Published in final edited form as:

Qual Life Res. 2013 April ; 22(3): 461–471. doi:10.1007/s11136-012-0171-1.

ASSESSMENT OF RESPONSE SHIFT USING TWO STRUCTURAL EQUATION MODELING TECHNIQUES

Pranav K. Gandhi, PhD,

Assistant Professor, South College, School of Pharmacy, 400 Goodys Lane, Knoxville, TN 37922,
Phone: (865) 288-5855

L Douglas Ried, PhD,

Professor and Chair, College of Pharmacy, 12901 Bruce B. Downs Blvd MDC 2, Suite 2128,
University of South Florida, Tampa, FL 33612, Phone: (813) 974-1309, FAX: (813) 905-9890

I-Chan Huang, PhD,

Associate Professor, Department of Health Outcomes and Policy, and Institute for Child Health
Policy, University of Florida, 1329 SW 16th Street, Room 5231, PO Box 100177, Gainesville FL
32610, Phone: (352) 265-2514, Fax: (352) 265-7221

Carole L. Kimberlin, PhD, and

Professor, Department of Pharmaceutical Outcomes and Policy, University of Florida College of
Pharmacy, HPNP Room 3324, Gainesville FL 32610, Phone: (352) 273-6263

Teresa Kauf, PhD

Associate Professor and Graduate Coordinator, Department of Pharmaceutical Outcomes and
Policy, University of Florida College of Pharmacy, HPNP Room 3325, Gainesville FL 32610,
Phone: (352) 273-6252

Pranav K. Gandhi: pgandhi@southcollegetn.edu; L Douglas Ried: lried@health.usf.edu; I-Chan Huang: ichuang@ufl.edu;
Carole L. Kimberlin: kimber@cop.ufl.edu; Teresa Kauf: kauf@cop.ufl.edu

Introduction

Self-report measures of health-related quality of life (HRQoL) are part of any assessment of interventions in randomized trials [1]. Although reliable and valid HRQoL measures are available, change in scores based on pre-test versus post-test data may reflect factors other than the intervention [2]. In other words, individuals experiencing improvement or deterioration over time may adopt a different frame of reference or the concept and meaning of HRQoL [3,4]. Response shift is “a change in the meaning of one’s self-evaluation of a target construct as a result of (a) a change in the respondent’s internal standards of measurement (i.e., recalibration); (b) a change in the respondent’s values (i.e., the importance of component domains constituting the target construct, reprioritization); or (c) a redefinition of the target construct (i.e., reconceptualization)” [3, 4].

Hypertension and coronary artery disease (CAD) are health conditions where symptoms persist over a long period of time [5]. CAD is likely to have a direct impact on HRQoL in

domains such as reduced walking distances, shortness of breath, trouble lifting and carrying objects, and doing housework. The choice of antihypertensive treatment may influence risk of depression and consequently clinical and health outcomes [5, 6]. Response shift is important to consider in treatment evaluations because patients may adjust their perceptions of the benefits and side-effects of various treatment modalities or disease progression.

Statistical techniques are useful in assessing response shift among individuals who respond differently to the same HRQoL measure over time. Covariance structures proposed by Schmitt [7–10] have been used with structural equation modeling (SEM) to identify both reconceptualization and recalibration response shift [7–10]. However, the approach [7–10] does not identify reprioritization or non-uniform recalibration response shift. On the other hand, the SEM approach proposed by Oort [11] identifies all types of response shift. The Oort and Schmitt SEM procedures use different parameters to define and test response shift. The type of response shift identified is based on the parameter operationalization. In addition, the difference in order (forward vs. backward approach) of releasing parameter constraints with the two SEM procedures may influence the results. The backward approach of releasing parameter constraints using the Oort procedure [11] provides a “greater power for detecting changes” [12], however, increases the Type I error probability.

One study directly compared two different SEM methods (i.e., the Oort [11] and Schmitt SEM procedures [7–10]) and found that the Oort procedure best detected uniform and non-uniform recalibration among patients enrolled in a chronic obstructive pulmonary disease (COPD) self-management program [12]. However, Type I error was not controlled which may lead to an overidentification of response shift [11–13]. In addition, a different order of testing was used. When releasing parameter constraints, the authors started with residuals, followed by intercepts, and then factor loadings [12]. The order (i.e., hierarchy) in which parameter constraints are released (factor patterns, followed by factor loadings and intercepts, and then residual variances) will influence whether subsequent parameters are different [12] and threaten an unbiased comparison of latent factor mean scores [14], asserting the need to follow a particular order when releasing parameter constraints.

The objectives of this study were to investigate the occurrence of response shift using both Oort [11, 13, 15] and Schmitt [8, 9] SEM approaches and examine if there was additional divergence based on the order of testing with the Ahmed et al. study [12].

Methods

Source of Data

The study is a secondary analysis using data from subjects enrolled in the Study of Antihypertensive Drugs and Depressive Symptoms (SADD-Sx) [16]. Briefly, SADD-Sx was a substudy of the International Verapamil-Trandolapril Study (INVEST). INVEST was a randomized, open-label, blinded end-point study of 22,576 hypertensive patients with CAD aged >50 years conducted from September 1997 to February 2003 [17, 18]. Patients were randomized to antihypertensive treatment with either a verapamil SR- or atenolol-based strategy to achieve blood pressure control according to the sixth report of the Joint National

Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (JNC VI) [17, 18].

The full sample of SADD-Sx patients residing in the United States were mailed surveys between April 1, 1999, and October 31, 1999 ($n = 2,317$). SF-36 Health Survey and demographic characteristics such as age, gender and race were included in the SADD-Sx survey. Patients were mailed the baseline survey the day after randomization. Follow-up surveys were mailed at one year. Patients were mailed a second survey if it was not returned within 10 working days. Each patient was mailed a reminder letter to enhance the response rate approximately 2 weeks before the follow-up surveys were mailed. If patients failed to respond to the second survey, no extra attempts were made to contact them. Only patients who completed both the baseline and one year follow-up surveys were included.

SADD-Sx was conducted according to the principles of the Declaration of Helsinki [19]. The University of Florida Institutional Review Board approved the study protocol.

SF-36 Health Survey

The SF-36 is a generic HRQoL measure that captures behavioral functioning, subjective well-being and perceptions of health [20, 21]. The SF-36 health survey consists of eight domains: physical functioning (PF), role limitations due to physical health (role physical, RP), bodily pain (BP), vitality (VT), general health perceptions (GH), social functioning (SF), role limitations due to emotional problems (role emotional, RE), and mental health (MH). Raw scores were transformed to a range of 0 to 100, with higher scores indicating better HRQoL [20, 21].

Statistical Analysis

LISREL 8.8 [22] was used to construct and test the SEM for both Oort and Schmitt procedures. SAS 9.1 software [23] was used for the remaining analyses. The eight domains SF-36 domains were assessed for normality. Maximum likelihood estimation (MLE) is robust for data that demonstrates small deviations from normality and was used given the slightly non-normal distribution of our data. A variety of fit indices were adopted to assess the appropriateness of the SEM, including the classic goodness-of-fit index χ^2 ; the standardized root mean square residual (SRMR) with a range between 0.0 and 1.0 where values less than 0.1 indicating an acceptable fit; root mean square error of approximation (RMSEA) with values below 0.08 indicating a good model fit and values below 0.05 indicating a close fit [24]; comparative fit index (CFI), normed fit index (NFI) and non-normed fit index (NNFI) with a range between 0.0 and 1.0. Values greater than 0.9 indicate better model fit for all these three criteria [24].

Oort's SEM Approach (Method 1)

The first method conducted to test response shift is a four-step procedure [11]. True change was defined as a change in a patient's target construct of HRQoL, representing improved or deteriorated health status [11].

Model 1: Establishing an Appropriate Baseline Model—Results from confirmatory factor analyses and the literature [20] were used to define the baseline SF-36 measurement model.

Model 2: No Response Shift Model—In model 2, we set the parameters of factor loadings, residual variances, and intercepts to be equal over time (i.e., a hypothesis of no response shift). The data fit was compared for Model 2 and Model 1. The χ^2 difference test evaluated the statistical significance in the comparison with the freed and fully constrained models. The null hypothesis was rejected if the χ^2 difference test was statistically significant, suggesting response shift.

Model 3: Testing for Specific Types of Response Shift—Parameter estimate differences between Model 3 and Model 2 were tested and guided by the modification indices to represent different types of response shift [11]. The detection of response shift based on model parameters threatens the unbiased comparison of latent factor mean scores [14]. Consistent with previous research using the Oort approach [11, 13, 15, 25], we chose the order by releasing parameter constraints on factor pattern, followed by factor loadings and intercepts, and then residuals (i.e., a backward approach). Parameter invariance was tested by releasing one parameter at a time and constraining all other parameters over time. The equality constraints were released on each factor loading while imposing equality constraints on the remaining factor loading parameters. After inspecting each result, a similar procedure was conducted by releasing equality constraints on intercepts while imposing equality constraints on the remaining intercepts and factor loadings when response shift was not identified. Subsequently, a similar procedure was conducted by releasing equality constraints on residual variances while imposing equality constraints on factor loadings, intercepts, and other residual variances. We tested the modification step indices with the χ^2 difference test. A χ^2 difference of 3.84 with df(1) indicates response shift.

Interpretation of the different model parameters are linked with different types of response shift (Table 1). First, reconceptualization is indicated if a change in the matrix pattern containing factor loadings at Time 1 differs from the matrix pattern at Time 2. Second, if the factor loading value of a domain changes over time, reprioritization has occurred. Third, uniform recalibration is indicated if there were differences between intercepts while still constraining the common factor loadings. Uniform recalibration occurs when patients adjust their perception to all response options in the same direction and to the same extent. Lastly, the differences between residual variances over time indicate non-uniform recalibration. In other words, non-uniform recalibration occurs if patients' responses to the scale are mixed (e.g., some lower or higher) [11].

Model 4: Identification of True Change and Effect Size Calculation—In model 4 we tested other types of invariance, including the equality of latent factor means, variances, and correlations. Model 4 parameter estimates were used to calculate effect size indices for true change and the effect of response shift. The obtained true change effect size divided by the estimated standard deviation of observed change yielded Cohen's effect-size indices d [26]. Effect size values (d) < 0.2, 0.2–0.49, 0.5–0.79, and 0.8 were considered negligible, small, medium, and large, respectively. Estimates from models that accounted for response

shift (Model 4) subtracted from models that did not account for response shift (Model 2). The absolute difference between the two models is the estimated effect of response shift.

Since multiple results would be necessary, we adjusted for possible Type I error where a Bonferroni-adjusted critical value [27] associated with a probability of 0.05/40 (based on the modification indices for all scales in regards to factor loadings, intercepts, and residual variances) was used in our study.

Schmitt's SEM Approach (Method 2)

We applied the same baseline measurement model (Model 1) for both the Schmitt and Oort SEM approaches. Parameter invariance over time was tested step-by-step, adding one parameter at a time (i.e., forward approach). Parameter estimate differences between the two models were tested to represent different types of response shift [8–10] (Table 1). The comparison between two models was guided by the modification indices. A χ^2 difference of 3.84 with df(1) indicated response shift.

The first step in Schmitt's SEM method assessed the change in the entire variance-covariance matrix over time [8]. Reconceptualization was indicated if these differences were due to the variance of the *a priori* pattern of factor loadings imposed on the scales (Model 2). If reconceptualization was not detected, we tested subsequent models for scale recalibration by testing the equivalence of the factor variances (Model 3) or the factor loadings (Model 4) over time. The equality constraint on each parameter added to the model was released step-by-step while imposing constraints on the remaining parameters where response shift was not identified. Finally, we tested Model 5 for change in error variance to evaluate if changes were due to random error. The goodness-of-fit of the SEM solution was evaluated for statistical significance [12]. Schmitt's SEM technique [8] precludes adjusting for Type I error. Additionally, changes in factor loadings are interpreted differently for the Schmitt and Oort approaches.

Results

The initial SADD-Sx sample consisted of 2,317 INVEST patients assigned to either the verapamil SR-based (n = 1184) or atenolol-based (n = 1133) treatment strategy [16]. Complete responses from both the baseline and one year follow-up surveys were obtained from 39.2% (n = 909). At baseline, patients were randomized to treatment groups. Among other reasons, differential dropout rates among patients assigned to one treatment arm of the study compared to the other may signal response shift. The final sample of patients in the two treatment groups was similar on their sociodemographic characteristics and SF-36 responses ($p > 0.05$).

Table 2 gives baseline and one year means scores and standard deviations for all SF-36 scales. Conventional t-tests indicate improvement in RP, BP, GH, VT, MH and RE, and no change in PF and SF scale mean scores.

Step 1: Measurement Model

General physical (Gen PHYS) and General mental (Gen MENT) represent two latent variables (Figure 1). Similar to the Oort et al. study [13] we found that the wording of the SF items captured both physical and mental aspects. Hence SF loaded both on Gen PHYS and Gen MENT. The residual factors for RP and RE were allowed to co-vary ($r = 0.22$) since the covariance was not adequately explained by the correlation between the two latent factors. Both RP and RE have similar wording and questions about social roles. The measurement model in this step had acceptable fit ($\chi^2(84) = 301.284$, $p < 0.001$; RMSEA = 0.054 (90% CI = 0.048, 0.061)) (Model 1, Table 3).

Step 2: No Response Shift Model

In Step 2, the model with all the parameter estimates freed (i.e., Model 1) was compared with the fully constrained model (i.e., Model 2, Table 3). The initial measurement model fit the data significantly better than the no response shift model ($\chi^2(25) = 64.128$, $p < 0.0001$) indicating that response shift was likely.

Step 3: Identification of Specific Types of Response Shift

Three cases of response shift were identified before controlling for Type I error. The fit of Model 3^{a1} after accounting for uniform recalibration for PF scale was acceptable (RMSEA = 0.049 (90% CI = 0.043 – 0.055), SRMR = 0.037, CFI = 0.991, NFI = 0.986, NNFI = 0.989), and significantly better than the fit of Model 2 (χ^2 difference test: $\chi^2(1) = 24.244$, Table 3). We found that the model fit significantly improved when the BP intercept parameter was freed to vary over time (χ^2 difference test: $\chi^2(1) = 4.05$, RMSEA = 0.049 (0.043 – 0.055), SRMR = 0.037, CFI = 0.991, NFI = 0.986, NNFI = 0.989, Model 3^{a2}, Table 3), indicating presence of uniform recalibration for BP scale. The measurement model also improved when residual variances were freed to vary (χ^2 difference test: $\chi^2(1) = 6.656$, RMSEA = 0.049 (0.043 – 0.055), SRMR = 0.037, CFI = 0.991, NFI = 0.987, NNFI = 0.989, Model 3^{a3}, Table 3), indicating presence of nonuniform recalibration for PF scale.

After controlling for Type I error, differences in intercepts over time were observed indicating uniform recalibration for PF. The fit of Model 3^b after accounting for one instance of response shift was acceptable (RMSEA = 0.049 (90% CI = 0.043 – 0.055), SRMR = 0.037, CFI = 0.991, NFI = 0.986, NNFI = 0.989) and significantly better than the fit of Model 2 (χ^2 difference test: $\chi^2(1) = 24.244$, Table 3).

Step 4: Contributions of Response Shifts and True Change to Change in the Observed Variables

In step 4, after testing for the equality of latent factor means, variances, and correlations, both models, i.e., Model 4^a obtained from Model 3^a and Model 4^b obtained from Model 3^b fitted the data closely (Table 3).

After adjusting for possible Type I error, the estimates of Model 4^b (Table 3) parameters are presented in Table 4. Based on these parameter estimates, the response shift effect size was -0.118 for the uniform recalibration and zero for the nonuniform recalibration. Recalibration

response shift had a negligible impact on the estimation of true change in the PF mean scores.

Schmitt's SEM Approach

We started with the same baseline measurement model (Model 1, Table 5) for both the Schmitt and Oort approaches. Based on the finding of invariance in factor structure, reconceptualization was not observed (Model 2, Table 5). The next step was to assess whether the factor variances or factor loadings were equivalent over time, indicating absence of recalibration. We found that releasing the factor variance parameter constraint did not significantly improve the model fit (Model 3, Table 5); they were constrained to be equal over time. However, we found that the model significantly improved when the PF factor loading parameter was freed to vary over time (χ^2 difference test: $\chi^2(1) = 3.906$, RMSEA = 0.05 (0.044 – 0.057), SRMR = 0.041, CFI = 0.991, NFI = 0.987, NNFI = 0.989, Model 4, Table 5), indicating presence of recalibration response shift. The measurement model also improved when error variances were freed to vary (χ^2 difference test: $\chi^2(1) = 8.446$, RMSEA = 0.049 (0.043 – 0.055), SRMR = 0.039, CFI = 0.991, NFI = 0.987, NNFI = 0.990, Model 5, Table 5), indicating presence of random error over time.

Discussion

Response shift was identified in patients' responses to the PF scale using both SEM procedures (uniform recalibration identified using Oort's approach [11] and recalibration identified using Schmitt's approach [8]); however, both approaches use different parameters to define and test response shift. When using the Oort approach, the uniform recalibration response shift had a marginal influence on the estimation of true change in the PF mean scores. Oort [11] defines uniform recalibration as a change in means of the observed variables over time compared to Schmitt [8] who defines recalibration as a change in factor loadings or factor variances over time. Changes in both the covariance and mean structures are accounted for when using the SEM approach proposed by Oort [11, 12]. On the other hand, the Schmitt approach [8] only accounts for covariance structure changes, due to which the Schmitt SEM approach may give a larger estimation of recalibration response shift [12] in our study. Consistent with previous recommendations [12], our study suggests that the differences in the findings between the Oort and Schmitt approaches may be due to method differences and not the sample.

The second objective was to examine if our findings regarding response shift were similar to those of Ahmed et al. [12]. We identified response shift using both the Oort and Schmitt SEM techniques. In contrast, Ahmed et al. [12] identified uniform and non-uniform recalibration response shift when using the Oort procedure. However, Ahmed and colleagues [12] did not detect response shift when using the Schmitt procedure. Our study differs from the Ahmed et al. [12] study in terms of the order of releasing parameter constraints, control of Type I error, and characteristics of subjects. Ahmed et al. [12] chose to release constraints on residuals, followed by intercepts, factor loadings, and then factor pattern. In contrast, we sequentially tested for invariance by releasing parameter constraints on factor patterns, factor loadings, followed by intercepts, and then residuals. Consistent

with prior recommendations, we believe that the order in which parameter constraints are released is of utmost importance and influence the unbiased comparison of latent factor mean scores [14]. If the order of releasing parameter constraints impacts the response shift findings, it is not known if the identification of response shift is true. In this regard, we also tested for response shift using a similar order of releasing parameter constraints as Ahmed et al. [12]. It was found that the order of testing did not alter our study results, so, the impact associated with order of testing may be minimal. The identification of response shift in our study appears credible since both order of sequential tests identify the same types of response shift.

An alternative reason for the difference in findings between our and Ahmed et al. study [12] may be related to the use of Bonferroni correction to control for Type I error. After controlling for Type I error, we retained only one instance of response shift, i.e. uniform recalibration response shift for the PF scale when using the Oort approach. Ahmed et al. [12] did not control for Type I error when using Oort's approach [11] which may have led to identification of response shift that was the result of an experiment-wise error. Finally, hypertensive CAD patients in our study and COPD patients in Ahmed et al. study [12] may experience different types and magnitude of response shift as a consequence of disease progression and different treatments. In this regard, the differences between the present study and Ahmed et al. [12] study may be due to both variation in the method and the study sample.

The PF scale scores revealed a non-significant decline one year after allocation to treatment strategies. However, after accounting for response shift, a positive true change contribution indicated that treatment effects may have been underestimated, leading to false negative results. It was only when response shift was accounted for that improvement in PF scores following treatment could be seen. Assessment of the treatment's "true" effects must include individuals' change in perception due to treatment or change in health, especially when using self-report measures in clinical trials. Identifying response shift has implications for interpreting the measurement of change in clinical and research endeavors.

The finding of response shift for the SF-36 PF scale may be explained as a consequence of the disease condition. A study by Bar-On and Amir [28] identified beta change (i.e., recalibration) over a one year period in hypertensive and normotensive male subjects randomly assigned to methyldopa, isradipine or placebo treatments. Vascular disease such as CAD mainly influences physical health [29]. Several studies report that hypertensive CAD patients feel limited in the PF domain. Mitchell et al. [30] report lower mean scores on the physical ability dimension among hypertensive CAD patients. Sevinc and Akyol [31] have found that HRQoL scores were lowest in PF domain for CAD patients compared to emotional and social domain scores.

The SF-36 PF consists of items describing walking distances, vigorous activities, moderate activities, lifting and carrying objects, climbing stairs, among others. A group of patients may have changed their perception for items as well as the response scale options (values, labels, or both) for the PF scale in the same direction and to the same extent due to aspects of functioning impacted by the disease condition [11]. It is plausible that new experiences

have changed their perception of the extent to which their health would physically limit when performing moderate or vigorous activities, lifting objects, walking or climbing stairs [11]. Moreover, it is likely that patients have learned coping skills relevant to their illness, so that their functioning improved more, or deteriorated less, than would be expected on account of their physical health.

An alternative explanation for the presence of response shift may be due to the assignment of antihypertensive treatment strategies. Atenolol has been reported to show adverse somatic side effects such as feelings of pressure and tightness in the head or body, dizziness and faintness, increased fatigue, and muscular weakness [16, 32]. The side effects may change the patient's assessment of their HRQoL concepts by reduced walking distances and trouble in lifting and carrying objects and doing housework, among others. The occurrence of uniform recalibration in PF scale may partly be explained as a consequence of the aforementioned effects. Future research examining the relationship between somatic side effects and the occurrence of uniform recalibration will support this proposition. Few studies report a stereotype implying that men are more physically oriented and invested in their ability to do heavier physical labor or strenuous activities [33, 34]. The studies [33, 34] report an inverse relation between age and SF-36 PF domain among CAD patients. The identification of response shift may thus be due to different age groups or gender roles which operate differently for males and females [33, 34]. It seems plausible that some subgroups in our study sample may have undergone response shift. Although this is beyond the scope of the present study, our ongoing work attempts to examine the role of catalysts, socio-demographic variables, depression, and anti-hypertensive treatment strategies on the presence of response shift over time among hypertensive CAD patients.

Evidence demonstrates the usefulness of both Schmitt and Oort SEM approaches to detect response shift for ordinal or multidimensional datasets. Applications to ordinal or multidimensional data are described by Ahmed et al. [7], Oort et al. [13], Visser et al. [15], and Barclay-Goddard et al. [25]. The SEM analysis in our study was purposefully conducted at the domain level rather than individual items. We chose this approach because, from the perspective of comparative effectiveness research, the SF-36 is best used to examine the overall impact of different treatments as a global generic measure, rather than assessing HRQoL item-by-item. If the SEM analysis was used in another context (e.g., scale development), detection of response shift on an item-level would be of greater interest.

Based on this and previous studies, we conclude that certain SF-36 HRQoL domains are more or less susceptible to response shift. For instance, we could find no reports of response shift for the SF-36 PF scale. Oort et al. [13] suggest that SF-36 PF items might be less prone to recalibration because they could be answered objectively. We checked for other types of response shift when freeing equality constraints on parameters one at a time, but no other selection provided a better model fit. Barclay-Goddard et al. [25] identified the presence of response shift post-stroke in PF construct. They hypothesized the word 'difficulty' in the response options for two of the four measured variables partly explained response shift. However, they acknowledged it was not the only plausible explanation [25]. Response shift research should be linked with other methodologies and fields of research, such as qualitative methodologies and cognitive interviewing, to rule out other explanations [35]. In

studies or settings where response shift is suspected and design methods are not feasible to assess response shift for group-level analysis, statistical methods should continue to be useful. Response shift assessment is particularly important in hypertensive CAD patients when impairments and activity limitations may never recover fully, but improvement in HRQoL is a therapeutic goal.

Limitations

SEM approaches detect response shift at a group level [11–13]. Only if a substantial number of participants undergo a response shift, will it be detected using any approach. Barclay-Goddard et al. [25, 35] suggest that not all individuals would experience response shift. Moreover, even those who do may not experience it at the same time [25, 35]. By definition, response shift is a time-dependent phenomenon and likely occurs at different times with different individuals. Therefore, its detection and effect-size are likely diminished when using cross-sectional group level analyses. With 3 time points, statistical approaches that examine response shift at an individual level (e.g., growth curve modeling [36] and latent trajectory analysis [37, 38]) may better detect response shift in subgroups of patients. [13]. Cognitive interviewing technique applied in a past study [39] can be used to demonstrate whether statistical approaches truly detect the occurrence of response shift [35]. Used together, individual level statistical and qualitative approaches will better elucidate the response shift phenomenon. Lastly, response shift was identified among hypertensive CAD patients assigned to antihypertensive treatment strategies. However, availability of a control group (potentially) not expected to experience a response shift would provide additional support.

Conclusion

Both Oort [11] and Schmitt [8] SEM approaches identified response shift (uniform recalibration using the Oort approach and recalibration using the Schmitt approach) in hypertensive CAD patients for the SF-36 PF scale. However, either or both variation in the method and the sample used may influence detection and characterization of response shift.

References

1. Goodwin PJ, Black JT, Bordeleau LJ, Ganz PA. Health-related quality of life measurement in randomized clinical trials in breast cancer – taking stock. *Journal of the National Cancer Institute*. 2003; 95(4):263–81. [PubMed: 12591983]
2. Campbell DT. Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*. 1957; 54(4):297–312. [PubMed: 13465924]
3. Sprangers MA, Schwartz CE. Integrating Response Shift into Health-Related Quality of Life Research: A Theoretical Model. *Social Science and Medicine*. 1999; 48:1507–15. [PubMed: 10400253]
4. Schwartz CE, Sprangers MA. Methodological Approaches for Assessing Response Shift in Longitudinal Health-Related Quality of Life Research. *Social Science and Medicine*. 1999; 48:1531–48. [PubMed: 10400255]
5. Ried LD, Tueth MJ, Taylor MD, Sauer BC, Lopez LM, Pepine CJ. Depressive symptoms in coronary artery disease patients after hypertensive treatment. *Annals of Pharmacotherapy*. 2006; 40:597–604. [PubMed: 16569807]

6. Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL Jr. for the National Heart, Lung, and Blood Institute Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure; National High Blood Pressure Education Program Coordinating Committee. The seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. *Journal of the American Medical Association*. 2003; 289:2560–72. [PubMed: 12748199]
7. Ahmed S, Mayo N, Corbiere M, Wood-Dauphinee S, Hanley J, Cohen R. Change in Quality of Life of People with Stroke Over Time: True Change or Response Shift? *Quality of Life Research*. 2005; 14:611–27. [PubMed: 16022056]
8. Schmitt N. The use of analysis of covariance structures to assess beta and gamma change. *Multivariate Behavioral Research*. 1982; 17:343–58.
9. Schmitt N, Pulakos ED, Lieblein A. Comparison of three techniques to assess group-level beta and gamma change. *Applied Psychological Measurement*. 1984; 8:249–60.
10. Schaubroeck J, Green SG. Confirmatory factor analytic procedures for assessing change during organizational entry. *Journal of Applied Psychology*. 1989; 74(6):892–900.
11. Oort FJ. Using structural equation modeling to detect response shifts and true change. *Quality of Life Research*. 2005; 14:587–98. [PubMed: 16022054]
12. Ahmed S, Bourbeau J, Maltais F, Mansour A. The Oort structural equation modeling approach detected a response shift after a COPD self-management program not detected by the Schmitt technique. *Journal of Clinical Epidemiology*. 2009; 62:1165–72. [PubMed: 19664904]
13. Oort FJ, Visser MRM, Sprangers MAG. 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*. 2005; 14:599–609. [PubMed: 16022055]
14. Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: Suggestions, practices and recommendations for organizational research. *Organizational Research Methods*. 2000; 3:4–70.
15. Visser M, Oort F, Sprangers M. Methods to Detect Response Shift in Quality of Life Data: A Convergent Validity Study. *Quality of Life Research*. 2005; 14:629–39. [PubMed: 16022057]
16. Ried LD, Tueth MJ, Handberg E, Kupfer S, Pepine CJ. the INVEST study group. A Study of Antihypertensive Drugs and Depressive Symptoms (SADD-Sx) in Patients Treated With a Calcium Antagonist Versus an Atenolol Hypertension Treatment Strategy in the International Verapamil SR-Trandolapril Study (INVEST). *Psychosomatic Medicine*. 2005; 67:398–406. [PubMed: 15911902]
17. Pepine CJ, Handberg-Thurmond E, Marks RG, Conlon M, Cooper-DeHoff R, Volkers P, et al. Rationale and design of the International Verapamil SR/Trandolapril Study (INVEST): an Internet-based randomized trial in coronary artery disease patients with hypertension. *Journal of the American College of Cardiology*. 1998; 32:1228–37. [PubMed: 9809930]
18. Pepine CJ, Handberg EM, Cooper-DeHoff RM, Marks RG, Kowey P, Messerli FH, et al. A calcium antagonist vs a non-calcium antagonist hypertension treatment strategy for patients with coronary artery disease. The International Verapamil-Trandolapril Study (INVEST): a Randomized Controlled Trial. *Journal of the American Medical Association*. 2003; 290:2805–16. [PubMed: 14657064]
19. World Medical Association Declaration of Helsinki. Recommendations guiding physicians in biomedical research involving human subjects. *Cardiovascular Research*. 1997; 35:2–3. [PubMed: 9302340]
20. Saris-Baglama, RN.; Dewey, CJ.; Chisholm, GB.; Kosinski, M.; Bjorner, JB.; Ware, JE. SF-36 Health Outcomes Scoring Software. 2004. p. 1-157.
21. Ware, JE.; Snow, KK.; Kosinski, M.; Gandek, B. SF-36 Health Survey: Manual and Interpretation Guide. Boston, MA: The Health Institute, New England Medical Center; 1993.
22. Jöreskog, K.; Sörbom, D. LISREL 8. Lincolnwood, IL: Scientific Software International Inc; 1993.
23. SAS Institute. SAS users guide, version 9.1. Cary, NC: SAS institute, Inc; 2004.
24. Brown, TA. Confirmatory factor analysis for applied research. Guilford Press; New York: London: Guilford Press; 2006.

25. Barclay-Goddard R, Lix LM, Tate R, Weinberg L, Mayo NE. Response shift was identified over multiple occasions with a structural equation modeling framework. *Journal of Clinical Epidemiology*. 2009; 62:1181–88. [PubMed: 19615858]
26. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. 2. Lawrence Erlbaum Associates, Inc; Hillsdale, NJ: 1988.
27. King-Kallimanis B, Oort FJ, Visser MRM, Sprangers MAG. Structural equation modeling of health-related quality-of-life data illustrates the measurement and conceptual perspectives on response shift. *Journal of Clinical Epidemiology*. 2009; 62:1157–64. [PubMed: 19595574]
28. Bar-On D, Lazar A, Amir M. Quantitative assessment of response shift in QOL research. *Social Indicators Research*. 2000; 49:37–49.
29. de Graff JC, Ubbink DTh, Kools EJC, Chamuleau SAJ, Jacobs MJHM. The impact of peripheral and coronary artery disease on health-related quality of life. *Annals of Vascular Surgery*. 2002; 16:495–500. [PubMed: 12085126]
30. Mitchell RA, Imperial E, Kelleher P, Brunner P, Gass G. Perceived health problems in subjects with varying cardiovascular diagnoses. *Journal of Behavioral Medicine*. 1991; 14:505–12. [PubMed: 1744912]
31. Sevinc S, Akyol AD. Cardiac risk factors and quality of life in patients with coronary artery disease. *Journal of Clinical Nursing*. 2010; 19:1315–25. [PubMed: 20500341]
32. Bass C, Akhras F, Upward J, Keates J, Lowe D, Harry J, et al. Does atenolol improve physical and psychological function after coronary artery bypass surgery: a controlled study. *Journal of Psychosomatic Research*. 1987; 31:521–9. [PubMed: 3312591]
33. Macintyre S, Ford G, Hunt K. Do women 'over-report' morbidity? Men's and women's responses to structured prompting on a standard question on long standing illness. *Social Science and Medicine*. 1999; 48:89–98. [PubMed: 10048840]
34. McCreary DR, Newcomb MD, Sadava SW. Dimensions of the male gender role: A confirmatory analysis in men and women. *Sex Roles*. 1998; 39:81–95.
35. Barclay-Goddard R, Epstein JD, Mayo NE. Response shift: a brief overview and proposed research priorities. *Quality of Life Research*. 2009; 18:335–46. [PubMed: 19241143]
36. Brossart D, Clay D, Willson V. Methodological and statistical considerations for threats to internal validity in pediatric outcome data: response shift in self-report outcomes. *Journal of Pediatric Psychology*. 2002; 27:97–107. [PubMed: 11726684]
37. Mayo NE, Scott SC, Ahmed S. Case management poststroke did not induce response shift: the value of residuals. *Journal of Clinical Epidemiology*. 2009; 62:1148–56. [PubMed: 19595568]
38. Mayo NE, Scott SC, Dendukuri N, Ahmed S, Wood-Dauphinee S. Identifying response shift statistically at the individual level. *Quality of Life Research*. 2008; 17:627–39. [PubMed: 18386161]
39. Bloem EF, van Zuuren FJ, Koenenman MA, Rapkin BD, Visser MRM, Koning CCE, et al. Clarifying quality of life assessment: do theoretical models capture the underlying cognitive processes? *Quality of Life Research*. 2008; 17:1093–1102. [PubMed: 18704756]

Figure 1.

The measurement model used in response shift detection

Notes: Circles represent latent variables (common and residual factors) and squares represent observed variables (SF-36 scales).

Abbreviations: PF – physical functioning; RP – role physical; BP – bodily pain; GH – general health; VT – vitality; SF – social functioning; RE – role-emotional; MH – mental health.

Table 1

Comparison of Oort and Schmitt's SEM approaches

	Oort's SEM approach	Schmitt's SEM approach
Recalibration	Two types – uniform and non- uniform recalibration. Uniform recalibration is indicated by change in intercepts, non-uniform recalibration is demonstrated by change in residual variances e.g., Perhaps new experiences after the first occasion have changed their idea of how much health should limit physical activity.	Indicated by change in factor loadings and factor variances are equal over time e.g., an increase or decrease in the factor variances would indicate that the respondents perceive more or less difference in the relevant constructs as a result of the intervention.
Reprioritization	Shown by change in factor loadings. e.g., Variable has become more indicative or less indicative of the concept involved.	Does not identify reprioritization
Reconceptualization	Shown by change in factor patterns over time. e.g., If a domain does not belong to the set of variables defining the concept of a latent construct at the first occasion, whereas at the second occasion it does, reconceptualization has occurred.	Factor patterns and factor covariances e.g., no reconceptualization means that the relationship among factors with similar intercorrelations should be the same before and after intervention.

Table 2

Means, standard deviations, and effect sizes for SF-36 scales at baseline and one year (n = 909)

Scale	Before assignment to treatment strategy		After assignment to treatment strategy		Post-Pre d-Index [#]
	Mean	Standard deviation	Mean	Standard deviation	
PF	55.142	29.013	54.526	29.030	-0.021
RP	47.992	40.866	51.375	41.201	0.082 *
BP	62.162	29.428	66.425	30.967	0.138 **
GH	52.031	21.851	54.334	21.994	0.105 **
VT	46.368	22.473	48.398	21.861	0.090 **
SF	70.421	23.826	71.741	23.101	0.055
RE	62.871	41.071	66.593	40.228	0.091 **
MH	72.095	18.772	73.627	18.762	0.082 **

N = 909;

[#] Standardized mean difference: 0.2, 0.5, and 0.8 indicate 'small', 'medium', and 'large' differences;

* p < 0.05,

** p < 0.01 in paired t-test.

Abbreviations: PF – physical functioning; RP – role physical; BP – role physical; BP – bodily pain; GH – general health; VT – vitality; SF – social functioning; RE – role-emotional; MH – mental health.

Table 3

Oort method: goodness of fit and χ^2 differences between models (n = 909)

Model	Df	CHISQ	RMSEA	SRMR	CFI	NFI	NNFI
Model 1	84	301.284	0.054 (0.048 – 0.061)	0.034	0.991	0.988	0.987
Model 2	109	365.412	0.051 (0.045–0.057)	0.038	0.989	0.985	0.989
Model 3 ^{a1}	108	341.168	0.049 (0.043–0.055)	0.037	0.991	0.986	0.989
Model 3 ^{a2}	107	337.118	0.049 (0.043 – 0.055)	0.037	0.991	0.986	0.989
Model 3 ^{a3}	106	330.462	0.049 (0.043 – 0.055)	0.037	0.991	0.987	0.989
Model 4 ^a	110	336.630	0.048 (0.042 – 0.054)	0.040	0.991	0.986	0.989
Model 3 ^b	108	341.168	0.049 (0.043–0.055)	0.037	0.991	0.986	0.989
Model 4 ^b	112	346.858	0.048 (0.042 – 0.054)	0.040	0.990	0.986	0.990

Abbreviations: RMSEA – root mean square error of approximation, SRMR – standardized root mean square residual, CFI – comparative fit index, NFI – normed fit index, NNFI – non-normed fit index. PF – physical functioning.

RMSEA <= 0.05 close fit; 0.05 to 0.08 reasonable fit; >= 0.10 poor fit; CFI, NFI, and NNFI > 0.90 reasonably good fit, SRMR < 0.10 is favorable

^aModels without controlling for Type I error;

^bModels using Bonferroni-adjusted critical value to control for Type I error

Table 4
Parameter estimates in the final model (Model 4^b, Table 3) (n = 909)

	Before assignment to treatment strategy		After assignment to treatment strategy	
	Gen. PHYSI	Gen. MENT1	Gen. PHYS2	Gen. MENT2
Factor loadings (1°)				
PF1	0.754		PF2	0.754
RPI	0.748		RP2	0.748
BPI	0.662		BP2	0.662
GHI	0.775		GH2	0.775
VT1	0.761		VT2	0.761
SF1	0.349	0.449	SF2	0.349
RE1		0.722	RE2	0.722
MH1		0.766	MH2	0.766
Intercepts (τ)				
PF1	55.142		PF2	51.720
RPI	47.661		RP2	47.661
BPI	62.971		BP2	62.971
GHI	52.082		GH2	52.082
VT1	46.276		VT2	46.276
SF1	69.991		SF2	69.991
RE1	63.188		RE2	63.188
MH1	72.115		MH2	72.115
Residual variances				
PF1	0.432		PF2	0.432
RPI	0.441		RP2	0.441
BPI	0.562		BP2	0.562
GHI	0.399		GH2	0.399
VT1	0.421		VT2	0.421
SF1	0.428		SF2	0.428
RE1	0.478		RE2	0.478
MH1	0.414		MH2	0.414

	Before assignment to treatment strategy		After assignment to treatment strategy	
	Gen. PHYS1	Gen. MENT1	Gen. PHYS2	Gen. MENT2
<hr/>				
Common factor correlations				
Gen. PHYS1	1			
Gen. MENT1	0.795	1		
Gen. PHYS2	0.795	0.642	1	
Gen. MENT2	0.642	0.795	0.795	1
Common factor means	0.000	0.000	0.131	0.104

Goodness of overall fit measures: χ^2 (112) = 346.858, RMSEA = 0.048 (0.042 – 0.054), SRMR = 0.040, CFI = 0.990, NFI = 0.986, NNFI = 0.990.

Abbreviations: PF – physical functioning; RP – role physical; BP – bodily pain; GH – general health; VT – vitality; SF – social functioning; RE – role-emotional; MH – mental health.

Gen Phys and Gen Ment are the General Physical and General Mental latent factors.

^aThe parameter estimates when using King-Kallimanis et al. approach to control for Type I error.

Table 5

Schmitt method: goodness of fit and χ^2 differences between models (n = 909)

Model		Df	CHSQ	RMSEA	SRMR	CFI	NFI	NNFI
Model 1		84	301.284	0.054 (0.048 – 0.061)	0.034	0.991	0.988	0.987
Model 2	Covariances constrained (reconceptualization)	87	305.641	0.053 (0.047–0.059)	0.038	0.991	0.988	0.988
Model 3	Factor covariances and variances constrained (recalibration)	91	310.750	0.052 (0.046–0.058)	0.040	0.991	0.987	0.988
Model 4	Factor covariances, variances, and loadings constrained (recalibration)	98	321.021	0.050 (0.044 – 0.057)	0.041	0.991	0.987	0.989
	PF	97	317.115	0.050 (0.044 – 0.057)	0.041	0.991	0.987	0.989
Model 5	Factor covariances, variances, loadings, and uniqueness constrained	106	338.198	0.049 (0.043 – 0.055)	0.039	0.991	0.986	0.989
	PF	105	329.752	0.049 (0.043 – 0.055)	0.039	0.991	0.987	0.990

Abbreviations: RMSEA – root mean square error of approximation, SRMR – standardized root mean square residual, CFI – comparative fit index, NFI – normed fit index, NNFI – non-normed fit index.

RMSEA $\leq$ 0.05 close fit; 0.05 to 0.08 reasonable fit; $\geq$ 0.10 poor fit

CFI, NFI, and NNFI $>$ 0.90 reasonably good fit, SRMR $<$ 0.10 is favorable.