link, B. G., Cullen, F. T., Struening, E., Shrout, P. E., & Dohrenwend, B. P. (1989). A modified labeling theory approach to mental disorders: an empirical assessment. *American Sociological Review, 54*, 400–423.

Link, B. G., Struening, E. L., Neese-Todd, S., Asmussen, S., & Phelan, J. C. (2002). On describing and seeking to change the experience of stigma. *Psychiatric Rehabilitation Skills, 6*(2), 201–231.

Link, B. G., Yang, L. H., Phelan, J. C., & Collins, P. Y. (2004). Measuring mental illness stigma. *Schizophrenia Bulletin, 30*, 511–541.

Mechanic, D., Mcalpine, D., Rosenfield, S., & Davis, D. (1994). Effects of illness attribution and depression on the quality-of-life among persons with serious mental-illness. *Social Science & Medicine, 39*, 155–164.

Mellenbergh, G. J. (1989). Item bias and item response theory. *Int J Ed R, 13*, 127–143.

Meredith, W. (1993). Measurement invariance, factor-analysis and factorial invariance. *Psychometrika, 58*, 525–543.

Neale, M. C. (2004). Mx32 [computer software]. Available from http://www.vcu.edu/mx/

Oort, F. J. (2005). Using structural equation modeling to detect response shifts and true change. *Quality of Life Research, 14*, 587–598.

Oort, F. J., Visser, M. R. M., & Sprangers, M. A. G. (2005). An application of structural equation modeling to detect response shifts and true change in quality of life data from cancer patients undergoing invasive surgery. *Quality of Life Research, 14*, 599–609.

Pescosolido, B. A., Martin, J. K., Long, J. S., Medina, T. A., Phelan, J. C., & Link, B. G. (2010). "A disease like any other?" A decade of change in public relations to schizophrenia, depression, and alcohol dependence. *Am J Psychiat* (in press). doi: 10.1176/appi.ajp.2010.09121743.

Radloff, L. S. (1977). The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385–401.

Reynolds, C. F., & Kupfer, D. J. (1999). Depression and aging: a look to the future. *Psychiatric Services, 10*, 1167–1172.

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: review practice and implications. *Human Resource Management Review, 18*, 210–222.

Sheehan, D., Lecrubier, Y., Sheehan, K., Amorim, P., Janavs, J., Weiller, E., et al. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of Clinical Psychiatry, 59* (Suppl 20), 22–33.

Sirey, J. A., Bruce, M. L., Alexopoulos, G. S., Perlick, D. A., Raue, P., Friedman, S. J., et al. (2001). Perceived stigma as a predictor of treatment discontinuation in young and older outpatients with depression. *American Journal of Psychiatry, 158*, 479–481.

Smith, S. M., Dawson, D. A., Goldstein, R. B., & Grant, B. E. (2010). Examining perceived alcoholism stigma effect on racial-ethnic disparities in treatment and quality of life among alcoholics. *Journal of Studies on Alcohol and Drugs, 71*, 231–236.

Sprangers, M. A. G., & Schwartz, C. E. (1999). Integrating response shift into health-related quality of life research: a theoretical model. *Social Science & Medicine, 48*, 1507–1515.

Substance Abuse and Mental Health Services Administration, Center for Mental Health Services, National Institutes of Health, & National Institute of Mental Health (1999). Mental health: A report of the Surgeon General U.S. Department of Health and Human Services (USDHHS).

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4–70.

Wagenaar, D. B., Mickus, M. A., Gaumer, K. A., & Colenda, C. C. (2002). Late-life depression and mental health services in primary care. *Journal of Geriatric Psychiatry and Neurology, 15*, 134–140.

Ware, J. E., Snow, K. K., Kosinski, M., & Gandek, B. (1993). *SF-36 Health survey Manual and interpretation guide*. Boston: The Health Institute, New England Medical Center.

Bellinda L. King-Kallimanis is a PhD candidate at Academic Medical Center, University of Amsterdam. Her project is titled "Unbiased measurement of health-related quality-of-life" and focuses on issues of measurement invariance and response shift using structural equation modeling. She also provides statistical consulting on a wide variety of projects both in the Netherlands and United States.

Frans J. Oort is full professor of Methods and Statistics in Educational Research, director of the Graduate School of Child Development and Education, and program director of the Research Master Educational Sciences. His research includes non-standard structural equation modeling, measurement and measurement bias, and non-standard applications of statistics in psychological research in medicine and education.

Nancy Lynn is a Financial Analyst at a large health insurance company. She received her Ph.D. in the field of Aging Studies from the University of South Florida in 2009. Her dissertation was titled "Correlates of Attitudes Toward Behavioral Health Services Among Older Primary Care Patients" and involved a secondary data analysis of the Primary Care Research in Substance Abuse and Mental Health for Elderly (PRISM-E) study.

Lawrence Schonfeld is Professor and Chair of the Department of Aging and Mental Health Disparities at the Louis de la Parte Florida Mental Health Institute. He is Principal Investigator/Evaluator of the "Florida BRITE Project" (BRief Intervention and Treatment for Elders) under contract with the Florida Department of Children and Families Substance Abuse Program Office funded through a $14 million "SBIRT" grant from SAMHSA's Center for Substance Abuse Treatment.