ASSESSMENT OF RESPONSE SHIFT IN PATIENTS TREATED WITH A CALCIUM ANTAGONIST- OR AN ATENOLOL-LED HYPERTENSION STRATEGY IN THE INTERNATIONAL VERAPAMIL-TRANDOLAPRIL STUDY

By

PRANAV KIRIT GANDHI

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2010

To my mom and dad

ACKNOWLEDGMENTS

Time spent as a graduate student at the University of Florida constituted some of the most important and in many ways the life shaping years of my life. I would like to thank my Ph.D. advisor Dr. L. Douglas Ried for his constant guidance and support in my work. His commitment to work was always inspiring and I have learned a lot under his guidance. I would also like to thank my Ph.D. committee consisting of Dr. Carole Kimberlin, Dr. Teresa Kauf, and Dr. I-Chan Huang for providing expertise and positive, constructive feedback. I am also thankful to Dr. Richard Segal who provided me constant support, guidance and appreciation for the work I did.

I believe unwavering support and encouragement from my family has been the key ingredient in all the successes throughout my life. I am and always shall remain short of words to thank my parents for all their love, care, and personal sacrifices. They have always supported and believed in me through everything. I would like to thank my younger sister and her husband for their constant support and a source of inspiration in my life. In the end I would like to thank my wife Poonam for her unconditional support, understanding, and for putting faith in me. She completes me and has always supported and believed in me.

TABLE OF CONTENTS

page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A	Attribute variables
BP	Bodily Pain
CABG	Coronary artery bypass graft
CAD	Coronary artery disease
CES-D	Center for Epidemiologic Studies Depression
CFA	Confirmatory factor analysis
CFI	Comparative fit index
CHF	Congestive heart failure
CI	Confidence Interval
COPD	Chronic Obstructive Pulmonary Disease
d	Effect size
DBP	Diastolic blood pressure
df	Degrees of freedom
Diag (Ω)	Residual factor variances
E	Explanatory variables
GH	General Health
H_a	Alternative hypothesis
H_0	Null hypothesis
HRQoL	Health related quality of life
IADL	Instrumental activities of daily living
ICF	International Classification of Function
INVEST	International Verapamil-Trandolapril Study
JNC VI	Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure
LISREL	Linear Structural Relationship

MCMC	Markov Chain Monte Carlo
MENT	Mental Health construct
MH	Mental Health
MI	Myocardial infarction
MLE	Maximum likelihood estimation
n	Sample size
NFI	Normed fit index
NHP	Nottingham Health Profile
NNFI	Non-normed fit index
p	Significance value
P	Number of indicator variables
Patt (Γ)	Common factor loadings
PF	Physical Functioning
PHYS	Physical Health construct
q	Number of parameters to be estimated
QOL	Quality of life
RMSEA	Root mean square error of approximation
RP	Role Physical
SADD-Sx	Study of Antihypertensive Drugs and Depressive Symptoms
SBP	Systolic blood pressure
SD	Standard deviation
SEM	Structural equation modeling
SF	Social Functioning
SF-36	Short form-36
SRMR	Standardized root mean square residual

Tau (τ)	Intercepts
UF	University of Florida
V	Potential violator variables
VT	Vitality
X	Domain scores
z-scores	Standardized scores
α_1	Common factor means at baseline
α_2	Common factor means at one year
β-blocker	Beta blocker
$\mu_2 - \mu_1$	Observed change
χ^2	Chi-square

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

ASSESSMENT OF RESPONSE SHIFT IN PATIENTS TREATED WITH A CALCIUM ANTAGONIST- OR AN ATENOLOL-LED HYPERTENSION STRATEGY IN THE INTERNATIONAL VERAPAMIL-TRANDOLAPRIL STUDY

By

Pranav Kirit Gandhi

December 2010

Chair: L. Douglas Ried
Major: Pharmaceutical Sciences – Pharmaceutical Outcomes and Policy

Numerous reliable and valid measures for HRQoL assessments are available but these assessments may be subject to change over time or after an intervention from factors separate from the effects of the intervention itself. These changes are referred to as response shifts. We investigated the occurrence of a response shift using short form-36 (SF-36) HRQoL instrument in hypertensive patients with coronary artery disease (CAD). First, two structural equation modeling (SEM) techniques were compared and examined for convergence to examine plausible occurrence of response shift over two time points (i.e., at baseline and at one year). We provide alternate explanations for the divergent results we found compared to the Ahmed et al. study.

Our study identified recalibration for the SF-36 physical function (PF) scale using both Oort and Schmitt approaches. Based on our results, the differences between the Oort and Schmitt approaches in our study may be due to variation in the method and not the sample used. Method differences may result from divergence in defining the type of response shift linked to changes in various parameters of the SEM measurement model. Our finding of response shift may be explained by a single or a

combination of factors including the disease condition, assignment of antihypertensive treatment strategies, age, gender, and/or the presence of depressive symptoms.

Second, we investigated measurement bias and response shift jointly with use of SEM over a one year period. We found that gender caused response shift in measurement. Women have a higher incidence of depression, and in combination with lower PF scores would worsen their quality of life (QOL). Thus, it is imperative to improve depression-caused reductions in PF and maintain an individual's ability and willingness to perform daily activities which may help reduce institutionalization. Our results suggest that the SF-36 PF scale may be susceptible to response shift – this deserves further research. By looking more closely at the scores for SF-36 PF domain in this study population will enable us to provide nuanced attention and direct treatment for the most impaired aspects of QOL.

CHAPTER 1
INTRODUCTION

Health related quality of life (HRQoL) is self-perceived and can change over time. It has been defined as 'the value assigned to duration of life as modified by the impairments, functional states, perceptions and social opportunities influenced by disease, injury, treatment or policy' [1]. Self-report measures of HRQoL are increasingly becoming part of the assessment profile for impact of drug interventions in randomized trials. Although numerous reliable and valid measures for HRQoL assessments are available, these assessments may be subject to change over time or after an intervention from factors separate from the effects of the intervention itself [2]. In other words, individuals experiencing change in health status owing to improvement or deterioration may adopt a different frame of reference over time and reevaluate the importance of fundamental elements of HRQoL [2]. These changes are referred to as response shifts. Response shifts refers to a change in meaning of quality of life (QOL) over time. Consequently, self-report assessments may over- or underrate the true effects of the intervention [3]. The identification of response shift may enable the assessment of relative importance of HRQoL outcomes from the point of view of patients and offer valuable insights into cost-effectiveness analysis of the various intervention strategies.

The work here investigated the occurrence of a response shift in hypertensive patients with coronary artery disease (CAD) treated with antihypertensive treatment strategies. First, two structural equation modeling (SEM) techniques were compared and examined for convergence to examine plausible occurrence of response shift in the study population over two time points (i.e., at baseline and at one year). Second, we

investigated measurement bias and response shift jointly with use of SEM in the study population over a one year period. The present study utilized short form-36 (SF-36) measurement instrument for HRQoL to achieve both objectives aforesaid.

Background

Types of Response Shift

In the beginning, response shift research was based in evaluation of educational training, organizational change, and management science. More recently, the concept has been utilized in HRQoL research [4-6]. Response shift is defined as “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., scale 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)” [4,5].

Reconceptualization. With reconceptualization, the patient revises the meaning of the item content. For instance, during post hip replacement surgery, an individual may consider the ability to work as important to his HRQoL. Prior to the injury, however, the ability to work was not even considered in the individuals’ conceptualization of HRQoL. The domains of importance to this individuals’ HRQoL changed, reflecting reconceptualization [4,5].

Recalibration. With recalibration, the patient revises the meaning of the scale’s item response values [4,5]. For instance, in the immediate period post-angioplasty, even though an individual is in pain and unable to engage in vigorous activity, he/she may judge current health as good. The same person who at 3 months post-angioplasty is able to perform vigorous activities may now look back and judge his/her initial health as

poor (scale recalibration). Recalibration can be uniform or non-uniform. Non-uniform recalibration occurs when a portion of the measurement scale is recalibrated (e.g. only some response options are associated with greater pain than before). Conversely, uniform recalibration occurs when the entire scale is recalibrated (e.g. all points of the response scale are associated with greater pain than before) [4,5].

Reprioritization. With reprioritization, an individual, for example, might initially value the HRQoL concepts of social network, work and family as important [4,5]. After a health scare (i.e., change in health status), the same individual may find the same concepts important, but the order of importance (i.e. change in values) may change to family first, social network second, and lastly, work, indicating reprioritization.

Importance of Response Shift in Clinical Trials

Randomized trials offer the benefit of balancing measured and unmeasured variables, which may include conceptualization of HRQoL and internal standards [2]. In the present study, in addition to hypertension, CAD and elderly age, most patients (89%) had one or more associated conditions (diabetes, dyslipidemia, cerebral or peripheral vascular disease, etc.) contributing to increased risk for adverse outcome [7,8]. Moreover, patients who deal with treatment strategies and chronic disease are faced with the necessity to cope with the consequences of the treatment modalities and/or the illness. As a result, a response shift may have occurred in these patients.

Response shift in measurement of HRQoL is defined as an adaptation to changing health [9]. It may be a beneficial process for patients because it can help in adapting to a new situation. To measure change in HRQoL, the baseline (pretest) score is usually subtracted from the score after implementation of the intervention (posttest) [9]. The supposition that the administration of the same HRQoL instrument at two points in time

or between two different groups is, in fact, measuring the same phenomenon may be untrue [2]. This strategy assumes that the individuals' understanding of HRQoL and their way of interpreting it has not changed over the relevant time frame [10]. Sprangers et al. [4] state that the variability observed due to response shift may reflect shifts in an individual's internal standards, in values and priorities, or in conceptualizations of QOL in addition to changes in actual health state. As a result, it becomes difficult to separate which component of change is due to response shift as opposed to true change in the HRQoL construct [2]. Thus, HRQoL research must take into account the importance of assessment of response shift in longitudinal/clinical trial studies.

Hypertension and Coronary Artery Disease as Chronic Conditions

Experiencing a response shift is seen as a natural response to change in health status. Ahmed et al. [2] suggest that response shift may attenuate or exaggerate findings from clinical trials that incorporate HRQoL to evaluate treatment interventions. Response shift is the result of mechanisms used to accept or adapt to changes in physical, emotional, social health status [9] and nature of impairments and activity limitations that many hypertensive CAD patients may experience. Hypertension and CAD are chronic health conditions that impact all aspects of function, perception, cognition, mood, and QOL [11-13].

Hypertension and CAD are health conditions where symptoms can persist over a long period of time [11-13]. Literature renders support for the contention that hypertension and CAD are health conditions with changes in physical, pain, and social health status [14]; a likely scenario in which response shift may occur. To obtain patient preferences and their individual adjustment between benefits and side-effects of various treatment options, estimation of response shift can be useful. For example, before

receiving the medication for hypertension the subjects may not have associated QOL with sleeping issues, but now, as a consequence of the medication's side effects, it may become a major problem [15].

Hypertensive medications from different pharmacologic classes sometimes have varying side effect profiles and impact patients' HRQoL in different ways. Several studies report that the choice of antihypertensive treatment may influence risk for depression and consequently clinical and health outcomes [11,16]. Depression is highly prevalent in hypertensive CAD patients and is associated with poorer outcomes [11,17-19]. However, literature regarding the potential relationship between depression and use of antihypertensive medications (e.g. β-blocker (beta blocker) and calcium channel antagonist) have inconsistent findings regarding whether medications from one or both pharmacologic categories are associated with depression [19,20]. Some studies report that rates of depression are higher for β-blockers than for calcium antagonists [21]. These conclusions are based on findings that antidepressant prescriptions were more likely to be prescribed following β-blocker treatment [22,23]. In other instances, these findings have not been replicated [24,25]. However, other studies report an association of calcium channel antagonists with depression [26]. Response shift is important to consider in treatment evaluations, especially as it may serve to attenuate or to exaggerate estimates of treatment effects as patients adapt to treatment modalities or disease progression over time.

Structural Equation Modeling and Response Shift

Confirmatory factor analysis (CFA) is a type of SEM that deals specifically with measurement models, i.e. the relationships between observed measures or indicators (e.g., test items, test scores) and latent variables or factors, i.e., variables which cannot

be measured directly due to presence of measurement error. Schmitt has described the analysis of covariance structures approach using SEM to identify both reconceptualization and recalibration by evaluating changes in factor structure and factor loading over time, respectively [2,27-30].

An alternate method of determining response shift described by Oort uses SEM to test changes over time [10]. Contrary to Schmitt's SEM approach, Oort's SEM approach provides a direct measure of true change and identifies the presence of all three types of response shift, i.e., reconceptualization, reprioritization and recalibration. Oort operationalizes various types of response shift (Table 1-1) as below.

- **Reconceptualization**. Response shift is suggested by changes in factor loading patterns over time, with an observed variable loading onto one latent variable at one point in time, and at a follow-up time loading onto a different latent variable.
- **Reprioritization.** Reprioritization is demonstrated when the factor loading of an observed variable may become stronger or weaker over time, reflecting changes in values or priorities.
- **Recalibration.** Uniform recalibration is determined by differences between intercepts across occasions whereas non-uniform recalibration is determined by differences between error variances across occasions.

Measurement Perspective of Response Shift

If scale scores of a test instrument are not fully determined by the common HRQoL factors (i.e., SF-36 physical and mental constructs), then it is called measurement bias [31,32]. In other words, when differences between the observed scale scores cannot be fully explained by true differences between respondents in the common HRQoL factors, it is termed as measurement bias [31,32]. In longitudinal research, when the relationships are not consistent across measurement occasions, i.e. over time, the measurement bias is considered to be response shift [31,32]. In

longitudinal analysis, measurement bias is investigated by testing the invariance of factor loadings and intercepts [31,32].

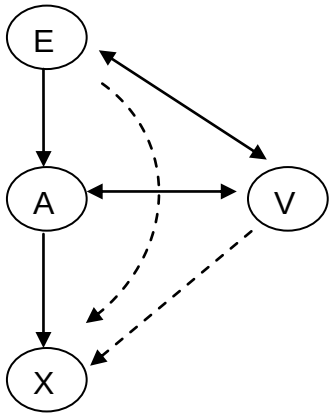

Figure 1-1. Graphical representation of measurement bias and response shift

All circles in Figure 1-1 represent sets of variables: A are the attributes of interest; X are measurements of A; E are explanations of A; and V are all other variables. One-sided arrows represent causal relationships and double-sided arrows represent mere associations. Dashed arrows indicate measurement bias. Example: A = SF-36 physical health and mental health constructs; X = domains of SF-36 self-report health questionnaire; V = sex, age, race, baseline co-morbid conditions; E = change in systolic blood pressure, etc.

To illustrate Figure 1-1 with an example, bias is introduced from the measurement perspective when the relationship between the measurement instrument (i.e., observed variables such as SF-36 domains, denoted by X) and other variables (potential violator variables such as age, sex, race, denoted by V) cannot be fully explained by the relationships between explanatory variable such as change in systolic blood pressure (denoted as E) to V, E to common HRQoL factors (denoted by A), A ⟷ V and A ⟶ X.

Similarly, if E → X relationships cannot be fully explained by the E↔V, E→A, A↔V and A→X relationships, then there is measurement bias with respect to variables E.

Importance of Response Shift Assessment in This Study

As aforementioned, in the present study, in addition to hypertension, CAD and elderly age, most patients (89%) had one or more associated conditions (diabetes, dyslipidemia, etc.) contributing to increased risk for adverse outcome [7,8]. These multiple medical conditions place patients at high risk of depressive symptoms. Bush et al. [33] report higher mortality at levels of depressive symptoms not generally considered clinically significant and below levels usually considered predictive of increased postacute myocardial infarction (MI) mortality [17]. We presumed that as time passes and people's functional ability improves, the individuals will retrospectively re-evaluate how they were in the past, subsequently leading to occurrence of a response shift.

In the absence of response shift evaluation, it cannot be determined whether temporal change observed in the HRQoL construct is due to true change or to changes in the structure or measurement of the HRQoL construct [34]. The effect size of HRQoL change can be estimated with the method described by Oort [10], taking response shift into account, to suggest the clinical importance of the response shift that occurs [34]. Some individuals may have made actual changes in HRQoL. Conversely, others may have undergone response shift, thereby influencing the effects of an intervention over time. Barclay-Goddard et al. [34] report that presence of response shift may over or under-estimate true change leading to biased estimates of the magnitude of change. Moreover, it is important to measure response shift in clinical trial research, as the

estimates of the treatment effects may be underestimated and therefore inaccurate, possibly leading to a conclusion that represents a false negative [34].

Methods that assess response shift are not only necessary in valuating treatment effects, but also provide assistance for distinguishing the impact of disease over time [34]. Ahmed et al. [35] demonstrate the convergence among Oort and Schmitt's SEM approaches to detect response shift in chronic obstructive pulmonary disease (COPD) patients. The authors report Oort procedure to be more sensitive in detecting a response shift compared to the Schmitt procedure. However, the Ahmed et al. study [35] did not control for Type I error even though the Oort's SEM approach has been suggested to increase chance of a Type I error [10]. As a result, their study may have wrongly identified a response shift when differences may be attributed to other types of changes. Contrary to the Oort procedure adopted by Ahmed et al. [35] in their study where they chose to release constraints on residuals, followed by intercepts, and then factor loadings, we conduct the Oort approach as recommended and tested by several other authors including Oort [10,36,37], Donaldson [38] and Barclay-Goddard [39]. Similar to the approach tested by these authors, we test sequentially for invariance of factor loadings, intercepts, and error variance. The divergence in the Oort procedure adopted for releasing constraints among our study and Ahmed's study may result in conflicting results. Thus, our first goal was to compare Oort and Schmitt's SEM approaches to test if conclusions regarding response shift were supported in our entire dataset.

A self-report instrument measuring change in health status over time highlights the concern that respondents may change their frame of reference when answering the

questionnaire or item. This may result in incomparable scores from separate measurement occasions. Previous studies [2,27,36,37,39] report paradoxical and inconsistent findings when using the SF-36 measurement instrument to examine response shift. The presence or absence of response shift in one or more of the SF-36 domains may thus be a reflection of an explanatory or violator variable explained previously (shown in Figure 1-1). Our second goal was to highlight the differences between the measurement perspective about bias and response shift in HRQoL data from hypertensive CAD patients using SEM approach suggested by Oort [31].

When investigating from the measurement perspective, the SF-36 scales should measure the physical and mental constructs exclusively. If, however, differences between respondents in SF-36 scales scores cannot be fully explained by true differences between respondents in the physical and mental constructs, it indicates measurement bias [31,32]. In other words, the scale scores will not only be indicative of the common HRQoL factors [31,32] but also of some other variable such as treatment strategies or comorbid conditions. The SF-36 instrument may thus include domain/s to which a variable responds differently to. This study is the first to illustrate the detection of bias and response shift from the measurement perspective in hypertensive CAD patients' HRQoL data, based on SF-36.

Study Objectives

- To determine convergent validity among Oort and Schmitt's SEM approach when examining response shift in the entire sample of hypertensive CAD patients over a one year period.
- To investigate measurement bias and response shift in measurement jointly in HRQoL data from hypertensive CAD patients using SEM over a one year period.

Study Significance

To our knowledge, this is the first study to assess the occurrence of response shift in hypertensive CAD patients assigned to antihypertensive treatment strategies. A previous study that demonstrated the comparison of Oort and Schmitt's SEM approaches to identify response shift did not account for Type I error [35]. In addition, they chose to release constraints on residuals, followed by intercepts, and then factor loadings. This approach is divergent compared to the Oort approach recommended and tested by several other authors including Oort [10], Barclay-Goddard [39], and Donaldson [38]. We thus chose to compare the two statistical techniques to test if conclusions regarding response shift were supported in our dataset, controlling for Type I error and using the order of testing for response shift as suggested by Oort which will strengthen the acceptability of the results. The comparison of the said statistical approaches will assist in developing a proposed set of HRQoL measurement recommendations under circumstances where response shift is expected to occur in a specific patient population. Previous studies [2,27,36,37,39] report paradoxical and inconsistent findings when using the SF-36 measurement instrument to examine response shift. Though both the antihypertensive treatment strategies in the International Verapamil-Trandolapril Study (INVEST) trial were shown to be clinically equal [8], the presence of response shift may reflect medications' different pharmacological properties and side effects, occurrence of stroke or MI over a one year period, depression scores, self-reported prior history of depression, or the presence of comorbid conditions leading to varying effects on HRQoL (as shown in Figure 1-1). Barclay-Goddard suggests that researchers need to come to a generally accepted definition of response shift, using prior work as a basis. Based on these

recommendations, we chose to apply Oort's definition to estimate both measurement bias and response shift using SEM. With response shift present, the respondents' observed scale scores cannot be fully explained by true changes in the latent construct that we want to measure. In other words, the SF-36 scale scores may not only be indicative of the latent construct but also of the variables such as treatment strategies, age, or sex, among others.

Research Questions

- Was convergent validity observed when investigating response shift between Oort's SEM approach and Schmitt's SEM approach in hypertensive CAD patients over a one year period?
- Was measurement bias and response shift observed using SEM in hypertensive CAD patients over a one year period?

Table 1-1. Types of response shift as observed with Oort's SEM approach

Matrix	Parameter	Type of change
		Response shift
Patt(Γ)	Common factor loadings	Reconceptualization
Γ	Common factor loadings	Reprioritization
τ (tau)	Intercepts	Recalibration (uniform)
Diag(Ω)	Residual factor variances	Recalibration (non-uniform)

CHAPTER 2
LITERATURE REVIEW

Unification of Incoherent Findings in Quality of Life and Emergence of Response Shift

Howard et al. [40] introduced the term response shift when investigating educational training interventions. Prior to their study, application of response shift was found in organizational change within the field of management sciences, where Golembiewski et al [41] discussed three types of observed changes in self report ratings: alpha, beta, and gamma change. “Alpha change involves a variation in the level of some existential state, given a constantly calibrated measuring instrument related to a constant conceptual domain [41].” Alpha change is considered true behavioral change, typically measured with pretest-posttest study designs [41]. “Beta change involves a variation in the level of some existential state, complicated by the fact that some intervals of the measurement continuum associated with a constant conceptual domain have been recalibrated [41].” “Gamma change involves a redefinition or reconceptualization of some domain, a major change in the perspective or frame of reference within which phenomena are perceived and classified, in what is taken to be relevant in some slice of reality [41].”

Sprangers and Schwartz [4,5] applied the idea of response shift to the field of HRQoL research. They combined and extended the previous definitions of response shift and proposed the following definition: response shift refers to a change in the meaning of one’s self-evaluation of a target construct as a result of: (1) recalibration, that is, a change in the respondent’s internal standards of measurement; (2) reprioritization, that is, a change in the respondent’s values (i.e., reevaluation of the importance of component domains constituting the target construct); or (3)

reconceptualization, that is, a redefinition of the target construct. A change in values (termed as reprioritization) was not described in the alpha, beta, gamma framework, but was added by Schwartz and Sprangers in 1999 [4,5].

Over the years, contradictory findings about what QOL measures and how scores obtained from these measures need to be interpreted have raised several questions. For example, people with severe chronic illnesses report QOL equal or superior to less severely ill or healthy people [42-44] and inconsistencies persist between clinical measures of health and patients' self-report [42,45,46]. Some studies on HRQoL found that patients score better than healthy people do, or that significantly disabled or terminally ill people report similar or higher levels of QOL after deteriorating health. These counterintuitive findings may reflect individual differences and intra-individual changes in internal standards, values, and meaning of QOL due to response shift. QOL research has a lot to gain by using methods that include the response shift phenomena. Theoretical and empirical evidence on response shifts in QOL support the concept that differences in evaluation enter into all self-report assessments of QOL. Next, we summarize the QOL studies that have identified response shift followed by studies that merit response shift assessment in our study population. We then demonstrate the literature illustrating use of Schmitt and Oort SEM techniques to detect response shift. Lastly, we elaborate on the issue of measurement bias and response shift in presence of explanatory and violator variables.

Summary of Studies that Identify Response Shift

Several studies [47-53] document the presence and importance of response shift in both treatment outcome research and naturalistic longitudinal observations of QOL. However, only a few studies are cited here. Researchers are recommended to refer to

Rapkin and Schwartz [42] for other studies that examine the presence of response shift. Sprangers and colleagues [47] report occurrence of recalibration for fatigue in two subgroups of cancer patients undergoing radiotherapy: in patients experiencing diminishing levels of fatigue and in patients facing early stages of adaptation to increased levels of fatigue. Schwartz and colleagues [52] report a harmful QOL effect of a psychosocial intervention was due to recalibration and reconceptualization. Rees and colleagues [53] report a 10% increase in QOL with addition of recalibration in prostate cancer patients. Several researchers have argued that beta change potentially threatens the internal validity of judgments about change based on self-reports [54-58].

King et al. [59] compared overall HRQoL of stroke survivors to that of normative controls and found that HRQoL scores were comparable. A comparison of studies evaluating the outcome of stroke found that persons with major stroke reported higher levels of HRQoL than individuals with moderate stroke [59,60]. In one other study, patients with cancer report better levels of HRQoL than the general population [59,61]. People with spinal cord injuries were found to report only slightly worse levels of HRQoL compared to controls [59,62]. Primarily, these findings provide substantiation that people's goals and values continue to evolve during progression of disease, thus warranting assessment of response shift in longitudinal research and clinical trials.

Importance of Response Shift Assessment among Hypertensive CAD patients

Hypertension and CAD are chronic health conditions that impact all aspects of function, perception, cognition, mood, quality of life [11-13]. In a study by Mitchell et al. [14], poorer mean scores were reported in hypertensive CAD patients when using the Nottingham Health Profile (NHP) questionnaire. The study was conducted among

elderly subjects that had differing cardiovascular status (cardiovascular normal, being hypertensive, having isolated CAD, or both being hypertensive and having CAD).

The NHP questionnaire incorporates questions related to activity limitations and impairments similar to the classification provided by the World Health Organization's International Classification of Function (ICF) [63]. Under the ICF model, difficulties with psychological functioning include impairments in cognitive or emotional functioning. Impairments such as decreased strength cause limitations in activities, such as walking, which eventually may lead to participation restrictions, i.e. restricted ability to carry out usual activities in the community, such as through work or volunteering. Response shift can result due to mechanisms used to accept or adapt to changes in health status and nature of impairments and activity limitations that many hypertensive CAD individuals are likely to experience.

In addition to hypertension, CAD and elderly age, most INVEST trial patients (89%) had one or more associated conditions (diabetes, myocardial infarction, abnormal coronary angiogram, etc.) contributing to increased risk for adverse outcome [7,8]. These patients with comorbidities place them at high risk of depressive symptoms. An essential outcome of hypertension and CAD, and their treatment is mental depression [17]. Literature supports the notion that depression is a risk factor for death and myocardial infarction among patients with CAD [17,33]. Moreover, higher mortality has been observed at levels of depressive symptoms not generally considered clinically significant and below levels usually considered predictive of increased postacute MI mortality [17,33].

Impact of Hypertension and CAD on QOL

QOL in patients with hypertension has been extensively evaluated in recent years. A study by Wang et al. [64] report hypertensive patients to score lower in five SF-36 domains compared to normotensives. The authors observed hypertensive subjects with comorbidities to score lower SF-36 scores compared to those without comorbidities. Boini et al. [65] report that CAD patients were more likely to score lower in four of the eight SF-36 domains. The authors agree that the study of HRQoL in CAD patients may be used to inform treatment decisions to practitioners. Cost-effectiveness studies of antihypertensive modalities often utilize HRQoL outcomes, thereby suggesting the need to assess HRQoL in these patients. During a change in patient's HRQoL over time there exist a possibility for the patient's to reevaluate their opinion toward past events. Therefore, facilitating response shift assessment is particularly important in hypertensive CAD patients when impairments and activity limitations are not expected to recover fully, but improved HRQoL is a goal.

Impact of Antihypertensive Treatment Strategies on QOL and Need to Examine Response Shift

Antihypertensive treatment focuses on individual attributes/domains of HRQoL which are important to the individual hypertensive CAD patient. Croog et al. [13] have demonstrated that effects on QOL vary among classes of antihypertensive agents: captopril improved QOL, whereas methyldopa and propranolol worsened it. In a separate study, Testa et al. [66] conclude that two angiotensin-converting enzyme inhibitors, captopril and enalapril, the same according to clinical assessments of efficacy and safety, had varying effects on QOL. Assessment of HRQoL in cardiovascular clinical trials is important to provide a more complete understanding of treatment effects.

Hypertensive medications from different pharmacologic classes sometimes have varying side effect profiles and impact patients' HRQoL in different ways. Several studies report that the choice of antihypertensive treatment may influence risk for depression and consequently clinical and health outcomes [11,16]. Response shift is thus important to consider in treatment evaluations, especially insofar as it may serve to attenuate or to exaggerate estimates of treatment effects as patients adapt to treatment toxicities or disease progression over time. To obtain patient's preferences and their individual adjustment between benefits and side-effects of various treatment options, estimation of response shift can be useful.

Based on INVEST trial [8] results, patients in the verapamil-SR group reported constipation more frequently than in the atenolol-led group. A study by Wald et al. [67] reports the impact of constipation on QOL in different cultural and national settings. HRQoL was assessed with the SF-36 questionnaire. There were significant differences in HRQoL between constipated and non-constipated individuals and a significant, negative correlation between the number of symptoms and complaints and SF-36 scores. The authors observed that constipated individuals consistently reported low scores in the psychological components of the SF-36 questionnaire. On the other hand, atenolol-led treatment strategy has been reported to show adverse somatic side effects such as decreased energy and increased fatigue [17]. The side-effects associated with verapamil and atenolol may thus cause a change in HRQoL over time in the study population.

Ried et al. [17] demonstrate a significant improvement in depressive symptoms for those assigned to the verapami-SR strategy; whereas there was little change for those

assigned to the atenolol strategy among a subgroup of INVEST patients. They explain that changes in depressive symptoms at one year in the verapamil-SR group could be due to verapamil's mood-related effects. Analogous to their [17] findings regarding depressive status changes [17], Ried and colleagues [68] in a separate study report that HRQoL improved among those assigned to the verapamil-SR-led strategy, but found no significant change in HRQoL for those assigned to atenolol-led strategy.

A study by Bar-On and Amir [15] identified the presence of response shift in hypertensive patients. Bar-On and Amir [15] examined beta change (i.e. recalibration) over the period of one year in hypertensive and normotensive male subjects randomly assigned to methyldopa, isradipine or placebo treatments. Among the 297 subjects in their study, 11% performed one of the two scales recalibrations. Nonetheless, there were no significant differences in those who showed recalibration between normotensives and hypertensives. When, however, the subjects who showed severe scale recalibration were excluded from the comparison, the difference between the normotensives and hypertensives became significant. To summarize, several studies demonstrate that taking antihypertensive medications over a one year period is potent enough to affect the QOL assessments. Nonetheless, there is modicum of evidence that response shift may be detected in these patients over time.

Assessment of Response Shift Using Schmitt and Oort's SEM Techniques

There are two broader alternatives for detecting response shift, the first is to address response shift phenomenon from imposing study design changes; second alternative is to address response shift from a statistical framework [5]. The study design methods are considered time consuming and burdensome for the individual

involved [5,69]. Statistical methods are often used where it is impractical to incorporate response shift evaluation into the study design or when doing secondary data analysis.

Schmitt's SEM Technique

The analysis of covariance structures approach using CFA, a statistical framework to assess response shift was proposed by Schmitt [28] in 1982. Schmitt's SEM approach [28] identified both reconceptualization and recalibration by evaluating changes in factor structure and factor loading over time, respectively. Ahmed et al. [2,27] apply Schmitt's SEM approach to assess invariance of the SF-36 measurement model over time among 238 individuals with stroke and 392 controls, separately. Response shift was not identified between one and six months even though it was suggested using individualized methods [2]. It was later argued that a reason for not finding response shift with SEM model in Ahmed et al. study [2] was related to the measurement model used in their study.

Contrary to Ahmed et al.'s study [2], a few studies demonstrate the presence of response shift using Schmitt's SEM approach. Schaubroeck and Green [30] demonstrate the presence of gamma change (i.e., reconceptualization) with use of Schmitt's SEM approach. In a separate study, Schmitt et al. [28] identify presence of gamma and beta change by analyzing data collected from 116 individuals who had lost their job, before and after they had secured subsequent employment. In one other study [29], three methods to evaluate response shift were compared – coefficient of congruence (assessment of the similarity between the factor structures of before and after measures), then-test approach and Schmitt's SEM approach. Coefficient of congruence and Schmitt's approach were more consistent in detecting reconceptualization response shift. In the same study, the then-test and Schmitt's

technique both showed indications of recalibration; the coefficient of congruence method did not. The authors recommend the use of then-test or Schmitt's approach compared to the coefficient of congruence technique to identify response shift.

Oort's SEM Technique

A more contemporary work by Oort and colleagues [10] attempt to rectify the issues inherent in previous research using covariance analysis. Oort [10] investigated all three types of response shift by considering variation in the factor loading patterns to reflect reconceptualization, significant differences in the values of the factor loadings to reflect reprioritization, and significant differences in the specific factor mean to reflect recalibration.

With application of Oort's SEM approach, a study [36] of newly diagnosed cancer patients identify the presence of response shift for five SF-36 scales: reconceptualization of general health (GH), reprioritization of social functioning (SF), and recalibration of role-physical (RP), bodily pain (BP), and vitality (VT). In one other study, Visser et al. [37] demonstrate the convergent validity of the then-test, anchor-recalibration and Oort's SEM approach to detect response shift in cancer patients using SF-36 and a Multidimensional Fatigue Inventory. Results showed agreement between the then-test and SEM approach on the absence (six scales) and presence (two scales, SF-36 BP and RP) of response shift in eight of the nine scales. For the ninth scale (SF-36 GH) both methods detected response shift, but in opposite directions. However, the anchor-recalibration task agreed with the other approaches on only the absence of response shift in 4 of the 7 scales. In a separate study using Oort's SEM approach, Barclay-Goddard [39] identify the presence of response shift among 678 individuals at 1, 3, 6, and 12 months post-stroke in mental health (MH) construct using multiple

measures. Uniform recalibration was identified at one year in the SF-36 role emotional (RE) and preference-based stroke index self-esteem. In addition, non-uniform recalibration was identified at 6 months in SF-36 MH and at one year in EQ-5D anxiety/depression, SF-36 MH, and Stroke Impact Scale emotional.

Differences between Schmitt and Oort's SEM Techniques

In recent past, although the Schmitt and Oort SEM approaches were the most often used statistical techniques to identify response shift, only Ahmed et al. [35] demonstrate the convergence among the said techniques. The authors found Oort procedure to be more sensitive in detecting a response shift compared to the Schmitt procedure in chronic obstructive pulmonary disease patients. However, the Ahmed et al. [35] study did not control for Type I error even though the Oort's SEM approach has been suggested to increase the chances of Type I error. As a result, their study may have wrongly identified a response shift when differences may be attributed to other types of changes. In addition, similar to Nolte et al. [70], Ahmed et al. [35] chose to release constraints on residuals, followed by intercepts, and then factor loadings. This approach was divergent compared to the Oort approach recommended and tested by several other authors including Oort [10], Barclay-Goddard [39], and Donaldson [38]. Moreover, Ahmed et al. [35] did not illustrate any approach to assess validation of the model. As Barclay-Goddard et al. [39] point out in their framework article, SEM requires validation of the model to ensure that it is truly a good model for the population and not just the sample. We thus chose to compare the two statistical techniques to test if conclusions regarding response shift are supported in our dataset, controlling for Type I error, releasing constraints as tested by several studies, and illustrating an approach to assess apparent validation which will strengthen the acceptability of the results. The

next paragraph briefly exemplifies the distinction between Schmitt and Oort's SEM approaches in assessment of response shift over time.

The SEM method described by Schmitt does not identify reprioritization (since the covariance analysis used was incapable of identifying changes in definition from change in values) or non-uniform recalibration and it interprets changes in factor loadings differently from the approach described by Oort. In the approach described by Schmitt, constraints are added to the model to identify response shift; in the approach described by Oort, constraints are removed to identify response shift. The constraints added or removed are equality constraints, where a parameter estimate at one time is made to be equal to another time. Given the inconsistent reports from individual studies using Schmitt and Oort's SEM approaches separately and the recent convergent validity study comparing the said statistical methods justify the need to compare these methods to identify response shift in our patient population.

Measurement Bias and Response Shift

Previous sections in this chapter summarize HRQoL studies that report patients score better than healthy people. On the other hand, some HRQoL studies [31,43,71] report that patients may score better after worsening health state. The self-report and multidimensional nature of HRQoL thus necessitates the idea of measurement bias and response shift particularly relevant to the evaluation of the said concepts over time. Patients undergoing a change in health status over time may have different frames of reference when answering a questionnaire or item [31]. Consequently, the measurement may be biased. As Oort et al. [31] explains, when observed differences between respondents' SF-36 domain scores reflect something other than true differences in the latent construct (i.e. HRQoL), it indicates measurement bias. Oort and

King-Kallimanis [31,32] in separate studies explain response shift as a special case of measurement bias which occurs when investigating the change in latent construct. With response shift, observed changes in respondents' test scores may reflect something other than true changes in the latent construct that we want to measure.

King-Kallimanis et al. [32] test an empirical example using cancer patients' HRQoL data with SF-36 to investigate measurement bias and response shift with a simplified model (Figure 2-1), i.e., with use of potential violator variables (V). Six measurement biases were found, five of which were considered response shift. The GH scale appeared more susceptible to response shift in their study. Moreover, they also found patients to have reconceptualized their perception of BP and GH scales. However, to detect measurement bias and response shift using multiple explanatory and violator variables, we chose to empirically investigate measurement bias and response shift using a joint model illustrated in Figure 1-1. To the best of our knowledge, ours is the first study to empirically investigate the presence of measurement bias and response shift using Oort's model as illustrated in Figure 1-1.

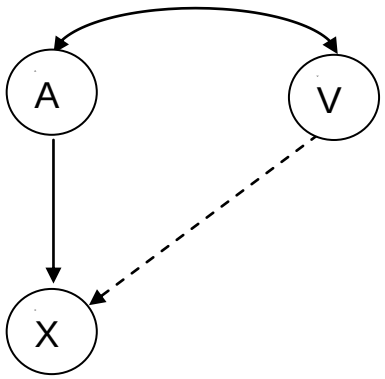

Figure 2-1. Graphical representation of measurement bias and response shift

To conclude, several studies [2,27,36,37,39] cited in preceding sections report paradoxical and inconsistent findings when using SF-36 measurement instrument to

examine response shift. Though both the antihypertensive treatment strategies in the INVEST trial were shown to be clinically equal [8], the presence of response shift on one or more of the SF-36 domains may reflect medications' different pharmacological properties and side effects, presence of comorbid conditions, or the disease condition itself leading to varying effects on HRQoL. The SF-36 instrument may demonstrate a domain to which a variable (e.g., treatment strategies, age, comorbid conditions, etc.) may respond differently to. In other words, the SF-36 scale scores may not only be indicative of the latent construct but also of the variable, thus warranting the need to assess measurement bias and response shift in our study population.

CHAPTER 3
METHODS

Subject and Data Collection Procedure

Source of Data

A description of the Study of Antihypertensive Drugs and Depressive Symptoms (SADD-Sx) is reported elsewhere [17]. Briefly, SADD-Sx was a substudy of 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 [7,8]. 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) [72].

SADD-Sx patients residing in the United States were mailed surveys between April 1, 1999, and October 31, 1999 (N = 2317). SF-36 HRQoL and demographic characteristics such as age, gender and race were included in the SADD-Sx survey and mailed to each study subject. Patients were mailed baseline surveys the day after randomization and follow-up surveys at six months and at one year. If surveys were not returned within 10 working days, they were mailed a second survey. Approximately 2 weeks before follow-up surveys were mailed, a letter asking for continued support was mailed to each patient to enhance response rate. If patients failed to respond to the second survey, no added attempts were made to contact them for the purposes of SADD-Sx. Only patients with complete survey responses from both the baseline and one year follow-up surveys were included in the study's initial sample of patients.

SADD-Sx was conducted according to the principles of the Declaration of Helsinki. The University of Florida (UF) Institutional Review Board approved the study protocol.

SF-36 HRQoL

The SF-36 is a generic measure (Table 3-1) of perceived HRQoL health status that incorporates behavioral functioning, subjective well-being and perceptions of health by assessing eight health concepts: Physical Functioning (PF) (10 items), Role-Physical (RP), limitations in role activities due to physical health problems (4 items), Bodily Pain (BP) (2 items), General Health (GH) (5 items), Vitality (VT) (energy & fatigue) (4 items), Social Functioning (SF) (2 items), Mental Health (MH) (5 items), Role Emotional (RE), limitations in usual role activities due to emotional problems (3 items) [73,74]. All items are measured using Likert-type scales. Response choices varied and range from two to six levels. The transformed scores on all of the health concept scales range from 0 to 100 with higher scores indicating better health [73,74]. The SF-36 scales formed the basis of developing the measurement models for each construct.

The eight SF-36 domains used in the measurement model were assessed for normality. The maximum likelihood estimation (MLE) method yields a χ^2 test of overall of goodness-of-fit, and standard errors for all parameter estimates when data is normally distributed, using the un-transformed scores of the SF-36. On the other hand, if the data was not normally distributed, the robust MLE method was used to account for non-normality in data. The Satorra-Bentler χ^2 is the commonly used fit statistic with the robust MLE method. When data are continuous or ordered categorical non-normal data, the Satorra-Bentler χ^2 performs better than the χ^2 from non-robust methods [75].

Statistical Approach for Research Question 1

Oort's SEM Approach

Oort's SEM framework was used to carry out a CFA using Linear Structural Relationship (LISREL) version 8.8 [76]. The latent factors in the model were the physical health (PHYS HRQoL) and mental health (MENT HRQoL) constructs. The eight scales of the SF-36 were the observed variables explained by the latent variables.

Procedure

The framework here was built on the methodology illustrated by Oort [10]. Specification of response shift detection and true change was done in four steps [10]: (1) establishment of an appropriate measurement model, (2) fitting a model of no response shifts, (3) detection of response shifts, and (4) assessment of true change.

Model 1: Establishing an Appropriate Measurement Model

Based on the content of the items, results of an exploratory factor analysis and based on the published results of principal components analyses of the SF-36, we created a measurement model. Model 1 had no across occasion constraints.

Model 2: No Response Shift Model

Building Model 2 included constraining the values of common factor loadings (reprioritization), the intercepts (uniform recalibration), and residual variances (nonuniform recalibration) across time. Two models were compared, the free model in which parameter estimates (paths, error variances, and intercepts) were all freely estimated, and the fully constrained model in which all of the paths, error variances, and intercepts were made to be equal across time. The χ^2 difference test evaluated the change in overall fit of the free and fully constrained (no response shift) models. If the difference in fit between models 1 and 2 was not significant, we concluded that there

were no response shifts, and skip model 3. When the χ^2 difference test was statistically significant, the null hypothesis that the two models fit equally well was rejected, providing evidence that response shift was present in the data.

Model 3: Testing for Response Shift

In Model 3 all measured response shifts were accounted for. The specification search of Model 3 started with Model 2, and was be guided by modification indices and standardized residuals. Inspection of modification indices and standardized residuals indicated which of the equality constraints were not feasible. Each modification was tested with the χ^2 difference test. First, factor loading constraints were released. This was followed by testing of differences between intercepts (uniform recalibration) while still constraining common factor loadings across time. Lastly, error variances (non-uniform recalibration) were released keeping equal intercepts and common factor loadings across time. Across occasion differences between common factor loadings (in Γ), intercepts (in τ), and residual variances (in Ω) were interpreted in accordance with the operationalization of response shifts. That is, common factor loadings that changed from zero into non-zero (or vice versa) indicated reconceptualization, other changes in common factor loadings indicated reprioritization, changes in intercepts indicated uniform recalibration, and changes in residual variances indicated non-uniform recalibration.

Model 4: Identification of True Change and Effect Size Calculation

To investigate change in the means, variances, and correlations of the common factors, we fitted additional models with Model 3 as the starting point. The across occasion invariance of parameters were tested step-by-step, maintaining all equality constraints that proved tenable. After establishing (partial) invariance of factor loadings,

intercepts, and residual variances, we tested other types of invariance, including the equality of common factor means, variances, and correlations.

Identification of all model parameters, scales and origins of the common factors were established by fixing the means at zero and the variances at one. In Steps 2-4 of the detection procedure, only the baseline factor means and variances were fixed; second occasion means and variances were then identified by constraining intercepts and factor loadings to be equal across occasions.

Oort describes true change as a label for a change in the respondent's level of the target construct. In case of HRQoL, positive true change refers to better health or improved HRQoL. The common factors were seen as operationalization of target constructs, so that the change in the common factor means α was used as operationalization of true change. The statistical significance of true change was tested with a χ^2 difference test. If the null hypothesis of invariant common factor means ($\alpha_1 = \alpha_2 = 0$) was rejected, then the across occasion difference between the common factor means $\alpha_2 - \alpha_1 = \alpha_2$ were taken as a measure of true change (since the first occasion factor means α_1 was fixed at zero).

Evaluation of response shifts and true change χ^2 difference tests were used to evaluate the statistical significance of response shifts and true change. It was also informative to evaluate the sizes of response shift and true change effects on observed change. According to Oort, $\mu_2 - \mu_1$ is the model for observed change, and we assumed $\alpha_1 = 0$ (for Models 3 and 4). Observed change was decomposed into three components, $\mu_2 - \mu_1 = (\tau_2 - \tau_1) + (\Gamma_2 - \Gamma_1)\alpha_2 + \Gamma_1\alpha_2$ where the first component, $(\tau_2 - \tau_1)$, constituted the contribution of recalibration, the second component, $(\Gamma_2 - \Gamma_1)\alpha_2$, constituted the

contribution of reconceptualization and reprioritization, and the third component, $\Gamma_1\alpha_2$, constituted the contribution of true change to observed change.

The parameter estimates of Model 4 (or, if preferred, those of Model 3), were used to estimate the components of observed change. The parameter estimates of the final model, Model 4, were used to calculate effect size indices for true change, as well as for the contributions of response shifts and true change to observed change. Division by the estimated standard deviation of observed change yielded effect-size indices d. Effect size values of d = 0.2, 0.5, and 0.8 were considered 'small', 'medium', and 'large' respectively. The response shift effect on the estimation of true change was investigated by comparing estimates of true change from a model in which response shifts were accounted for (Model 4) with estimates from a model in which response shifts were not accounted for (Model 2).

Controlling for Type I Error

The probability of making at least one Type I error among a family of tests (i.e., familywise error rate) increases when multiple tests are conducted [39]. A family can be defined to include all of the inferences that are made in a study; in the present study, all the constraints freed in the model constituted a family. The Bonferroni correction is one well-known approach to control the familywise error rate to the nominal level of significance, α [39]. Under this method, the statistical significance of each result is evaluated at the α/p level of significance, where p is the number of tests (i.e. hypotheses) being conducted.

Model Evaluation

The acceptability of the fitted CFA solution was evaluated on the basis of three major aspects: (1) overall goodness-of-fit; (2) the presence or absence of localized

areas of strain in the solution (i.e. specific points of ill fit); and (3) the interpretability, size, and statistical significance of the model's parameter estimates. Two statistics frequently used to identify focal areas of misfit in a CFA solution are residuals and modification indices [75] were also applied in our study.

Overall goodness-of-fit. The fit was assessed in numerous ways. When model χ^2 increases, the fit of a model become worse. A statistically significant χ^2 supports the alternative hypothesis, meaning that the model estimates do not sufficiently reproduce the sample variances and covariances (i.e. the model does not fit the data well). A non-significant χ^2 means there is little difference between the models, suggesting a good fit.

The normed χ^2 statistic, the ratio of the χ^2 statistic to its degrees of freedom, decreases the effect of the sample size on the χ^2; values up to 5.0 suggest a reasonable model fit [75]. Fit indices can be broadly characterized as falling under three main categories: absolute fit, adjusting for model parsimony, and comparative or incremental fit. We reported at least one index from each category when evaluating the fit of their models.

Standardized root mean square residual (SRMR) falls in the category of absolute fit. The SRMR can take a range of values between 0.0 and 1.0, with 0.0 indicating a perfect fit (i.e. the smaller the SRMR, the better the model fit). A widely used index from the category of parsimony correction is the root mean square error of approximation (RMSEA). According to a generally accepted rule of thumb, an RMSEA value below 0.08 indicates 'reasonable' fit and one below 0.05 'close' fit [75]. As with the SRMR, RMSEA values of 0 indicate perfect fit and values very close to 0 suggest good model fit. In addition, a 90% interval is typically used to obtain confidence intervals for RMSEA.

The confidence interval indicates the precision of the RMSEA point estimate. Comparative fit index (CFI) is an example of comparative fit indices. The CFI has a range of possible values of 0.0 to 1.0 with values closer to 1.0 implying good model fit [75]. Similarly, normed fit index (NFI) and non-normed fit index (NNFI) have a range of possible values of 0.0 to 1.0 with values closer to 1.0 implying good model fit [75].

Residuals. Standardized residuals were computed using the LISREL 8.8 software [75]. Large, positive standardized residuals indicated that additional parameters are needed in the model to better account for the covariance between the indicators. Conversely, a negative standardized residual suggested that the model's parameters overestimate the relationship between two indicators to some extent. Since standardized residuals can be roughly interpreted as standardized scores (z-scores), the z-score values that correspond to conventional statistical significance levels were employed as practical cutoffs. For instance, we scanned for standardized residuals that were equal to or greater than the absolute value of 1.96 because this value corresponds to a statistically significant z score at $p < .05$. In general, larger N's are associated with larger standardized residuals because the size of the standard errors of the fitted residuals is inversely related to sample size. For this reason, we looked for larger cutoff values (i.e. 2.58) [75].

Modification Indices. Modification indices were computed for each fixed parameter (e.g. parameters that are fixed to zero such as indicator cross-loadings and error covariances) and constrained parameter in the model. The modification index reflects an approximation of how much the overall model χ^2 would decrease if the fixed or constrained parameter was freely estimated. A good-fitting model should produce

modification indices that are small in magnitude. Modification indices of 3.84 or greater suggest that the overall fit of the model could be significantly improved ($p < 0.05$) if the fixed or constrained parameter was freely estimated [75].

Sample Size Consideration

The measurement model that was used in the response shift detection procedure assumed the extreme value of RMSEA = 0.08 which indicates the fit of the model is at least mediocre. With a RMSEA value of <0.08, $\alpha = 0.05$, a desired power of at least 0.80 and assuming the degrees of freedom (df) to be > 20 would require a minimum sample size of 421 [75]. Degrees of freedom (df) = $p (p + 1)/2 – q$, where, p = number of indicator variables (i.e. observed scale variables) and q = number of parameters to be estimated (i.e., factor loadings, error variances, error covariances, etc.).

We also reported the power of the overall test of response shifts, which guard against capitalization on the chance of Type I error, and the power of the tests of the particular response shifts that we found in our study. All power calculations [77-79] were based on a 5% level of significance. This procedure increased our confidence that we found a correct measurement model, and that response shifts if detected were indeed present.

Schmitt's SEM Approach

The first step in the Schmitt's SEM method [28] was to assess change in the entire variance-covariance matrix between time 1 and time 2. If a significant difference was found, then the subsequent steps of the analysis were to evaluate if these differences were because of reconceptualization, configural invariance, which is invariance of the factor structure (the a priori pattern of factor loadings imposed on the scales). If configural invariance did not hold, then the observed scores represents different

constructs assessed from one time to the next, and it would not make sense to test further differences over time. If there was configural invariance, then the subsequent step was to assess if there was scale recalibration. This was evaluated by testing the equivalence of the factor variances and the factor loadings over time as they represent the scale metric or units of measurement. Finally, change in uniqueness (error variances) was tested to evaluate if changes were because of change in measurement error over time. The acceptability of the fitted SEM solution was evaluated on the basis of overall goodness-of-fit indices and validation approach stated for Oort's SEM approach.

Missing Data Evaluation

Useable survey responses were obtained for the baseline survey of study patients. We compared missingness in the patients who responded to the baseline SF-36 surveys with gender, race, age, history of diabetes, angina, abnormal coronary angiogram, myocardial infarction, stroke, coronary bypass, parkinsons, cancer, peripheral vascular disease, left ventricular hypertrophy, congestive heart failure, arrhythmias, alzheimers, transient ischemic attack, renal insufficiency, hypercholesteremia, and smoking. If any of the above said variables differed based on return of baseline surveys, the groups were compared within the two antihypertensive treatment strategies to determine whether randomization took place appropriately.

Validation is the process of determining that a model is good. In our study, we demonstrate an approach to assess sensitivity analysis, in which model fit was assessed for four imputed data sets using MLE. The four imputed data sets were compared to the data set with complete cases (i.e. complete baseline and one year cases) based on presence of response shift and model fit statistics. For the first

imputation approach, we looked at patients who scored on the PF and MH scales at both baseline and one year time points, and imputed the remaining domains using markov chain monte carlo (MCMC) approach.

For the second imputation approach, we chose to impute PF domain scores with instrumental activities of daily living (IADL) complete scores and impute the remaining seven domains at both time points using MCMC approach. For the third approach, we analyzed cases with complete PF scores at both time points and imputed the seven domain scores with complete IADL scores. For the fourth approach, we chose to impute all missing cases using MCMC approach.

Statistical Analysis for Research Question 2

Variables

We distinguished between outcome variables measuring HRQoL, change in systolic BP (SBP) over one year, change in diastolic BP (DBP) over one year, non-fatal stroke within one year of enrollment, non-fatal MI within one year of enrollment, baseline CES-D scores, one year CES-D scores, and self-reported prior history of depression were included as explanatory variables (E). Antihypertensive treatment strategies, baseline comorbid conditions, age, sex, race, depression diagnosis by the physician at one year, living status, and educational level were included as potential violator variables (V). From here on, the SF-36 physical and mental latent constructs are denoted by A and the eight SF-36 domains are denoted by X. All variables were collected at baseline. However, SBP, DBP, CES-D scores and SF-36 scale scores were also collected at one year. Patient's sex, age, race, living status, and educational level were obtained by the survey. Baseline comorbid conditions were measured through

self-report. We described latent constructs (A) for the first research question and thus will not be reiterated here.

Hypothesis#1

- H_0: The relationships between explanatory variables (E) and SF-36 scales (X) will be fully explained by the relationships between explanatory variables and violator variables (V), explanatory variables and latent constructs (A), latent constructs and violator variables, and latent constructs and SF-36 scales, no measurement bias will be found with respect to the explanatory variables.

- H_a: The E → X relationships will not be fully explained by the E ↔ V, E → A, A ↔ V and A → X relationships, measurement bias will be found with respect to variables E.

Hypothesis#2

- H_0: The V → X relationships will be fully explained by the E ↔ V, E → A, A ↔ V and A → X relationships, no measurement bias will be found with respect to variables V.

- H_a: The V → X relationships will not be fully explained by the E ↔ V, E → A, A ↔ V and A → X relationships, measurement bias will be found with respect to variables V.

Hypothesis#3

- H_0: The E → X relationships will be consistent across measurement occasions (i.e. over two time points), no response shift will be observed.

- H_a: The E → X relationships will not be consistent across measurement occasions (i.e. over two time points), response shift will be observed.

Hypothesis#4

- H_0: The V → X relationships will be consistent across measurement occasions (i.e. over two time points), no response shift will be observed.

- H_a: The V → X relationships will not be consistent across measurement occasions (i.e. over two time points), response shift will be observed.

Explanatory variables (E)

- Change in systolic blood pressure and diastolic blood pressure over one year: Change in SBP and change in DBP over one year reported for patients assigned to verapamil-SR and atenolol-led treatment strategies were taken from the INVEST trial dataset.

- Stroke or Myocardial infarction within one year of enrollment: Patients who had stroke or MI within one year of enrollment (i.e. after allocation to treatment strategies) were coded as (1) or (0) if they did not have an occurrence of stroke or MI within the same time period. These two variables will be combined to form a single variable with patients who had a stroke or MI will be coded as (1) or (0) if they did not have the same within one year of enrollment.

- Baseline and one year CES-D scores: Respondents' depressive symptoms were assessed using the CES-D scale [80] included in the baseline mail questionnaire. The CES-D is a 20-item self-reported scale that is considered a reliable (Cronbach's α range 0.83-0.91) and valid instrument. CES-D scores range from 0 to 60, with higher scores indicating more depressive symptoms. Scores > 16 are generally consistent with depressive symptoms of clinically depressed patients. People with major chronic medical conditions are most likely to score in the high depressive symptoms range, so a higher threshold of 23 is recommended for studies of older persons with chronic illnesses [11,17].

- Self-reported prior history of depression: Patients' self-reported prior history of depression was obtained at baseline. Patient's who reported prior history of depression was coded as (1) or (0) if they did not report depression.

Potential violator variables (V)

- Pharmacologic Hypertension Treatment Strategy: INVEST patients were randomly assigned to either the atenolol-led or the verapamil-led treatment strategy [8]. Patients assigned to the atenolol-led strategy were coded as zero (0) and those assigned to the verapamil-SR strategy were coded as one (1).

- Baseline comorbid conditions: History of coronary artery bypass graft, angina, arrhythmia, abnormal coronary angiogram, Parkinson disease, peripheral vascular disease, myocardial infarction, stroke, alzheimers, cancer, left ventricular hypertrophy, hypercholesterolemia, congestive heart failure, transient ischemic attack, renal insufficiency, smoking, and diabetes were noted at baseline (coded as '1' if the condition was present). The original Charlson Index includes a few conditions that are different to the ones present in this study. A modified Charlson Index (which removes the points for the CAD complications of myocardial infarction and heart failure from the original index) (Table 3-2) has been applied in patients with CAD [81,82]. Each of the indicated diagnoses is assigned a weight and summed to provide a patient's total score.

- Age: Patient's age was obtained at the baseline INVEST visit. Patients age was dichotomized as (1) for $\geq$75 years old and (0) for those $\leq$74 years old.

- Gender: Patients who were females were coded zero (0) and those who were males were coded one (1). Data regarding gender were obtained during the baseline INVEST visit.

- Race: Patients who were Caucasians were coded zero (0) and those who were Non-Caucasians were coded one (1). Data regarding race were obtained during the baseline INVEST visit.

- Depression diagnosis by the physician at one year: Patients were asked if, "….a medical doctor or psychiatrist has told you that you were depressed during the past year". The respondents were coded as (1) if they reported yes or (0) if they reported as no or I don't know.

- Education level: Patients who completed a high school graduation were coded as (1) or (0) if they did not complete a high school graduation.

- Living status: Patients who were living alone were coded as (0) or (1) if they were living with someone.

A two-step procedure was used to investigate measurement bias and response shift in the HRQoL data gathered at baseline and at one year period. In each of the two steps, SEM described by Oort et al. [31,32] was applied.

Step 1: Establishing a Measurement Model

SEM was used to fit a confirmatory factor model to the 16 x 16 variance-covariance matrix of the eight HRQoL scales measured at two occasions. We used the MLE method to fit a model with both measurement occasions and two common factors, PHYS HRQoL and MENT HRQoL, with a pattern of factor loadings that was similar to that developed for the first research question. Overall goodness-of-fit was evaluated with the χ^2 test of exact fit and the RMSEA as a measure of approximate fit. A nonsignificant χ^2 test indicated good model fit. An RMSEA value of less than 0.08 suggests satisfactory fit, and a value of less than 0.05 suggests close fit [75]. In addition to overall goodness-of-fit, component fit was evaluated through inspection of modification indices and standardized residuals [73].

Step 2: Detecting Measurement Bias and Response Shift in Measurement

In the second step, we extended the four-factor model to include the following: change in SBP and DBP, stroke and MI within one year of enrollment, baseline CES-D scores, one year CES-D scores, self-reported prior history of depression, depression diagnosis by the physician at one year, antihypertensive treatment strategies, age, sex, race, education level, living status, and baseline comorbid conditions. There were six explanatory variables and eight potential violator variables that were included in the model as exogenous variables, with residual variances fixed at zero. All violator variables were correlated with each other, with the explanatory variables and with the common HRQoL factors.

Measurement bias was indicated by significant modification indices for direct effects of the six explanatory variables and the eight violator variables on the eight SF-36 scales and for across occasion constraints on factor loadings and intercepts [31,32]. In all, there were 264 modification indices to consider ([16 x 6 direct effects fixed at zero] + [16 x 8 direct effects fixed at zero] + [8 x 4 across occasion constraints on factor loadings] + [8 across occasion constraints on intercepts]), however, eight modification indices were subtracted because of impossible effects of the second occasion health state on first occasion SF-36 scales. This left 256 modification indices for constrained parameters that were interpreted as measurement bias if unconstrained. As there were a large number of tests, to maintain a family wise Type I error rate of 5%, a Bonferroni-adjusted critical value [78,79] of 12.9 (associated with a probability of 0.05/256) was used. The Step 2 model was modified by changing one parameter at a time and constantly checking the estimates and results to ensure that the changes were meaningful and interpretable. This process was continued until the largest modification

index was less than 12.9. We should note that the modification index often underestimates the actual improvement of fit (χ^2 difference). Therefore, we also checked for parameter changes that may be associated with modification indices less than 12.9 and investigated the actual χ^2 difference.

Missing Data Evaluation

For research question#2, data was analyzed to determine the type of missingness. Useable survey responses were obtained for the baseline survey of study patients. We compared missingness in the patients who responded to the baseline SF-36 surveys with gender, race, age, history of diabetes, angina, abnormal coronary angiogram, myocardial infarction, stroke, coronary bypass, parkinsons, cancer, peripheral vascular disease, left ventricular hypertrophy, congestive heart failure, arrhythmias, alzheimers, transient ischemic attack, renal insufficiency, hypercholesteremia, and smoking. If any of the above said variables differed based on return of baseline surveys, they were compared within the two antihypertensive treatment strategies to determine whether randomization took place appropriately.

Validation is the process of determining that a model is good. In our study, we demonstrated an approach to assess sensitivity analysis, in which model fit was assessed for four imputed data sets using MLE. The four imputed data sets were compared to the data set with complete cases (i.e. complete baseline and one year surveys) based on presence of response shift and model fit statistics. For the first imputation approach, we looked at patients who scored on the PF and MH scales at both baseline and one year time points, and imputed the remaining domains and variables using MCMC approach.

For the second imputation approach, we chose to impute PF domain scores with IADL complete scores and impute the remaining seven domains at both time points and other variables using MCMC approach. For the third approach, we chose cases with complete PF scores at both time points and imputed the seven domain scores with complete IADL scores and remaining variables using MCMC. For the fourth approach, we chose to impute all cases at both time points using MCMC approach.

Table 3-1. SF-36 HRQoL measure

Measure	Short-Form 36
Type	General health profile
Purpose	Evaluative
Domains	Eight Physical functioning, role limitations due to physical, bodily pain, vitality, social functioning, role limitations due to emotional, mental health, general health
Number of items	36
How administered	Interview – in person or phone
Scaling	2 points to 6 points depending on item
Scoring	Domains 0-100 Higher = Better Physical and mental health components 0-100 with mean = 50, SD = 10 Algorithm required
Used in response shift studies	Yes

Table 3-2. Weights assigned to different comorbid states in the modified Charlson Index

Points	Condition
0	Coronary artery bypass graft, angina, myocardial infarction, congestive heart failure
1	Peripheral vascular disease, arrhythmia, abnormal coronary angiogram, alzhemier, gastrointestinal bleed, hypercholesteremia, left ventricular hypertrophy, Parkinson, transient ischemic attack
2	Stroke, cancer, renal insufficiency, diabetes

Table 3-3. Analytical approach

Task	Measure	Software
Characterizing study participants	Descriptive statistics, cross-sectional at two points in time, missing data excluded	SAS 9.14
Characterizing variables	Univariate tests for normality of the subscales for HRQoL indices	SAS 9.14
Missing data evaluation	Chi-square comparison of specific outcomes and missingness to determine missing data pattern	SAS 9.14
Model building	Based on Medical Outcomes Study Framework of Health Indicators and incorporating modification indices	LISREL 8.8
Identification of response shift	For research question#1 SEM approach described by Oort [10] to identify reconceptualization, recalibration, or reprioritization response shift SEM approach described by Schmitt [28] For research question#2 SEM approach described by Oort [31] and King-Kallimanis [32]	LISREL 8.8 LISREL 8.8
Model estimation	Variance-covariance data analyzed	LISREL 8.8
Model evaluation	Overall goodness-of-fit χ^2 (df, p-value) Multiple fit indices (SRMR, RMSEA, CFI, NFI, NNFI) Localized areas of ill fit (modification indices, standardized residuals)	LISREL 8.8

CHAPTER 4
RESULTS

Schmitt and Oort's SEM Approaches to Detect Response Shift (Research Question 1)

Description of the Population

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. Complete survey responses from both the baseline and one year follow-up surveys were obtained from 39.23% (n = 909) of the study's initial sample of patients (Figure 4-1). Nearly 56.5% (n = 514) of the study's final sample of patients was male and the majority were Caucasian (80.7%).

Comparison of Treatment Groups at Baseline

At baseline, patients were randomized to treatment group. The final sample of patients in the two treatment groups was similar on their sociodemographic characteristics and other variables (Table 4-1) indicating success of the randomization procedures.

Measurement Model

Below we first explain the measurement model that was used in the response shift identification procedure. The same measurement model was used for both the Oort and Schmitt SEM procedures to detect response shift. We present the results of response shift detection and true change evaluation conducted with the Oort procedure and we conclude with an evaluation of the size of response shifts and true change. Lastly, we present the results of response shift detection with the Schmitt procedure.

Table 4-3 gives baseline and one year means scores and standard deviations for all SF-36 scales. The last column of Table 4-3 presents the standardized difference

between baseline and one year mean scores (i.e., Cohen's d-index) that is the change in the observed scale mean scores without accounting for potential response shifts. Conventional t-tests indicate improvement in RP, BP, GH, VT, MH and RE, and no change in PF and SF scale mean scores.

Results from exploratory factor analyses gave rise to the measurement model displayed in Figure 4-2. Two latent variables are the latent factors PHYS HRQoL and MENT HRQoL. PHYS HRQoL is measured by PF, RP, BP, GH, VT, and SF, MENT HRQoL is measured by MH, RE, and again SF. Other latent variables are the residual factors ResPF, ResRP, ResBP, etc. The residual factors represent all that is specific to individual scales (i.e., PF, RP, BP, etc) in each latent factor, plus random error variation.

The measurement model portrayed in Figure 4-2 resembles the principal components model of the SF-36 scales described by Ware et al. [74]. The general physical and mental components in this study are similar to Ware's model, with largely the same indicators. Similar to the Oort et al. study, we found that the wording of the SF items combines physical and mental aspects, causing SF to load on both PHYS HRQoL and MENT HRQoL. We allowed the residual factors for RP and RE to co-vary (0.22 correlation) since the covariance between RP and RE was not adequately explained by the correlation between the two latent factors.

Oort's SEM Approach to Detect Response Shift

Detection of Response Shift and True Change

Fit results for the four models that resulted from carrying out the four-step procedure are given in Table 4-4 and Table 4-5.

Step 1: Measurement Model

The measurement model of Figure 4-2 was the basis of Model 1, a SEM for measurements at two occasions, but without any across occasion constraints. The χ^2 test of exact fit was significant (χ^2 (84) = 301.284), but the RMSEA measure indicated reasonable fit (RMSEA = 0.054 (90% CI = 0.048, 0.061), Table 4-4). The CFI was 0.991; a value of >.90 suggest a reasonably good fit. Other fit indices, such as the NFI, NNFI, and SRMR had values 0.99, 0.99, and 0.03, respectively; all indicating good fit (i.e., the data fit the proposed model well).

Step 2: No Response Shift Model

In Model 2, all response shift parameters were held invariant across occasions. This means that all across occasion invariance constraints on parameter estimates (factor loadings, intercepts, and residual variances) were imposed. Two models were compared, the model in which parameter estimates were all freely estimated (i.e., Model 1), and the fully constrained model (i.e., Model 2) in which all of the parameter estimates were made to be equal across time. The χ^2 difference test evaluates the change in overall fit of the free and fully constrained (no response shift) models. The fit of Model 2, although still satisfactory (RMSEA = 0.051, Table 4-4), was significantly worse than the fit of Model 1, indicating the presence of response shifts (χ^2 difference test: χ^2 (25) = 64.128, $p < 0.0001$). Therefore, we reject the null hypothesis that the two models fit equally well and conclude that there may be evidence of response shift present in the data.

Step 3: Identification of Response Shift

Inspection of modification indices and standardized residuals indicated which of the equality constraints were not tenable. Step by step adjustment of Model 2 yielded

Model 3, which showed three cases of response shift. The fit of Model 3 was good (RMSEA = 0.049 (90% CI = 0.043 – 0.055)), and significantly better than the fit of Model 2 (χ^2 (3) = 34.950), Table 4-4.

The next step was to address the possibility of making a Type I error with multiple testing; familywise Type I error rate. Using a significance level of $\alpha = 0.05/3 = 0.0167$, we identified two cases of response shift. The fit of the new Model 3 was good (RMSEA = 0.049 (90% CI = 0.043 – 0.055)), and significantly better than the fit of Model 2 (χ^2 (2) = 30.892), Table 4-5.

Step 4: Final Model

To investigate change in the means, variances, and correlations of the common factors, we fitted additional models with the new Model 3 as the starting point. The across occasion invariance of parameters was tested step-by-step, maintaining all equality constraints that proved tenable. This procedure finally yielded Model 4, which fitted the data closely (χ^2 (111) = 340.681, RMSEA = 0.048 (90% CI = 0.042 – 0.054)), Table 4-5.

Evaluation of Response Shifts and True Change

Recalibration Response Shift

Based on Oort's definition, intercepts and residual variances contain information about uniform and non-uniform recalibration, respectively. For PF, we found differences between first and second occasion intercepts, indicating uniform recalibration for the PF scale. According to Oort, uniform recalibration is reflected by change in the means of the observed variables that cannot be attributed to change in the common factor means.

We also found a change in the variance of the residual factor PF, indicating non-uniform recalibration for PF. Based on Oort's definition, non-uniform recalibration is

reflected by the change in the variances of the observed variables that cannot be attributed to change in the common factor variances.

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

In addition to significance test results, Table 4-7 provides effect-sizes for observed change, and the response shift and true change contributions to observed change, as implied by the parameter estimates of Model 4 (in Table 4-5 and Table 4-6). From Table 4-7 it appears that the response shift effects on observed change were only small: -0.118 for the uniform recalibration for PF scale and zero for the non-uniform recalibration for PF scale. The effects of true change were smaller for PF. For PF the effects of response shifts and true change were in opposite directions.

Impact of Response Shifts on the Measurement of True Change

We found that most Model 4 parameters were invariant across occasions, except for the latent means that did change over time (Table 4-6). Common factor variances and common factor correlations did not change across occasions, but the common factor means did. Common factor means were fixed at zero for the first occasion (because of identification requirements) so that the second occasion estimates were a direct representation of change.

Schmitt's SEM Approach to Detect Response Shift

The Schmitt approach started with the same baseline longitudinal measurement model as the Oort method. As the longitudinal model fit well, there was no reconceptualization, configural variance over time, which is invariance of factor structure (the a priori pattern of factor loadings imposed on the scales). Since configural invariance was absent, the subsequent step was to assess if there was scale

recalibration. This was evaluated by testing the equivalence of the factor variances and the factor loadings (metric invariance) over time as they represent the scale metric or units of measurement. When constraints were applied and added one at a time, variances and factor loadings (recalibration), we found one significant change between subsequent models (Table 4-8). The fit of Model 4, was satisfactory (RMSEA = 0.050 (90% CI, 0.044 – 0.057), Table 4-8). However, we found that freeing the factor loading for PF significantly improved the fit of the Model (χ^2 difference test: χ^2 (1) = 3.906, p = 0.048). We also found an improvement to the model when PF error variances were constrained (χ^2 difference test: χ^2 (1) = 8.446, p = 0.004), which according to the Schmitt's definition, represents a change in random error over time. Based on Schmitt's definition, a significant decrease in the variance accounted for by the models signals a difference with respect to the reliability of measurement before and after assignment to treatment strategies.

Missing Data Evaluation and Imputation

Useable survey responses were obtained for the baseline survey from 68.1% (n = 1578) of study patients. Gender (χ^2 = 7.06, p = 0.008), race (χ^2 = 69.19, p<0.0001), history of angina (χ^2 = 10.10, p = .001), history of abnormal coronary angiogram (χ^2 = 11.04, p = 0.001), history of coronary bypass (χ^2 = 7.81, p=0.005), history of cancer (χ^2 = 4.50, p = 0.03) history of left ventricular hypertrophy (χ^2 = 5.64, p = 0.02), history of smoking (χ^2 = 5.65, p = 0.02) were the least likely to respond to the baseline survey. Within gender, race, history of angina, history of abnormal coronary angiogram, history of coronary bypass, history of cancer, history of left ventricular hypertrophy, and history of smoking, baseline return rates were similar for the two antihypertensive treatment strategies. Figure 4-1 details reasons for not completing the baseline survey.

In this study, we chose missing data imputation using sensitivity analysis with four different approaches; model fit and response shift was assessed across the four imputed datasets using MLE. All longitudinal models using Oort and Schmitt's SEM approaches had a reasonable fit and showed presence of response shift (Table 17-Table 24).

Oort SEM procedure. When using the first imputation approach (Table 17), we found presence of uniform recalibration for BP in addition to the presence of uniform recalibration and non-uniform recalibration for PF. It seems that for some patients (using the first imputation approach) the meaning of the response scale anchors for the BP scale changed since assignment to treatment strategy. Similar to our first imputation approach, we found presence of uniform recalibration for BP using the second imputation approach (Table 18). However, we did not identify non-uniform recalibration for PF compared to our original results. When using the third imputation approach (Table 19), we found reprioritization and non-uniform recalibration for VT in addition to the presence of uniform and non-uniform recalibration for PF. On the other hand, when using the fourth imputation approach, we found several instances of response shift as shown in Table 20.

Schmitt's SEM procedure. When using the first imputation approach (Table 21), we found random error for BP and VT in addition to the presence of recalibration for PF and random error for PF. Comparatively, when using the second imputation approach (Table 22); we found that the variances of the physical function construct changed over time indicating presence of recalibration. We also found evidence of random error for PF, BP, and VT. Its seems that using imputation approach 2, some respondents are

perceiving more or less difference in the relevant constructs due to the disease condition or treatment strategy assigned.

When using the third imputation approach (Table 23), we found random error for VT in addition to the response shift obtained for our original analyses. On the other hand, when using the fourth imputation approach (Table 24), we found several instances of response shift. It seems that each imputation approach led to slightly different results compared to our original analyses. Since there is no best single imputation technique available, these results should be interpreted with caution. This deserves further investigation.

Measurement Bias and Response Shift (Research Question 2)

Description of the population

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. Complete survey responses from both the baseline and one year follow-up surveys were obtained from 34.01% (n = 788) of the study's initial sample of patients (Figure 4-3). Nearly 57% (n = 451) of the study's final sample of patients was male and the majority were Caucasian (82.9%).

Comparison of Treatment Groups at Baseline

At baseline, patients were randomized to treatment group. The final sample of patients in the two treatment groups was similar on their sociodemographic characteristics and other variables (Table 4-10) indicating success of the randomization procedures.

Below we first explain the measurement model that was used in the response shift identification procedure. We present the results of response shift detection and true

change evaluation conducted with the Oort procedure and we conclude with an evaluation of the size of response shifts and true change.

Step 1: Measurement Model

The measurement model for this research question was analogous to the first research question, Figure 4-2. Briefly, PHYS HRQoL was measured by PF, RP, BP, GH, VT, and SF, MENT HRQoL was measured by MH, RE, and again SF. Table 4-11 gives baseline and one year means scores and standard deviations for all SF-36 scales. The last column of Table 4-11 presents the standardized difference between baseline and one year mean scores (i.e., Cohen's d-index) that is the change in the observed scale mean scores without accounting for potential response shifts. Conventional t-tests indicate improvement in RP, BP, GH, VT, MH and RE, and no change in PF and SF scale mean scores.

In Model 1, factor loadings and intercepts were not constrained to be equal across occasions. Identical to the measurement model in the first research question, the χ^2 test of exact fit was significant (χ^2 (84) = 287.554), but the RMSEA measure indicated reasonable fit (RMSEA = 0.055 (90% CI = 0.048, 0.062), Table 4-12). We concluded that this model was satisfactory and applied this model for the investigation of measurement bias and response shift in measurement.

Step 2: Measurement Bias and Response Shift

In the second step, all factor loadings and intercepts were constrained to be equal across occasions. Explanatory variables were included – occurrence of non-fatal stroke or non-fatal MI, baseline CES-D scores, one year CES-D scores, change in SBP, change in DBP, and prior history of depression with direct effects on the latent factors, PHYS HRQoL and MENT HRQoL. Potential violators of measurement invariance were

included as exogenous variables – age, gender, race, treatment strategy, education, living alone, depression diagnosis by physician at one year, and modified charlson's comorbidity score. The exogenous variables were allowed to correlate with the explanatory variables and with the common HRQoL factors, but not directly affect the observed HRQoL subscale scores.

Two models were compared, the model in which parameter estimates were all freely estimated (i.e. Model 1), and the model specified in Step 2. The χ^2 test of exact fit was significant (χ^2 (273) = 744.881) and the RMSEA measure indicated reasonable fit (RMSEA = 0.045 (90% CI = 0.041, 0.049), Table 4-12). The fit of Model 2, although still satisfactory, was significantly worse than the fit of Model 1, indicating the presence of response shifts (χ^2 difference test: χ^2 (189) = 457.327, $p < 0.0001$). Modification indices revealed that the fit of the model could be further improved by accounting for instances of measurement bias. We found one intercept that was not equal across measurement occasions, which indicated recalibration response shift. For PF, we found differences between first and second occasion intercepts, indicating uniform recalibration for the PF scale.

The relationship between the potential violator variables and the observed variables should be explained via their relationships with the common HRQoL factors. In addition, the relationships between the explanatory variables and the observed variables should be explained via their relationships with the common HRQoL factors. When this does not occur, measurement bias has been found. The results indicated four instances of measurement bias which could be considered as response shift in measurement. For the model that identified four instances of measurement bias, age

was used as a continuous variable. However, when age was dichotomized ($\leq$ 74 years and $\geq$ 75 years), we identified only two instances of measurement bias/response shift. Below we elaborate and provide explanations on both these approaches. Subsequently, we demonstrate which of these two approaches provide a better explanation of the presence of measurement bias and/or response shift in our study population.

For the first approach when age was used as a continuous variable, we found the relationship between age and PF was not fully determined by their relationship with the PHYS HRQoL latent factor. This indicated that PF was not just indicative of PHYS HRQoL but also of age. Therefore, a direct relationship between age and PF was included. The violation of measurement invariance was not consistent across occasions (estimated at -0.177 at baseline and -0.199 at one year) which indicated that older patients reported worse PF than younger patients, even if their PHYS HRQoL was similar. We also found that the relationship between age and RP was not fully determined by their relationships with the PHYS HRQoL factor. This indicated that RP was not just indicative of PHYS HRQoL but also of age. Therefore, a direct relationship between age and RP was included. The violation of measurement invariance was not consistent across occasions (estimated at -0.140 at baseline and -0.119 at one year) and indicated that older patients reported worse RP than younger patients, even if their PHYS HRQoL was similar. Lastly, a direct relationship between gender and PF was included. The violation of measurement invariance was not consistent across occasions (estimated at 0.100 at baseline and 0.131 at one year) which indicated that male patients reported better PF than female patients, even if their PHYS HRQoL was similar. After accounting for the three response shifts stated above, the modified model

showed improvement and close fit (RMSEA = 0.039, see Model 2, Table 4-12). The largest modification indices were well below the critical value of 12.9.

Contrary to our first approach, when age was dichotomized, we did not find any relationship between age and PF or age and RP. On the other hand, similar to the first approach, we found a relationship between gender and PF that was not fully determined by their relationships with the common PHYS HRQoL factor. This indicated that PF was not just indicative of PHYS HRQoL but also of gender. Therefore, a direct relationship between gender and PF was included. The violation of measurement invariance was not consistent across occasions (estimated at 0.113 at baseline and 0.150 at one year, Table 4-14) which indicated that male patients reported better PF than female patients, even if their PHYS HRQoL was similar. After accounting for response shifts, the modified model showed improvement and close fit (χ^2 = 638.562, RMSEA = 0.041 (90% CI = 0.036 – 0.045), Table 4-13). The largest modification indices were well below the critical value of 12.9. Based on the findings in our study, the second approach resulted in another model, albeit more parsimonious. Similar to the approach adopted by King-Kallimanis [32], here, we chose the model (second approach when age was dichotomized) with measurement bias and response shift in the smallest number of scales, providing a more parsimonious model.

True Change

Common factor means were fixed at zero for the first occasion (because of identification requirements) so that the second occasion estimates were a direct representation of change. In model 2, the estimates of the common factor means indicated that after assignment to treatment strategies patients reported worsened PHYS HRQoL and improved MENT HRQoL.

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

In addition to significance test results, Table 4-16 provides effect-sizes for observed change, and the response shift and true change contributions to observed change, as implied by the parameter estimates of Model 2 shown in Table 4-13. From Table 4-16 it appears that the response shift effects on observed change were only small: -0.097 for the uniform recalibration for PF scale. The effects of true change were smaller for PF. For PF the effects of response shifts and true change were in opposite directions.

Missing Data Evaluation and Imputation

Useable survey responses were obtained for the baseline survey from 68.1% (n = 1578) of study patients. Gender ($\chi^2 = 7.06$, $p = 0.008$), race ($\chi^2 = 69.19$, $p<0.0001$), history of angina ($\chi^2 = 10.10$, $p=.001$), history of abnormal coronary angiogram ($\chi^2 = 11.04$, $p = 0.001$), history of coronary bypass ($\chi^2 = 7.81$, $p = 0.005$), history of cancer ($\chi^2 = 4.50$, $p = 0.03$) history of left ventricular hypertrophy ($\chi^2 = 5.64$, $p = 0.02$), history of smoking ($\chi^2 = 5.65$, $p = 0.02$) were the least likely to respond to the baseline survey. Within gender, race, history of angina, history of abnormal coronary angiogram, history of coronary bypass, history of cancer, history of left ventricular hypertrophy, and history of smoking, baseline return rates were similar for the two antihypertensive treatment strategies. Figure 4-3 details reasons for not completing the baseline survey.

In this study, we chose missing data imputation using sensitivity analysis with four different approaches; model fit and response shift was assessed across the four imputed datasets using MLE. All longitudinal models using Oort SEM approach had a reasonable fit and showed presence of response shift (Table 25 – Table 28).

When using the first imputation approach (Table 25), we found presence of uniform recalibration for BP in addition to the presence of uniform recalibration and non-uniform recalibration for PF. It seems that for some patients (using the first imputation approach) the meaning of the response scale anchors for the BP scale changed since assignment to treatment strategy. We also found that high school education had a direct effect on BP at the second occasion. The effect of high school education on BP at the second measurement occasion was negative suggesting that at one year patients with high school education reported worse BP than those with no high school education, even if their true HRQoL was similar. Similar to the first imputation approach, we found high school education to have a direct effect on BP using imputation approach 2 (Table 26). However, we found no effect of gender on PF, but, found age to have a direct effect on PF at the second occasion. Imputation approach 3 (Table 27) gave us similar results to those using imputation approach 2. Using imputation approach 4 (Table 28), we found uniform recalibration for PF and BP scales, gender to have a direct effect on PF, and education to have a direct effect on baseline PF, among several others.

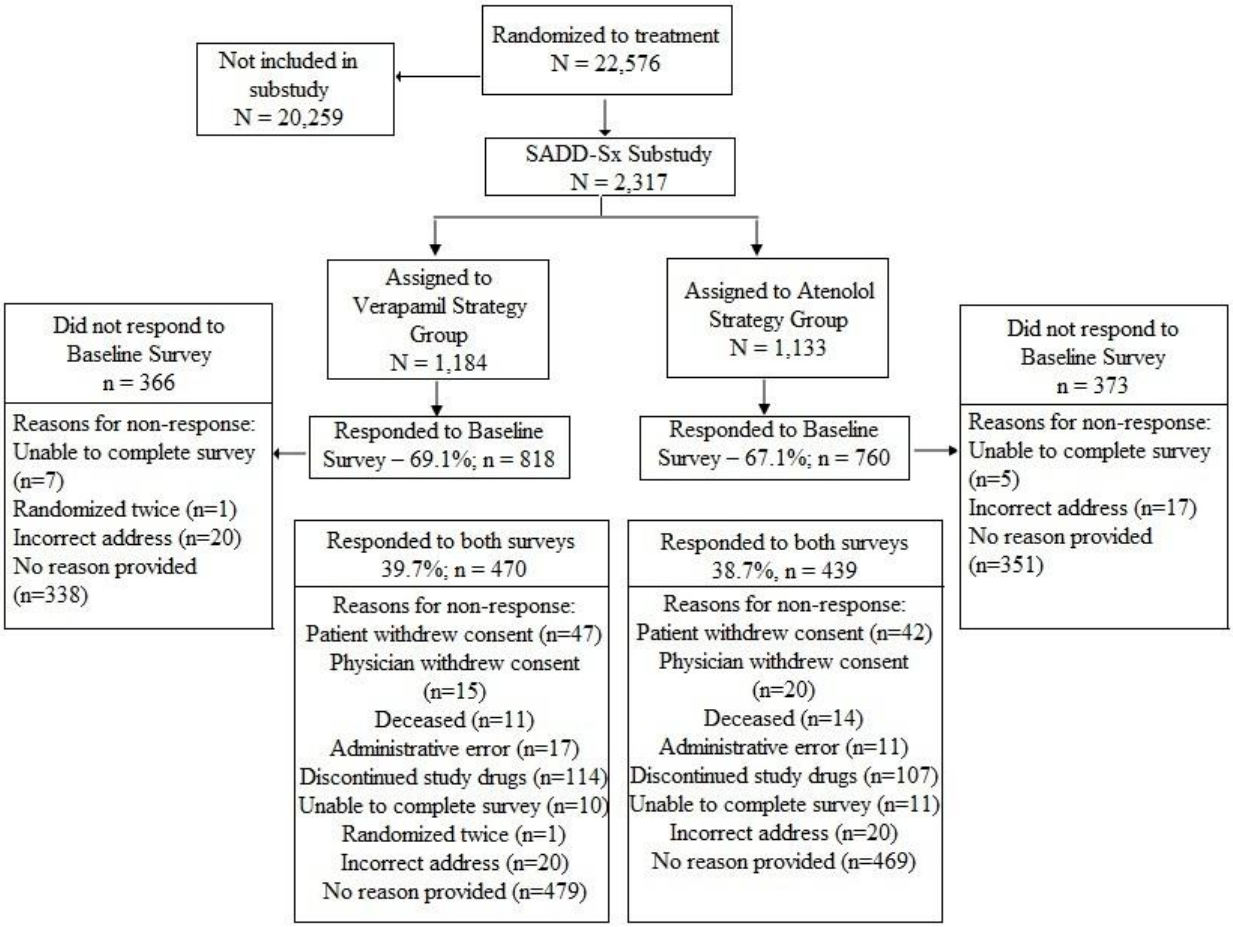

Figure 4-1. Flow diagram of patients' progression through the SADD-Sx substudy of the International Verapamil SR-Trandolapril Study (n = 2317).
Abbreviation: SADD-Sx – Study of Antihypertensive Drugs and Depressive Symptoms

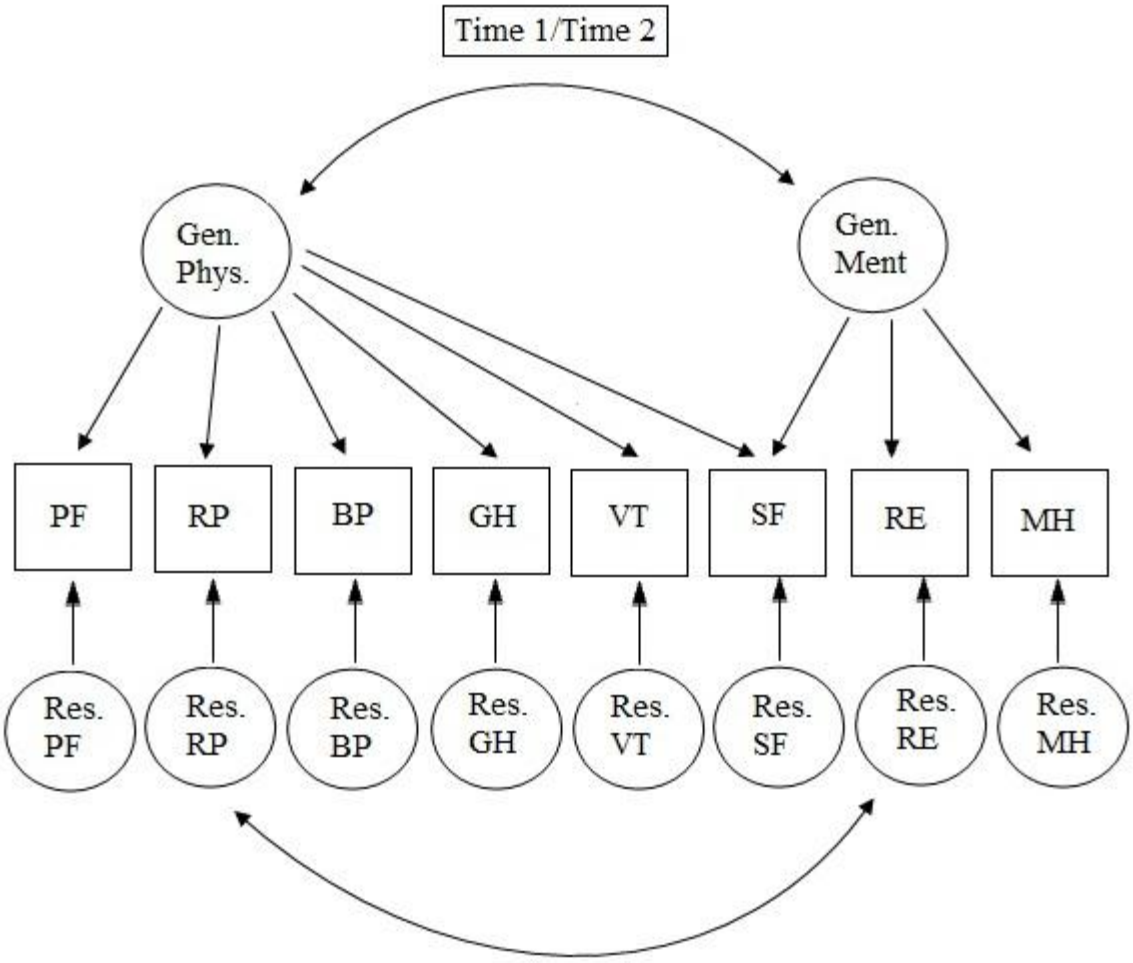

Figure 4-2. The measurement model used in response shift detection

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.

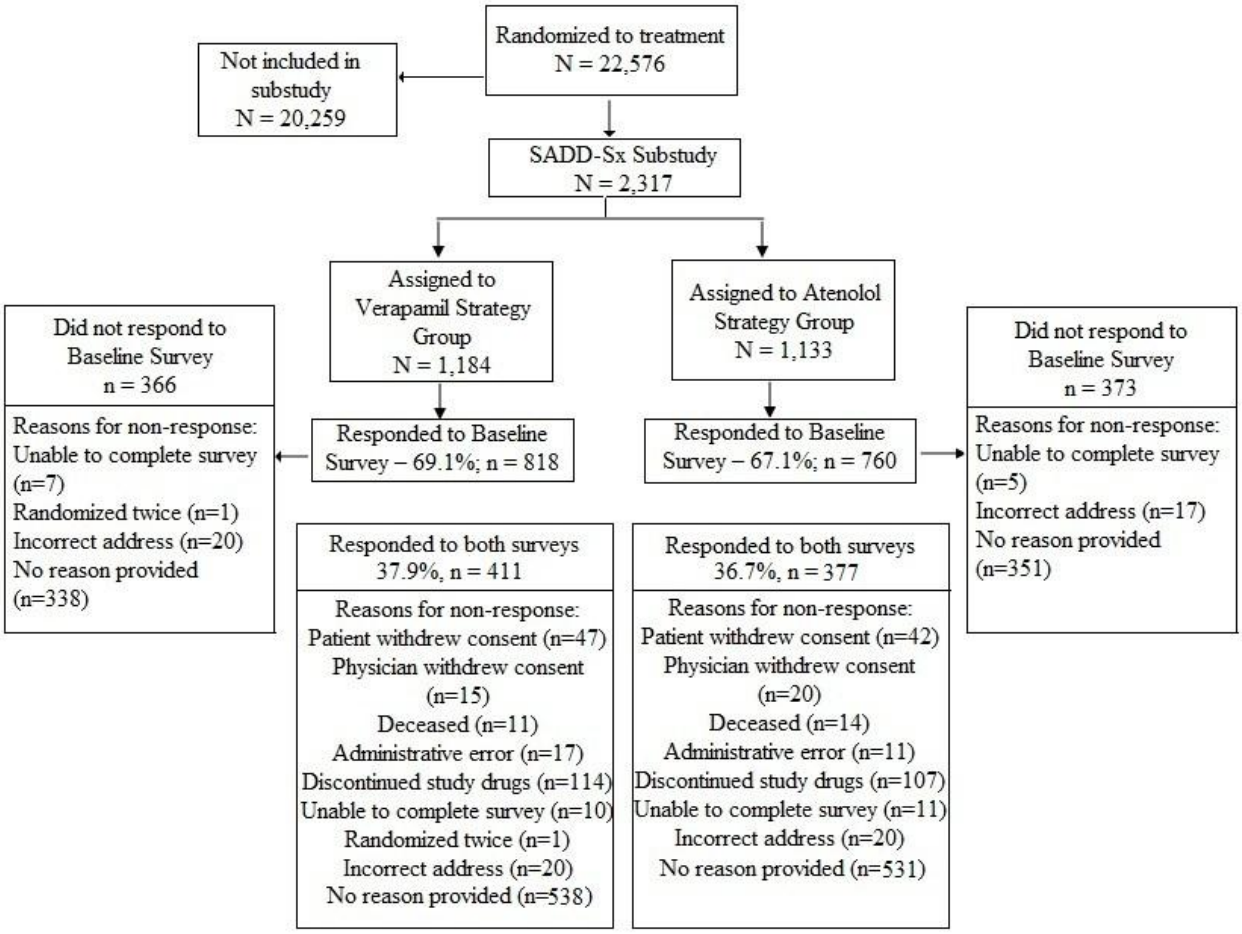

Figure 4-3. Flow diagram of patients' progression through the SADD-Sx substudy of the International Verapamil SR-Trandolapril Study (n = 2317).
SADD-Sx = Study of Antihypertensive Drugs and Depressive Symptoms.

Table 4-1. Comparison of patients at baseline assigned to the verapamil-SR strategy and the atenolol strategy among the SADD-Sx patients returning both baseline and one year surveys (N = 909)

Variable	Atenolol Strategy (N = 429)	Verapamil- SR strategy (N = 465)	p value
Demographic characteristics			
Male	61.3%	55.7%	0.09
Caucasian	82.5%	82.8%	0.91
Patient average age at entry – years (SD)	67.4 (9.0)	67.5 (9.4)	0.91[a]
Medical History at Baseline			
Coronary Artery Bypass Graft (CABG)	30.3%	27.7%	0.23
Congestive Heart Failure (CHF)	7.7%	5.6%	0.21
Angina	45.7%	45.0%	0.82
Peripheral vascular disease	13.5%	14.8%	0.57
Arrhythmia	7.5%	6.9%	0.74
Abnormal coronary angiogram	65.5%	65.2%	0.92
Myocardial Infarction (MI)	42.7%	46.5%	0.25
Stroke	5.1%	6.5%	0.40
Alzheimer[b]	0.5%	0.4%	0.38
Cancer	5.6%	4.3%	0.37
Hypercholesteremia	68.1%	65.6%	0.43
Left Ventricular Hypertrophy	19.1%	17.4%	0.51
Diabetes	29.6%	25.8%	0.21
Parkinson[b]	0.0%	0.2%	0.52
Transient Ischemic Attack	4.9%	4.3%	0.67
Renal Insufficiency	3.0%	3.7%	0.60
Smoking	55.7%	54.8%	0.79

SADD-Sx = Study of Antihypertensive Drugs and Depressive Symptoms; SD = Standard Deviation
a Independent t test
b Fisher's Exact test

Table 4-2. Testing for normality in HRQoL indices

Variable	Normality
Subscale scores	
SF-36 physical health, role physical, role emotional, social function, mental health, pain, vitality, general health	Moderately non-normal to normally distributed

Moderately non-normal = skewness <2, kurtosis<7, severely non-normal = skewness>2, kurtosis>7

Table 4-3. 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-Pred-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 – bodily pain; GH – general health; VT – vitality; SF – social functioning; RE – role-emotional; MH – mental health.

Table 4-4. Oort method: goodness of fit and χ^2 differences between models without controlling for Type I error (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	Uniform recalibration (PF)	108	341.168	0.049 (0.043-0.055)	0.037	0.991	0.986	0.989
	Uniform recalibration (BP)	107	337.118	0.049 (0.043 – 0.055)	0.037	0.991	0.986	0.989
	Non-uniform recalibration (PF)	106	330.462	0.049 (0.043 - 0.055)	0.037	0.991	0.987	0.989
Model 4	Final Model (all tenable constraints imposed)	110	336.630	0.048 (0.042 – 0.054)	0.040	0.991	0.986	0.989

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, BP – bodily pain.
RMSEA <= 0.05 close fit; 0.05 to 0.08 reasonable fit; >= 0.10 poor fit
CFI > 0.90 reasonably good fit, SRMR < 0.10 is favorable

Table 4-5. Oort method: goodness of fit and χ^2 differences between models when controlling for Type I error (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	Uniform recalibration (PF)	108	341.168	0.049 (0.043-0.055)	0.037	0.991	0.986	0.989
	Non-uniform recalibration (PF)	107	334.520	0.049 (0.043 - 0.055)	0.038	0.991	0.986	0.989
Model 4	Final Model (all tenable constraints imposed)	111	340.681	0.048 (0.042 – 0.054)	0.040	0.991	0.986	0.989

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 > 0.90 reasonably good fit, SRMR < 0.10 is favorable

Table 4-6. Parameter estimates in the final model (Model 4, Table 4-5) (n = 909)

	Before assignment to treatment strategy			After assignment to treatment strategy	
	PHYS HRQoL1	MENT HRQoL1		PHYS HRQoL2	MENT HRQoL2
Factor loadings (Γ)					
PF1	0.742		PF2	0.769	
RP1	0.748		RP2	0.748	
BP1	0.661		BP2	0.661	
GH1	0.775		GH2	0.775	
VT1	0.761		VT2	0.761	
SF1	0.348	0.450	SF2	0.348	0.450
RE1		0.722	RE2		0.722
MH1		0.766	MH2		0.766
Intercepts (τ)					
	55.142		51.709		
	47.660		47.660		
	62.973		62.973		
	52.081		52.081		
	46.276		46.276		
	69.991		69.991		
	63.187		63.187		
	72.115		72.115		
Residual variances					
	0.449		0.409		
	0.440		0.440		
	0.563		0.563		
	0.399		0.399		
	0.421		0.421		
	0.428		0.428		
	0.478		0.478		
	0.414		0.414		

	PHYS HRQoL1	MENT HRQoL1	PHYS HRQoL2	MENT HRQoL2
Common factor correlations				
PHYS HRQoL1	1			
MENT HRQoL1	0.795	1		
PHYS HRQoL2	0.795	0.643	1	
PHYS HRQoL2	0.643	0.795	0.795	1
Common factor means				
	0.000	0.000	0.131	0.104

Goodness of overall fit measures: χ^2 (111) = 340.681, RMSEA = 0.048 (0.042 – 0.054), SRMR -= 0.040, CFI = 0.991, NFI = 0.986, NNFI = 0.989.

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 4-7. Significance tests of response shifts, and effect sizes of observed change, response shift, and true change in the final model (From tables 4-5 and 4-6)

Scale	Response shift	Significance test		Effect-sizes*		
		χ^2 (df =1)	p value#	Observed change	Response shift contribution	True change contribution
PF	Uniform recalibration	24.244	p < 0.0001	-0.021	-0.118	0.097
	Non-uniform recalibration	6.648	p = 0.01		0.000**	
RP				0.082		0.082
BP				0.138		0.138
GH				0.105		0.105
VT				0.090		0.090
SF				0.055		0.055
RE				0.091		0.091
MH				0.082		0.082

#α significance level p < 0.0167

*effect size has been described as small (0.2), medium (0.5), and large (0.8).

**non-uniform recalibration was identified, but non-uniform recalibration is not estimated in the effect size calculation

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 4-8. Schmitt 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	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	0.989

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 <= 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.

Table 4-9. 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 across measurements 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. 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. 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 4-10. Comparison of patients at baseline assigned to the Verapamil-SR strategy and Atenolol-led treatment strategy among SADD-Sx patients returning both baseline and one year surveys (n = 788)

Variable	Atenolol strategy	Verapamil-SR strategy	P value
Stroke or MI within one year of enrollment	1.6%	1.9%	0.706
Baseline CES-D score[b,c]	13.837	13.338	0.523
Baseline SBP[b,c]	147.988	146.608	0.282
Baseline DBP[b,c]	82.765	82.450	0.681
Self reported prior history of depression[b]	18.0%	17.5%	0.849
Age (≥ 75 years)[a]	21.8%	20.4%	0.652
Male[a]	59.4%	55.2%	0.236
Non-Caucasian[a]	17.5%	16.8%	0.789
High school graduate[b]	69.5%	71.5%	0.531
Living status alone or someone[b]	75.9%	75.9%	0.987
Modified Charlson's comorbidity index[c]	2.584	2.628	0.719

SD = standard deviation
aInformation regarding these variables was obtained from INVEST data.
bInformation regarding these variables was obtained from the SADD-Sx baseline mail survey from those patients who returned both the baseline and one-year follow-up surveys.
cIndependent t-test.

Table 4-11. Means, standard deviations, and effect sizes for SF-36 scales (n = 788)

Scale	Before assignment to treatment strategy		After assignment to treatment strategy		Post-Pre d-Index[#]
	Mean	Standard deviation	Mean	Standard deviation	
PF	55.557	29.260	55.065	29.284	-0.017
RP	49.228	40.780	51.840	41.250	0.064
BP	63.047	29.648	66.284	31.195	0.109**
GH	52.834	22.041	55.051	22.143	0.101**
VT	46.937	22.736	48.617	22.052	0.074*
SF	70.963	24.063	72.303	23.071	0.056
RE	64.150	40.786	67.936	39.895	0.093**
MH	72.475	18.744	74.132	18.920	0.088**

N = 788; # 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 – bodily pain; GH – general health; VT – vitality; SF – social functioning; RE – role-emotional; MH – mental health.

Table 4-12. Goodness of fit of models in measurement bias and response shift detection procedure (when age was used as a continuous variable) (n = 788)

Step	Model	Df	CHISQ	RMSEA (90% CI)	SRMR	CFI	NFI	NNFI
Step 1:	Model 1: Measured Model	84	287.554	0.055 (0.048 – 0.062)	0.035	0.991	0.987	0.987
Step 2: Detect measurement bias	Model 2: First model, without accounting for measurement bias	273	744.881	0.045 (0.041 – 0.049)	0.038	0.986	0.978	0.978
	Uniform recalibration for PF	272	729.136	0.045 (0.040 – 0.049)	0.037	0.987	0.979	0.979
	Age – TPF	271	703.021	0.043 (0.047 – 0.048)	0.037	0.987	0.980	0.980
	Age – BPF	270	669.555	0.042 (0.038 – 0.046)	0.035	0.988	0.981	0.981
	Age – TRP	269	658.643	0.042 (0.037 – 0.046)	0.035	0.989	0.981	0.981
	Age – BRP	268	631.316	0.040 (0.036 – 0.045)	0.033	0.989	0.982	0.983
	Gender – TPF	267	614.376	0.040 (0.036 – 0.044)	0.033	0.990	0.982	0.983
	Gender – BPF (Final model)	266	600.530	0.039 (0.035 – 0.043)	0.032	0.990	0.983	0.984

Abbreviations: PF, physical functioning; RP, role limitations because of physical health; BPF, baseline physical functioning; TPF, one year physical functioning; BRP, baseline role limitations because of physical health; TRP, one year role limitations because of physical health.
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 <= 0.05 close fit; 0.05 to 0.08 reasonable fit; >= 0.10 poor fit
CFI > 0.90 reasonably good fit, SRMR < 0.10 is favorable

Table 4-13. Goodness of fit of models in measurement bias and response shift detection procedure (when age was used as a dichotomous variable) (n = 788)

Step	Model	Df	CHISQ	RMSEA (90% CI)	SRMR	CFI	NFI	NNFI
Step 1:	Model 1: Measured Model	84	287.554	0.055 (0.048 – 0.062)	0.035	0.991	0.987	0.987
Step 2: Detect measure-ment bias	Model 2: First model, without accounting for measurement bias	273	692.494	0.043 (0.039 – 0.047)	0.036	0.988	0.980	0.980
	Uniform recalibration for PF	272	676.847	0.042 (0.038 – 0.046)	0.035	0.988	0.980	0.981
	Gender – TPF	271	657.167	0.042 (0.037 – 0.046)	0.035	0.989	0.981	0.982
	Gender – BPF (Final model)	270	638.562	0.041 (0.036 – 0.045)	0.034	0.989	0.981	0.982

Abbreviations: PF, physical functioning; RP, role limitations because of physical health; BPF, baseline physical functioning; TPF, one year physical functioning.
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 <= 0.05 close fit; 0.05 to 0.08 reasonable fit; >= 0.10 poor fit
CFI > 0.90 reasonably good fit, SRMR < 0.10 is favorable.

Table 4-14. Parameter estimates in the final model (Model 2, Table 4-13) (n = 788)

	Before assignment to treatment strategy			After assignment to treatment strategy	
	PHYS HRQoL1	MENT HRQoL1		PHYS HRQoL2	MENT HRQoL2
Factor loadings (Γ)					
PF1	0.734		PF2	0.752	
RP1	0.727		RP2	0.742	
BP1	0.664		BP2	0.646	
GH1	0.781		GH2	0.787	
VT1	0.749		VT2	0.778	
SF1	0.404	0.372	SF2	0.424	0.383
RE1		0.668	RE2		0.681
MH1		0.838	MH2		0.823
[a]Intercepts (τ)					
	55.557			52.711	
	48.895			48.895	
	63.520			63.520	
	53.004			53.004	
	46.819			46.819	
	70.561			70.561	
	64.542			64.542	
	72.457			72.457	

Common factor variances				
	PHYS HRQoL1	MENT HRQoL1	PHYS HRQoL2	MENT HRQoL2
	1.000	1.000	1.000	1.000
Common factor correlations				
PHYS HRQoL1	1			
MENT HRQoL1	0.762	1		
PHYS HRQoL2	0.797	0.609	1	
MENT HRQoL2	0.563	0.738	0.750	1

Abbreviations: PF, physical functioning; RP, role limitations because of physical health; BP, bodily pain; SF, social functioning; MH, mental health; RE, role limitations because of emotional problems; VT, vitality; GH, general health.

Notes: (a) across time differences in bold between factor loadings indicate (recalibration) response shift.

Table 4-15. Direct effects indicating measurement bias and response shift in measurement (based on Table 4-13)

Measurement bias	Baseline	One year
Gender – PF	0.113	0.150

Abbreviations: PF, physical functioning.
Notes: across time differences in measurement bias indicate response shift in measurement.

Table 4-16. Significance tests of response shifts, effect sizes of observed change, response shift, and true change in the final model (Table 4-13 and Table 4-14)

Scale	Response shift	Significance test χ^2 (df =1)	p value[#]	Effect-sizes* Observed change	Response shift contribution	True change contribution
PF	Uniform recalibration	15.647	p < 0.0001	-0.017	-0.097	0.080
RP				0.064		0.064
BP				0.109		0.109
GH				0.101		0.101
VT				0.074		0.074
SF				0.056		0.056
RE				0.093		0.093
MH				0.088		0.088

#α significance level p < 0.05
*effect size has been described as small (0.2), medium (0.5), and large (0.8).
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 4-17. Oort method: Imputation using approach 1 (PF and MH present and other domain scores imputed using MCMC) (n=988)

Model		Df	CHISQ	RMSEA	SRMR	CFI	NFI	NNFI
Model 1		84	322.618	0.054 (0.048 – 0.060)	0.034	0.991	0.988	0.987
Model 2		109	406.214	0.053 (0.047-0.058)	0.038	0.989	0.985	0.988
Model 3	Uniform recalibration (PF)	108	373.719	0.050 (0.045-0.056)	0.038	0.990	0.986	0.989
	Uniform recalibration (BP)	107	366.927	0.050 (0.044 – 0.055)	0.037	0.990	0.986	0.989
	Non-uniform recalibration (PF)	106	358.522	0.050 (0.044 - 0.055)	0.038	0.990	0.986	0.989
Model 4	Final Model (all tenable constraints imposed)	110	366.436	0.049 (0.044 – 0.054)	0.040	0.990	0.986	0.989

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, BP – bodily pain.
RMSEA <= 0.05 close fit; 0.05 to 0.08 reasonable fit; >= 0.10 poor fit
CFI > 0.90 reasonably good fit, SRMR < 0.10 is favorable

Table 4-18. Oort method: Imputation using approach 2 (PF imputed with IADL scores and remaining domain scores imputed using MCMC) (n=1057)

Model		Df	CHISQ	RMSEA	SRMR	CFI	NFI	NNFI
Model 1		84	354.338	0.056 (0.050 – 0.062)	0.034	0.990	0.987	0.986
Model 2		109	421.830	0.052 (0.047-0.058)	0.038	0.989	0.985	0.988
Model 3	Uniform recalibration (PF)	108	397.402	0.051 (0.045-0.056)	0.037	0.990	0.986	0.988
	Uniform recalibration (BP)	107	391.856	0.051 (0.045 – 0.056)	0.037	0.990	0.986	0.989
Model 4	Final Model (all tenable constraints imposed)	111	399.377	0.050 (0.045 – 0.055)	0.040	0.990	0.986	0.989

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, BP – bodily pain.
RMSEA <= 0.05 close fit; 0.05 to 0.08 reasonable fit; >= 0.10 poor fit
CFI > 0.90 reasonably good fit, SRMR < 0.10 is favorable

Table 4-19. Oort method: Imputation using approach 3 (PF imputed with MCMC approach and remaining domain scores with IADL scores) (n=1037)

Model		Df	CHISQ	RMSEA	SRMR	CFI	NFI	NNFI
Model 1		84	346.451	0.055 (0.049 – 0.061)	0.034	0.990	0.987	0.986
Model 2		109	420.019	0.052 (0.047-0.057)	0.038	0.988	0.984	0.987
Model 3	Reprioritization (VT)	108	413.882	0.052 (0.046-0.057)	0.038	0.988	0.984	0.987
	Uniform recalibration (PF)	107	392.839	0.050 (0.045 – 0.056)	0.037	0.989	0.985	0.988
	Non-uniform recalibration (PF)	106	385.458	0.050 (0.045 – 0.055)	0.037	0.989	0.985	0.988
	Non-uniform recalibration (VT)	105	374.760	0.049 (0.044 – 0.055)	0.037	0.990	0.986	0.988
Model 4	Final Model (all tenable constraints imposed)	109	380.925	0.049 (0.043 – 0.054)	0.039	0.990	0.986	0.989

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, BP – bodily pain.
RMSEA <= 0.05 close fit; 0.05 to 0.08 reasonable fit; >= 0.10 poor fit
CFI > 0.90 reasonably good fit, SRMR < 0.10 is favorable

Table 4-20. Oort method: Imputation using approach 4 (N = 2317)

Model		Df	CHISQ	RMSEA (90% CI)	SRMR	CFI	NFI	NNFI
Model 1		84	775.190	0.069 (0.064 – 0.073)	0.037	0.988	0.986	0.983
Model 2		109	1574.624	0.076 (0.073 – 0.080)	0.051	0.980	0.979	0.978
Model 3	Reprioritization (VT)	108	1545.216	0.076 (0.073 – 0.079)	0.048	0.980	0.979	0.978
	Reprioritization (SF)	107	1532.692	0.076 (0.073 – 0.079)	0.046	0.981	0.979	0.978
	Uniform recalibration (PF)	106	1380.884	0.072 (0.069 – 0.076)	0.044	0.983	0.981	0.980
	Uniform recalibration (BP)	105	1365.455	0.072 (0.069 – 0.076)	0.044	0.983	0.981	0.980
	Non-uniform recalibration (PF)	104	1290.516	0.071 (0.067 – 0.074)	0.043	0.984	0.982	0.981
	Non-uniform recalibration (RP)	102	1217.155	0.069 (0.065 – 0.072)	0.043	0.985	0.983	0.982
	Non-uniform recalibration (GH)	102	1190.308	0.068 (0.065 – 0.072)	0.042	0.985	0.984	0.983
	Non-uniform recalibration (VT)	101	1136.287	0.067 (0.064 – 0.071)	0.042	0.986	0.985	0.983
	Non-uniform recalibration (SF)	100	1087.854	0.066 (0.063 – 0.070)	0.041	0.987	0.985	0.984
	Non-uniform recalibration (RE)	99	1061.802	0.066 (0.063 – 0.069)	0.040	0.987	0.986	0.984
	Non-uniform recalibration (MH)	98	1027.187	0.065 (0.061 – 0.068)	0.038	0.987	0.986	0.984
Model 4	Final Model (all tenable constraints imposed)	102	1095.932	0.066 (0.062 – 0.069)	0.045	0.986	0.985	0.984

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, BP – bodily pain.
RMSEA <= 0.05 close fit; 0.05 to 0.08 reasonable fit; >= 0.10 poor fit
CFI > 0.90 reasonably good fit, SRMR < 0.10 is favorable

Table 4-21. Schmitt method: Imputation using approach 1 (PF and MH present and other domain scores imputed using MCMC) (n=988)

Model		Df	CHISQ	RMSEA	SRMR	CFI	NFI	NNFI
Model 1		84	322.618	0.054 (0.048 – 0.060)	0.034	0.991	0.988	0.987
Model 2	Covariances constrained (reconceptualization)	87	326.875	0.053 (0.047 – 0.059)	0.038	0.991	0.988	0.987
Model 3	Factor covariances and variances constrained (recalibration)	91	333.806	0.052 (0.046-0.058)	0.040	0.991	0.987	0.988
Model 4	Factor covariances, variances, and loadings constrained (recalibration)	98	345.920	0.051 (0.045 – 0.057)	0.041	0.991	0.987	0.988
	PF (factor loading)	97	339.732	0.051 (0.045 – 0.057)	0.040	0.991	0.987	0.989
Model 5	Factor covariances, variances, loadings, and uniqueness constrained	107	370.565	0.050 (0.045 – 0.056)	0.041	0.990	0.986	0.989
	PF	106	359.650	0.050 (0.044 – 0.055)	0.041	0.990	0.986	0.989
	BP	105	354.438	0.049 (0.044 – 0.055)	0.040	0.991	0.987	0.989
	VT	104	349.062	0.049 (0.044 – 0.055)	0.040	0.991	0.987	0.989

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

Table 4-22. Schmitt method: Imputation using approach 2 (PF imputed with IADL scores and remaining domain scores imputed using MCMC) (n=1057)

Model		Df	CHISQ	RMSEA	SRMR	CFI	NFI	NNFI
Model 1		84	354.338	0.056 (0.050 – 0.062)	0.034	0.990	0.987	0.986
Model 2	Covariances constrained (reconceptualization)	87	358.203	0.055 (0.049 – 0.060)	0.037	0.990	0.987	0.987
Model 3	Factor covariances and variances constrained (recalibration)	91	365.847	0.054 (0.048-0.060)	0.039	0.990	0.987	0.987
	Factor variance	90	361.756	0.054 (0.048 – 0.060)	0.038	0.990	0.987	0.987
Model 4	Factor covariances, variances, and loadings constrained (recalibration)	97	370.365	0.052 (0.047 – 0.058)	0.039	0.990	0.987	0.988
Model 5	Factor covariances, variances, loadings, and uniqueness constrained	107	393.537	0.051 (0.045 – 0.056)	0.040	0.990	0.986	0.988
	PF	106	389.523	0.051 (0.045 – 0.056)	0.040	0.990	0.986	0.988
	BP	105	384.445	0.050 (0.045 – 0.056)	0.039	0.990	0.986	0.989
	VT	104	378.731	0.050 (0.045 – 0.056)	0.039	0.990	0.986	0.989

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

Table 4-23. Schmitt method: Imputation using approach 3 (PF imputed with MCMC approach and remaining domain scores with IADL scores) (n=1037)

Model		Df	CHISQ	RMSEA	SRMR	CFI	NFI	NNFI
Model 1		84	346.451	0.055 (0.049 – 0.061)	0.034	0.990	0.987	0.986
Model 2	Covariances constrained (reconceptualization)	87	351.215	0.054 (0.048 – 0.060)	0.038	0.990	0.987	0.986
Model 3	Factor covariances and variances constrained (recalibration)	91	355.871	0.053 (0.047-0.059)	0.039	0.990	0.987	0.987
Model 4	Factor covariances, variances, and loadings constrained (recalibration)	98	368.402	0.051 (0.046 – 0.057)	0.041	0.990	0.986	0.987
	PF	97	361.927	0.051 (0.046 – 0.057)	0.040	0.990	0.986	0.988
Model 5	Factor covariances, variances, loadings, and uniqueness constrained	107	396.937	0.051 (0.045 – 0.056)	0.041	0.989	0.985	0.988
	PF	106	387.352	0.050 (0.045 – 0.056)	0.041	0.989	0.985	0.988
	VT	105	374.368	0.050 (0.044 – 0.055)	0.040	0.990	0.986	0.988

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

Table 4-24. Schmitt method: Imputation using approach 4 (N = 2317)

Model		Df	CHISQ	RMSEA	SRMR	CFI	NFI	NNFI
Model 1		84	775.190	0.069 (0.064 – 0.073)	0.037	0.988	0.986	0.983
Model 2	Covariances constrained (reconceptualization)	87	978.013	0.067 (0.064 – 0.071)	0.038	0.988	0.987	0.983
Model 3	Factor covariances and variances constrained (recalibration)	91	1051.605	0.068 (0.065 – 0.072)	0.045	0.987	0.986	0.983
		90	1041.211	0.069 (0.065 – 0.072)	0.043	0.987	0.986	0.983
		89	1015.815	0.068 (0.065 – 0.072)	0.041	0.987	0.986	0.983
		88	984.569	0.067 (0.064 – 0.071)	0.039	0.988	0.987	0.983
Model 4	Factor covariances, variances, and loadings constrained (recalibration)	95	1021.200	0.066 (0.062 – 0.070)	0.041	0.987	0.986	0.984
	PF	94	1010.544	0.066 (0.062 – 0.069)	0.041	0.988	0.986	0.984
	VT	93	989.275	0.065 (0.062 – 0.069)	0.039	0.988	0.987	0.984
Model 5	Factor covariances, variances, loadings, and uniqueness constrained	103	1347.050	0.072 (0.069 – 0.076)	0.043	0.983	0.982	0.980
	PF	102	1264.030	0.070 (0.067 – 0.074)	0.042	0.984	0.983	0.981
	RP	101	1196.048	0.069 (0.065 – 0.049)	0.042	0.985	0.984	0.982
	GH	100	1171.043	0.068 (0.065 – 0.072)	0.042	0.985	0.984	0.982
	VT	99	1118.205	0.067 (0.064 – 0.071)	0.042	0.986	0.985	0.983
	SF	98	1068.300	0.066 (0.063 – 0.070)	0.040	0.987	0.985	0.984
	RE	97	1048.951	0.066 (0.062 – 0.070)	0.040	0.987	0.986	0.984
	MH	96	1020.859	0.065 (0.062 – 0.069)	0.040	0.987	0.986	0.984

Table 4-25. Measurement bias and response shift detection (when age was used as a dichotomous variable) using imputation approach 1 (n = 988)

Step	Model	Df	CHISQ	RMSEA (90% CI)	SRMR	CFI	NFI	NNFI
Step 1:	Model 1: Measured Model	84	322.125	0.054 (0.048 – 0.060)	0.034	0.991	0.988	0.987
Step 2: Detect measurement bias	Model 2: First model, without accounting for measurement bias	273	765.708	0.046 (0.042 – 0.050)	0.040	0.982	0.974	0.973
	Uniform recalibration for PF	272	732.423	0.045 (0.041 – 0.048)	0.040	0.984	0.975	0.975
	Uniform recalibration for BP	271	725.631	0.044 (0.041 – 0.048)	0.040	0.984	0.976	0.975
	Gender – TPF	270	689.581	0.043 (0.039 – 0.047)	0.039	0.985	0.977	0.977
	Gender – BPF	269	665.857	0.042 (0.038 – 0.046)	0.039	0.986	0.978	0.978
	Educ – BPF	268	650.763	0.041 (0.037 – 0.045)	0.039	0.986	0.978	0.978

Table 4-26. Measurement bias and response shift detection (when age was used as a dichotomous variable) using imputation approach 2 (n = 1057)

Step	Model	Df	CHISQ	RMSEA (90% CI)	SRMR	CFI	NFI	NNFI
Step 1:	Model 1: Measured Model	84	360.044	0.056 (0.050 – 0.062)	0.034	0.990	0.987	0.986
Step 2: Detect measurement bias	Model 2: First model, without accounting for measurement bias	273	827.930	0.047 (0.043 – 0.051)	0.041	0.981	0.974	0.972
	Uniform recalibration for PF	272	802.372	0.046 (0.043 – 0.050)	0.041	0.982	0.975	0.973
	Uniform recalibration for BP	271	795.967	0.046 (0.043 – 0.050)	0.041	0.982	0.975	0.973
	Age – TPF	270	749.505	0.044 (0.040 – 0.048)	0.041	0.984	0.976	0.975
	Educ – BPF	269	732.859	0.043 (0.040 – 0.047)	0.041	0.984	0.977	0.976

Table 4-27. Measurement bias and response shift detection (when age was used as a dichotomous variable) using imputation approach 3 (n = 1037)

Step	Model	Df	CHISQ	RMSEA (90% CI)	SRMR	CFI	NFI	NNFI
Step 1:	Model 1: Measured Model	84	346.451	0.055 (0.049 – 0.061)	0.034	0.990	0.987	0.986
Step 2: Detect measurement bias	Model 2: First model, without accounting for measurement bias	273	941.096	0.052 (0.048 – 0.055)	0.041	0.975	0.966	0.962
	Uniform recalibration for PF	272	894.588	0.050 (0.047 – 0.054)	0.041	0.976	0.968	0.964
	Uniform recalibration for BP	271	875.707	0.050 (0.046 – 0.053)	0.041	0.977	0.969	0.965
	Age – TPF	270	831.231	0.048 (0.044 – 0.051)	0.041	0.979	0.970	0.967
	Educ – BPF	269	816.748	0.047 (0.044 – 0.051)	0.041	0.979	0.971	0.968

Table 4-28. Measurement bias and response shift detection (when age was used as a dichotomous variable) using imputation approach 4 (n = 2317)

Step	Model	Df	CHISQ	RMSEA (90% CI)	SRMR	CFI	NFI	NNFI
Step 1:	Model 1: Measured Model	84	775.190	0.069 (0.064 – 0.073)	0.037	0.988	0.986	0.983
Step 2: Detect measurement bias	Model 2: First model, without accounting for measurement bias	273	2459.390	0.062 (0.059 – 0.064)	0.046	0.973	0.970	0.959
	Reprioritization for PF	272	2446.359	0.062 (0.059 – 0.064)	0.045	0.973	0.970	0.959
	Reprioritization for RP	271	2439.532	0.062 (0.059 – 0.064)	0.045	0.973	0.970	0.959
	Reprioritization for BP	270	2434.171	0.062 (0.060 – 0.064)	0.045	0.973	0.970	0.959
	Reprioritization for GH	269	2420.597	0.062 (0.059 – 0.064)	0.045	0.973	0.971	0.959
	Uniform recalibration for PF	268	2323.380	0.060 (0.058 – 0.063)	0.045	0.975	0.972	0.961
	Uniform recalibration for BP	267	2306.358	0.060 (0.058 – 0.063)	0.045	0.975	0.972	0.961
	Gender – TPF	266	2221.249	0.059 (0.057 – 0.062)	0.044	0.976	0.973	0.962
	Gender – BPF	265	2174.330	0.059 (0.057 – 0.061)	0.044	0.976	0.974	0.963
	Age – TPF	264	2077.683	0.057 (0.055 – 0.060)	0.043	0.978	0.975	0.965
	Age – BPF	263	2045.344	0.057 (0.054 – 0.059)	0.043	0.978	0.975	0.965
	Educ – BPF	262	1989.029	0.056 (0.054 – 0.058)	0.043	0.979	0.976	0.966
	Gender – TRP	261	1953.894	0.056 (0.053 – 0.058)	0.042	0.979	0.976	0.967
	Age – BRP	260	1921.734	0.055 (0.053 – 0.058)	0.042	0.979	0.977	0.967
	Age – TRP	259	1884.471	0.055 (0.052 – 0.057)	0.041	0.980	0.977	0.967

CHAPTER 5
DISCUSSION

The Oort procedure showed evidence of uniform recalibration. Apparently, the meaning of the response scale anchors for the PF scale changed since assignment to treatment strategy. The Oort approach also showed evidence of non-uniform recalibration for PF scale. Based on Oort's definition, possibly the meaning of some of the anchors for PF response scale has changed in a group of hypertensive CAD patients. The Oort approach also identified uniform recalibration for the BP scale. However, using a Bonferroni correction $\alpha = 0.05/3 = 0.0167$, only uniform and non-uniform recalibration for the PF scale was identified. All response shifts occurred within PHYS HRQoL construct.

We found that most Model 4 parameters were invariant across occasions, except for the latent means that did change over time (Table 4-6). Common factor variances and common factor correlations did not change across occasions, but the common factor means did. We thus did not find any evidence of reconceptualization of a possible higher order factor representing HRQoL, or of reprioritization of its components. Common factor means were fixed at zero for the first occasion (because of identification requirements) so that the second occasion estimates were a direct representation of change. In addition to significance test results, Table 4-7 provides effect-sizes for observed change, and the response shift and true change contributions to observed change, as implied by the parameter estimates of Model 4 (in Table 4-5). From Table 4-7 it appears that the response shift effects on observed change were only small: -0.118 for the uniform recalibration for PF scale. The effects of true change were smaller for PF. For PF the effects of response shifts and true change were in opposite directions. In

the present data set, we observe a change in the respondent's internal standards which had a marginal impact on the estimation of true change in the means of the physical factor. Although minimal, we did identify response shift among these patients which warrant the assessment of response shift.

Difference in Results between Oort and Schmitt Procedures in this Study

We now elaborate on alternate interpretations for the disparate results among the Oort and Schmitt procedures in our study. With Schmitt procedure, constraints were added to the model to identify a response shift. Contrary to the Schmitt procedure, constraints were removed to identify a response shift with the Oort procedure. The constraints added or removed are equality constraints, where a parameter estimate at one time is made to be equal to another time. All types of response shift (reconceptualization, reprioritization, and uniform and non-uniform recalibration) can be detected using the Oort procedure. On the contrary, the Schmitt procedure does not identify reprioritization (since the covariance analysis used is incapable of identifying changes in definition from change in values) or non-uniform recalibration and it interprets changes in factor loadings differently from the approach described by Oort.

Comparable to the Oort approach identifying a response shift in this study, the Schmitt procedure was also supportive of a response shift. Both the Oort and Schmitt SEM procedures showed evidence of a recalibration for PF scale. Based on our results, the differences between the Oort and Schmitt approaches in our study are because of variation in the method and not the sample used. Method differences may result from divergence in defining the type of response shift linked to changes in various parameters of the SEM measurement model. Although both the statistical approaches seem apt to identify response shift, researchers are recommended to clearly

demonstrate the approach used and limitations for the said approaches. Bonferroni correction is strongly recommended when using the Oort procedure to detect response shift. If the order of testing does not matter, we recommend use of both sequences – (1) release constraints on residuals, followed by intercepts, and then factor loadings; and (2) release constraints on factor loadings, intercepts, and then residuals.

The Oort procedure operationalizes various forms of response shift differently relative to the Schmitt's approach. Therefore, although we found significant changes using the Schmitt approach, our conclusions about the type of a response shift are different between the two approaches. Oort defines uniform recalibration as a change in means of the observed variables over time compared to Schmitt who defines recalibration as a change in factor loadings or factor variances over time. According to Oort, respondents change their interpretation of the response scale options (values, labels, or both). If this change affects all response options in the same direction and to the same extent, then the recalibration is explained as uniform. One explanation for this change may be that new experiences after the assignment of treatment strategies may have changed the idea of how much does health limit those performing moderate or vigorous activities, lifting objects, walking or climbing stairs. An alternate explanation may be if patients have learned how to cope with their illness, so that their functioning improved more, or deteriorated less, than would be expected on account of their physical health.

In addition to the identification of uniform recalibration, non-uniform recalibration for the PF scale was also identified using the Oort procedure. Based on Oort's definition, if only some points of a response scale are associated with change in the

same direction and to the same extent, then the shift may be non-uniform. Conversely, the Schmitt technique showed evidence of a change in random error over time for the PF scale. According to Schmitt, the test for recalibration concerns the equality of scaling units in the factors (when the measures are congeneric, i.e., have the same factor loadings). In other words, absence of recalibration indicates that the elements of factor loadings are equal across measurements.

Difference between the Present Study and Ahmed et al. Study

Ahmed et al. [35] is the only study to compare Oort and Schmitt's SEM procedures in the same sample. Converse to the results of our study, Ahmed et al. [35] found that the Oort procedure detected a response shift after a chronic obstructive pulmonary disease self-management program not detected by the Schmitt procedure. The authors report the Oort procedure to be more sensitive in detecting a response shift compared to the Schmitt procedure. We elaborate and provide alternate explanations on the disparity in the results between our study and Ahmed's study [35]. First, our study applied the Bonferroni correction to avoid making a Type I error when identifying a response shift with the Oort procedure. The Ahmed et al. [35] study did not control for Type I error even though the Oort procedure [10] has been suggested to increase the chances of Type I error. As a result, their study may have wrongly identified a response shift when differences may be attributed to other types of changes.

Second, similar to the approach taken by Nolte and Osborne [70], Ahmed et al. [35] chose to release constraints on residuals, followed by intercepts, and then factor loadings when using the SEM procedure. Contrary to Ahmed's approach [35], we conducted the Oort SEM approach as recommended by several other authors. As reflected in the Oort study and the order tested by the authors, and also recognized by

Donaldson [38] and Barclay-Goddard [39], we chose to test sequentially for invariance of factor loadings, intercepts, and error variance. Ahmed et al. [35] may have found divergent results if they would have tested for the presence of response shift by releasing parameter constraints on factor loadings, followed by intercepts, and then residuals.

Ahmed et al. [35] advocate that the order in which the constraints are released make a difference as to whether or not subsequent parameters would be significant. However, it has been questioned that if the sequence of testing does not alter the end result, order of testing may not matter. Therefore, we tested for presence of response shift using a similar approach to Ahmed et al. [35] and Nolte and Osborne [70], where we chose to release constraints on residuals, followed by intercepts, and then factor loadings. Using either of the two sequence of testing did not alter our results; order of testing did not matter in this study.

Based on our results, the differences between our study and Ahmed et al. study are because of variation in the method and the sample used. Method differences resulted due to use of Bonferroni and order of testing for response shift as suggested by previous authors. Sample differences may have resulted due to a larger proportion of the study sample experiencing a response shift compared with individuals with COPD, making response shift among hypertensive CAD patients more detectable.

Validating the Presence of Response Shift in Hypertensive CAD Patients

One explanation of the finding of response shift for the SF-36 PF scale may be demonstrated as a consequence of the disease condition. Hypertension and CAD are chronic health conditions that impact all aspects of function, perception, cognition, mood, QOL [11-13]. All of the negative components of health (impairments, activity

limitations, and participation restrictions) are grouped under the term disability. Under the ICF model [63], impairment, such as decreased strength or coordination of the limbs, causes limitations in activities, such as walking, which eventually leads to participation restrictions. The aspects of functioning that are impacted by ischemic heart disease (which commonly refers to CAD) [63] have been agreed upon by a consensus process conducted by the ICF Research Branch of the WHO Family of International Classifications Collaborating Center. Examples of the content areas included in the ICF Core Set that are likely to have a direct impact on HRQoL include: reduced walking distances, shortness of breath, energy and drive functions, lifting and carrying objects, moving around, and doing housework. The SF-36 PF consists of items describing walking distances, vigorous activities, moderate activities, lifting and carrying objects, climbing stairs, among others. We suggest that change in interpretation may have affected the response scale options (values, labels, or both) for the PF scale in the same direction and to the same extent for a group of patients due to aspects of functioning impacted by the disease condition.

We also found presence of non-uniform recalibration for the PF scale. We recommend that evaluation of specific items in the PF scale may explain why non-uniform response shift occurred in PF. WHO ICF recommends aspects that describe difficulty in PF include impairments (such as decreased strength) or activity limitations (difficulty walking, lifting and carrying objects). Analogous to the aspects that describe difficulty in PF, we believe that items describing vigorous activities such as running, lifting heavy objects; lifting or carrying groceries, and walking more than a mile or several blocks may explain why non-uniform response shift occurred for PF. The items

describing aspect of ‘how much does health limit in various activities’ could potentially undergo change in interpretation over time. Perhaps new experiences after the assignment to treatment strategy have changed their idea of how much does health limit these patients in their physical activities. Presence of response shift may be explained due to mechanisms used to accept or adapt to changes in health status and nature of impairments and activity limitations that many individuals with hypertension and CAD are likely to experience.

Some studies have shown lower HRQoL in individuals with CAD compared to those without CAD. Mitchell et al. [14] conducted a study among elderly subjects with differing cardiovascular status (cardiovascular normal, being hypertensive, having isolated CAD, or both being hypertensive and having CAD). The NHP questionnaire was administered to these participants. The NHP questionnaire incorporates questions related to activity limitations and impairments similar to the classification provided by the WHO’s ICF [63]. Impairments such as decreased strength cause limitations in activities, such as walking, which eventually may lead to participation restrictions, i.e. restricted ability to carry out usual activities in the community, such as through work or volunteering. Mitchell et al. [14] observed lower mean scores on the physical ability dimension in hypertensive CAD patients. In a separate study, Meyer-Rosberg et al. [83] have shown that the physical ability scale of NHP is substantially correlated (-0.79) to the SF-36 PF scale. Vascular disease such as CAD mainly influences physical health [84]. Sevinc and Akyol [85] have found that QOL scores were lowest in PF domain for CAD patients compared to emotional and social domain scores. By analogy, several

studies report that hypertensive CAD patients feel limited in the physical functioning domain.

An alternate explanation for the presence of response shift may be due to the assignment of antihypertensive treatment strategies. Antihypertensive treatment focuses on individual attributes/domains of HRQoL which are potentially important to the hypertensive CAD patients. Bass et al. [86] conducted a double-blind prospective randomized trial on patients awaiting coronary artery bypass graft surgery assigned to atenolol or placebo treatment. Atenolol was reported to show adverse somatic side effects such as feelings of pressure and tightness in the head or body, dizziness and faintness, fatigue, and muscular weakness. In a separate study, Ried et al. [17] report that patients assigned to atenolol-led treatment strategy are more likely to have decreased energy and increased fatigue. These aspects are analogous to the content areas included in the ICF Core Set that describe difficulty in physical functioning including impairments (such as decreased strength) or activity limitations (difficulty walking, lifting and carrying objects). These aspects may partly explain why uniform and non-uniform response shift occurred for PF. The items describing - ‘health limit in various activities’ could potentially undergo changes in interpretation over time. Response shift is thus important to consider in treatment evaluations, especially insofar as it may serve to attenuate or to exaggerate estimates of treatment effects as patients adapt to treatment side effects or disease progression over time.

In addition to the alternative explanations stated above, several studies suggest a relationship between gender and PF and age and PF among CAD patients. De Graff et al. [84] found that gender had a significant impact on the SF-36 PF domain among CAD

patients. The authors also report a small but significant inverse relation between age and SF-36 PF domain. Sevinc and Akyol [85] have found better QOL scores among male CAD patients compared to female CAD patients. The authors also found that the scores for the SF-36 PF domain for CAD patients in the 70-79 year age group were significantly lower than those of the other age groups. In a separate study, Norris et al. [87] report, when compared to men, women with CAD had lower scores in all QOL domains. Few studies [88,89] report of a stereotype implying that men are more physically oriented and invested to do heavier physical labor or strenuous activities. The authors of these studies allege that if PF (e.g. walking long distances, heavy lifting) forms an integral part of the average male's self-image, then it would not be surprising to observe men underreporting their factual physical limitations (i.e. higher PF scores) to preserve their self-image [88,89]. In other words, men seem to be over-reporting their PF to maintain their perceptions of their masculinity and to protect their egos by reducing gender role conflict. The identification of response shift may also be due to different age groups or gender roles which operate differently for males and females [88,89].

Identification of Response Shift for the PF Domain

None of the previous studies have found response shift for the SF-36 PF scale. Oort et al. [36] suggest that SF-36 PF items are less prone to recalibration because the PF items could be answered objectively. In Oort's study [36], after allowing for possible recalibration for BP items, most of their significant modification indices disappeared. In our study, we checked for alternative options in specification searches, but no other selection provided a better model fit. We thus chose to free the intercept constraint for PF scale, indicating evidence of recalibration.

Barclay-Goddard et al. [39] identified the presence of response shift post-stroke in PF construct over time. Based on their results, Barclay-Goddard et al. [39] suggest that 'changes in scores of measurement scales over time that are not considered to be true change may be interpreted as response shift in evaluation-based (how difficult is it to…) methods, while it may be measurement error in performance based (such as a timed walking test) measures'. They recommend that the evaluation of difficulty in their study partly explained why response shift occurred in the model of PF construct. They, however, acknowledge that it was not the only explanation; two of the four measured variables for PF construct in which response shift was suggested contained the evaluation of difficulty. We propose that presence of response shift for the SF-36 PF scale in our study population may be explained due to a single or a combination of the factors including the disease condition, treatment strategies assigned, presence of depressive symptoms, age, and/or gender [39]. This concept deserves further evaluation.

Barclay-Goddard et al. [39] point out in their framework article that SEM requires validation of the model. We thus chose to conduct a sensitivity analysis comparing the complete cases data set (n=909) with the four different imputed data sets as an example of validation. Model fit and presence of response shift was compared across the imputed datasets using MLE. When comparing the results from complete cases and those with different imputation approaches we found somewhat similar, but not identical identification of the location of response shift in our study. Table 17-Table 24 demonstrates assessment of response shift using Oort and Schmitt approaches using

the four imputation approaches. The uncertainty with results obtained after imputation cannot be ignored and should be interpreted with caution.

Implications

The findings in this study can serve to inform future research of response shift for PF in hypertensive CAD patients. Barclay-Goddard et al. [39] suggest that response shift presence in PF will affect the way in which one evaluates PF in the clinic and in research studies, the type of measures that we use, and the way in which change in physical function is analyzed.

Schwartz and Rapkin [90] describe the difference between performance based measures (such as a timed walking test), perception based measures (such as an individual describing how many times a day they walk), and evaluation based/self perceived measures (i.e., how difficult is it to walk?). They describe changes over time that is not true change may be: measurement error in performance based measures, response bias in perception based measures, or response shift in evaluation-based methods [90]. We assert that the evaluation of 'health limits physical activity' in SF-36 PF scale also interpreted as 'difficulty in PF' is self perceived and therefore may be susceptible to response shift. A comparison of models of PF with only performance based measures and models with self perceived measures would be of interest in CAD and other chronic conditions.

Barclay-Goddard et al. [39] recommend that 'if measures of PF demonstrate response shift, choice of self perceived or performance based measures need to be carefully chosen by the clinician to avoid bias in change estimates'. Identification of response shift in SF-36 PF in our study and PF construct in Barclay-Goddard study [39]

warrant assessment of response shift when using self-perceived measures in similar populations and other chronic populations.

Measurement of change in HRQoL will be inaccurate if response shift is present, but not accounted for. The effect size of HRQoL change can be estimated, taking response shift into account, after evaluation with the Oort SEM technique. In the present study, HRQoL change was small after accounting for response shift. This was demonstrated in PF, where observed change (true change plus response shift) in physical activity had a small effect size which increased slightly when response shift was removed. Moreover, for PF, the effects of response shifts and true change were in opposite directions. In other words, after accounting for response shift, we observe a reversal in sign from negative (for observed score) to positive (for true change). According to Rosenberg [91] a distorter variable reveals that the correct interpretation is precisely the reverse of that suggested by the original data. Analogous to Rosenberg's report [91], response shift acts a distorter variable, which, when controlled, causes the relationship between the independent and dependent variables to change direction. In our case, it changed the direction from negative sign for the observed scores to positive sign for true change scores.

The literature on the SF-36 health survey in a variety of disease states has suggested a minimal clinically meaningful difference to be 7-16 points on the SF-36 PF scale as they reflect a magnitude of change perceptible to patients [92]. After accounting for response shift with the Oort procedure, the mean PF score at one year was estimated to be 58, resulting in a mean change of ~3 points. Our finding of an average 3 point change was a little less than one half to one fifth of the 7-16 point

difference deemed clinically significant. Previous research has shown that a lower SF-36 Physical Component Summary score (note: physical component score consists of physical functioning) may be associated with an increased risk of stroke and predicts higher mortality [93]. Although there was an improvement in mean scores for the PF scale (3 points), presence of response shift resulted in slight worsening of PF scale scores over time. Initially, we found a decrease in change in scores from pre- to post-test. However, this relationship did not take into account the effects of response shift (acting as a distorter). After controlling for response shift, we found that the change from pre- to post-test was positive, i.e., there was an improvement in PF scores over time. It should be noted that the true change was marginal, warranting the need for future studies to investigate the role of response shift as a distorter variable.

According to Wyrwich et al. [94], measuring clinically meaningful differences will likely facilitate the clinical interpretation of health status changes experienced by patients with heart disease and improve clinical decision-making. The authors suggest that HRQoL measures provide valuable evidence on changes in emotional, social and physical functioning that help assess relevant component of a patient's condition and highlight different perspectives concerning treatment options that can be balanced against changes in the patient's health status [94]. However, the authors agree that clinically meaningful differences are not sufficient to drive subsequent clinical, treatment, and reimbursement decisions. They recommend use of clinically meaningful differences in research settings to demonstrate the clinically important effects of interventions on HRQoL [94]. Future studies measuring clinical importance has

implications for interpreting the measurement of change in both clinical and research endeavors.

Barclay-Goddard et al. [34,39] recommend that clinically, if recalibration response shift is suspected in a particular construct of HRQoL (as demonstrated in PF in our study), a design method such as a then-test could be used to capture response shift over time during treatment. They further propose to measure response shift in clinical trial research, as the estimates of the treatment effects may be underestimated and therefore inaccurate, possibly leading to a conclusion that is false negative. In studies or settings where response shift is suspected and design methods are not feasible to assess the amount of response shift, statistical methods should be applied such as those applied here.

Response shift assessment is particularly important in hypertensive CAD patients when impairments and activity limitations may never recover fully, but improved HRQoL is a goal. If poor PF observed in patients with hypertension and CAD is associated with poor QOL, this information may be used by health professionals to target recovery interventions to those who might be most responsive. Increased knowledge of response shift will therefore affect the way in which HRQoL measures are used in clinical, research, and policy decisions.

Limitations

Limitations should be noted. Ahmed et al. [35] report that SEM approaches detect response shift at a group level. Only if a substantial number of participants undergo a response shift, will it be detected using either of the Oort [10] or Schmitt SEM [28] procedure. Thus, if barely a few individuals undergo a response shift on the SF-36 domains, it may not be detected using either of the two statistical approaches. Barclay-

Goddard et al. [34,39] suggest, not all individuals would experience response shift. However, those individuals who do experience response shift may not experience it at the same time. This may influence study results when using group level analyses similar to our study.

Converse to the group level analyses, Mayo et al. [95,96] propose a method that uses the residuals generated from a growth model where health behaviors measured or functions were used to predict perceived health. Ahmed et al. [35] suggest that if there are enough time points (i.e. greater than three time points), analytical approaches that examine response shift at an individual level, such as growth curve modeling [97], and latent trajectory analysis described by Mayo et al. [95,96] will be more helpful to identify subgroups of patients who experience a response shift. Future comparisons with individual level methods are warranted.

The use of residuals to identify those who potentially underwent response shift is an intriguing area of study. This would enable multi-group analysis in SEM with, for example, those who appear to have undergone response shift based on residuals, and those who do not appear to have undergone response shift. This is a promising area of study with secondary data analysis. Cognitive interviews, used in previous response shift studies [98] are shown to provide insight into whether the statistical, design, and individualized approaches are a true reflection of response shift.

We identified response shift among hypertensive CAD patients assigned to antihypertensive treatment strategies. However, availability of a control group not expected to experience a response shift would provide support for our results. Power and sample size in large longitudinal models needs to be considered. In our study, the

sample size was 909 with more than 100 parameters estimated. It has been suggested that the number of cases to free parameters estimated be at least ten to one. For the unconstrained longitudinal model, a sample size of 909 might be considered somewhat small. The measurement model that we used in the response shift detection procedure using both Oort and Schmitt procedures fitted the data closely (RMSEA = 0.05). With a sample size of 909, the statistical power to reject such a model if it had been unreasonable was 95% [78]. We did also take a look at presence of response shift when comparing complete cases data set with imputed data sets. We found similar, but not identical identification of the location of response shift in our study when using imputed data sets. This deserves future study.

For Research Question# 2

Based on the findings in this study, we found that patients recalibrated their perception of PF after assignment to treatment strategies, i.e. at one year. When investigating from the measurement perspective, we found that the relationship between the SF-36 scale (PF) and gender cannot be explained via their relationship with PHYS HRQoL. We chose to allow gender to have a direct effect on PF indicating that the measurement bias is in the PF scale. The violation of measurement invariance was not consistent across occasions (estimated at 0.113 at baseline and 0.150 at one year, Table 4-15) which indicated that male patients reported better PF than female patients, even if their PHYS HRQoL was similar. King-Kallimanis et al. [32] recommend that the PF scale contains specific questions which can be answered almost objectively. According to the authors, it is difficult for individualized interpretation of the PF items. They believe that other SF-36 scales measuring PHYS HRQoL are more prone to subjective interpretations. However, to obtain a more parsimonious model with the

same fit, they allowed for a direct effect of age on PF scale in their study. By analogy to their study, we chose the final model based on the presence of measurement bias and response shift in the smallest number of scales, providing a more parsimonious model.

The order and absolute value of modification indices assist in subjective decision-making of the SEM model. In addition to the modification indices in our study, past research played a role in deciding the best model fit when we allowed for a direct effect of gender on PF scale. De Graff et al. [84] found a relationship between gender and the SF-36 domains for PF, RP, SF, and RE among CAD patients. In agreement to our study results, Sevinc and Akyol [85] found better QOL scores among male CAD patients compared to female CAD patients. In a separate study, Norris et al. [87] report, when compared to men, women with CAD had lower scores in all QOL domains. Few studies [88,89] report of a stereotype implying that men are more physically oriented and invested to do heavier physical labor or strenuous activities. The authors of these studies allege that if PF (e.g., walking long distances, heavy lifting) is an integral part of the average male's self-image, then it would not be surprising to observe men underreporting their factual physical limitations (i.e., higher PF scores) to preserve their self-image [88,89]. In other words, men seem to be over-reporting their PF to maintain their perceptions of their masculinity and to protect their egos by reducing gender role conflict. Our study results are consistent with others' findings that gender roles operate differently for males and females [88,89]. Further assessment into the reasons for the apparent gender difference is warranted.

For exploratory purposes, we operationalized presence of depressive symptoms as change in CES-D scores over time (note: for the original analysis we used baseline

and one-year CES-D scores as separate variables). We found that change in CES-D scores had a direct effect on PF at the second occasion. The effect of change in CES-D scores on PF at the second measurement occasion was positive (0.086), suggesting that at one year depressed patients reported better PF than less depressed patients, even if their true HRQoL was similar. However, use of change in CES-D scores as a covariate led to problems with interpretation since it is not known if the direct effect is an indication of measurement bias and/or presence of response shift. The differences in results warrants further investigation on the operationalization of depression as a variable to detect response shift. Future studies should define depression in different ways to enable proper interpretation of the results.

We found that depressed patients were not likely to undergo response shift. Few studies corroborate our study findings where they suggest that depressed individuals will not undergo response shift. Sprangers and Schwartz [4] claim that depression might impede the occurrence of response shift. A separate study by Bar-On et al. [15] found that depressed individuals were less likely to undergo recalibration response shift. It may be plausible that depression may reduce a patient's tendency to think about other events such as side effects, therefore, these patients are less likely to undergo response shift.

However, contrary to our findings, Lenert et al. [99] suggest depressive patients may have systematic differences in how they report their values in psychometric testing. Gibbons et al. [100] report that depressed patients may report higher QOL since they may re-evaluate their baseline status and decide that they were much worse in the past compared to now (i.e., negative response shift) resulting in response shift. The authors

suggest downward comparison among persons who are depressed and individuals who are ill and have a pessimistic prognosis i.e., those expecting to get worse may result in a negative response shift. They explain that learning of others whose deterioration is more advanced or more evident can remind individuals of their own decline. They suggest that a pessimistic shift in focus is more likely among persons who are depressed to begin with. Moreover, they recommend that by preventing depression one may moderate the type of response shift experienced.

In the framework, validation of the full model only is suggested. There is an opportunity to assess validation through the process of the identification of location of response shift. We used MCMC and found similar, but not identical identification of the location of response shift in our study which deserves future study.

Implications

Our findings have several clinical and public policy implications. Women have a higher incidence of depression, and in combination with lower PF scores would worsen their QOL. Ameliorating depression-caused reductions in PF and maintaining an individual's ability and willingness to perform activities of daily living may reduce the need for earlier institutionalization and increase independent living time.

Boini et al. [65] advocate that impairment in a particular HRQoL dimension should be followed up by the implementation of specific strategies. For instance, Boini et al. [65] suggest that when impairment is observed on mental dimensions, psychologist should play a major role in patient's medical management. Likewise, we advocate that when impairment is observed on physical dimensions, cardiac rehabilitation programs including doctors and exercise physiologists should play a larger part in patient's medical management. Patients should be educated that self-management must take

into account not only classical aspects such as diet low in transfat, saturated fat, cholesterol weight, and exercise programs, but also adaptation to the illness and its treatment, that is strategies such as stress or anger management for self-care of the disease [101]. In addition, cardiac rehabilitation which includes a structured program of education and activity guided toward lifestyle modification, increasing functional capabilities, improving QOL, and emotional support should be made available to improve impact of CAD on patient daily life [101].

Comparing the influence of different diseases on HRQoL is relevant, since the attitude of the doctor (and society) should be influenced by the impact of a disease on patient's QOL. Furthermore, by looking more closely at the scores for each domain of the SF-36, it is possible to give specific attention to and direct treatment for the most impaired aspects of QOL of patients with CAD and those vulnerable to response shift; our findings suggest that SF-36 PF domain was vulnerable to bias in this population. Sevinc and Akyol [85] advocate that nurses dealing with CAD patients should constantly incorporate their knowledge about the changes in patients' risk factors and QOL requirements into their clinic and nursing plans. Understanding how changes in magnitude and direction of the response shift affects perception of health can provide valuable information when comparing recovery among CAD populations. Assessment of response shift thus is pivotal and warrants further understanding of how different group of patients perceive their QOL over time.

The findings from our study also have implications for cost-effectiveness studies. By analogy to the findings from Lenert et al. [99], our results suggest that bias in self reports of HRQoL could result in inadvertent discrimination against patients with poor

health. Assuming other domain scores are similar across males and females, if females were less optimistic in their PF rating scale, it would lead to over-evaluation of difference in utility between males and females. In agreement to Lenert et al. [99], this would result in an increase in the cost-effectiveness of treatments for the females and potentially direct health resources away from the males.

Analogous to this, let's assume that treatment A and treatment B have similar utilities at baseline and at one year resulting in similar cost-effectiveness ratios for both treatment strategies. If however, after accounting for response shift, we may find that change in utilities for treatment B was smaller than previously estimated. Assuming costs to be constant, we observe that there is an increase in cost-effectiveness for patients assigned to treatment A. In one other instance, let's presume that patients assigned to treatment A have utility values of 0.5 and 0.7 at baseline and at one year, respectively. Comparatively, patients assigned to treatment B have utility values of 0.3 and 0.4 at baseline and at one year, respectively. We may conclude that treatment A is more cost-effective compared to treatment B. However, after accounting for response shift we may find that patients assigned to treatment B had a higher change in utility equal to 0.3 instead of 0.1 as shown above. In this case, contrary to our previous results, we observe that treatment B is more cost-effective. With our study results and the two examples stated above we propose the need to demonstrate the effect of response shift on the estimation of utilities and cost-effectiveness studies.

Limitations

The limitations of the study should be noted when interpreting and applying findings to other populations. First, the CES-D score is not a diagnosis of depression, but a measure of depressive symptoms experienced in the previous two weeks.

Second, missing data may introduce potential bias. Respondents who completed both baseline and one year surveys may differ from non-respondents. Third, the relationship between depression and physical functioning is likely reciprocal. The model presented in this study assumes that the causal path goes from depressive symptoms to physical functioning and not vice versa. The recursive nature of this model should be tested.

It should be noted that the sample size of 788 patients might be considered small relative to the number of parameters estimated and statistical tests. However, we have almost 100% power to reject the hypothesis that the model does not fit our data (according to a power analysis based on RMSEA values of 0.05 and 0.10).

By analogy to the King-Kallimanis et al. study [32], an issue specific to the model we selected for the measurement perspective is that the distinction between explanatory variables and violating variables may not always be clear. We distinguished between explanatory variables and potential violating variables based upon prior knowledge and handled variables during model-fitting. However, we agree with King-Kallimanis et al. [32], to avoid unjust data exploration and chance capitalization, researchers should have a well-defined and clear understanding of the research questions and hypotheses.

CHAPTER 6
CONCLUSION

The overall objective of this thesis was to assess response shift in hypertensive CAD patients. For the first objective, we chose to compare Oort [10] and Schmitt [28] SEM approaches to identify response shift in hypertensive CAD patients. Both the Oort and Schmitt approaches identified recalibration for the SF-36 PF scale. We found divergent results in our study compared to the Ahmed et al. study when identifying response shift using Oort and Schmitt procedures. First, our study applied the Bonferroni correction to avoid making a Type I error when identifying a response shift with the Oort procedure. The Ahmed et al. [35] study did not control for Type I error even though the Oort procedure [10] has been suggested to increase the chances of Type I error. As a result, their study may have wrongly identified a response shift when differences may be attributed to other types of changes.

Second, contrary to the Oort procedure adopted by Ahmed et al. [35] in their study where they chose to release constraints on residuals, followed by intercepts, and then factor loadings, we conducted the Oort approach as recommended and tested by several other authors including Oort [10], Donaldson [38] and Barclay-Goddard [39]. Similar to the approach tested by these authors, we chose to test sequentially for invariance of factor loadings, intercepts, and error variance. However, even when using the Oort procedure adopted by Ahmed et al. [35], we found similar results. Using either of the two different sequence of testing did not alter our results; order of testing did not matter in our study. Our results were robust to either of the two order of testing, justifying the presence of response shift in our study population. Ahmed et al. [35] may have found divergent results if they would have tested for the presence of response shift

by releasing parameter constraints on factor loadings, followed by intercepts, and then residuals.

Based on our results, the differences between our study and Ahmed et al. study appear due to variation in the method and the sample used. Method differences resulted due to use of Bonferroni and order of testing for response shift as suggested by previous authors. Sample differences may have resulted due to a larger proportion of the study sample experiencing a response shift compared with individuals with COPD, making response shift among hypertensive CAD patients more detectable.

To the best of our knowledge this is the second study to identify response shift in PF. Nonetheless, ours is the only study to identify response shift for SF-36 PF scale. The issue of response shift in PF among hypertensive CAD patients and gender role differences deserves further evaluation. The findings in this study can serve to inform future research of response shift in PF in hypertensive CAD patients. According to Barclay-Goddard et al. [34,39], response shift presence in PF will affect the way in which we evaluate PF in the clinic and in research studies, the type of measures that we use, and the way in which change in PF is analyzed. Facilitating response shift can be particularly important when impairments and activity limitations are not expected to recover fully in hypertensive CAD patients, but improved HRQoL is a goal.

Given the divergent results from Ahmed et al. [35] and our study comparing the Schmitt and Oort SEM approaches to detect response shift justify the need to compare these methods to identify response shift in future studies. Groups of individuals with different health conditions and catalysts for change should be evaluated with the Oort and Schmitt SEM procedures to allow further understanding of response shift over time

in various health conditions. Response shift may influence the measurement characteristics of HRQoL measurement instruments, so it is important to measure response shift to assist in understanding assessment of HRQoL.

Our findings suggest that the SF-36 PF scale was vulnerable to bias in this study population. By looking more closely at the scores for SF-36 PF domain in this study population will enable us to provide nuanced attention and direct treatment for the most impaired aspects of QOL. We also found that gender caused response shift in measurement. Females were less optimistic in their PF rating scale which would lead to over-evaluation of difference in utility between males and females. This would result in an increase in the cost-effectiveness of treatments for the females and potentially direct health resources away from the males. Women have a higher incidence of depression, and in combination with lower PF scores would worsen their QOL. Thus, it is imperative to improve depression-caused reductions in PF and maintain an individual's ability and willingness to perform daily activities which may help reduce institutionalization.

Knowledge of multiple comorbid conditions present in an individual may help us obtain a better understanding of how one may approach the issues that come along with these conditions, specifically in elderly patients. Assessment of response shift is pivotal and warrants further understanding of how different group of patients perceive their QOL over time.

LIST OF REFERENCES

1. Rabins PV, Black BS. Measuring quality of life in dementia: purposes, goals, challenges, and progress. Int Psychogeriatr 2007;19:401-07.

2. Ahmed S, Mayo N, Wood-Dauphinee S, Hanley J, Cohen S. The structural equation modeling technique did not show a response shift, contrary to the results of the then test and the Individualized approaches. J Clin Epidemiol 2005;58:1125-33.

3. Ahmed S, Mayo N, Wood-Dauphinee S, Hanley J, Cohen S. Response Shift influenced estimates of change in Health-Related Quality of Life Poststroke. J Clin Epidemiol 2004;57:561-70.

4. Sprangers MA, Schwartz CE. Integrating Response Shift into Health-Related Quality of Life Research: A Theoretical Model. Soc Sci Med 1999;48:1507-15.

5. Schwartz CE, Sprangers MA. Methodological Approaches for Assessing Response Shift in Longitudinal Health-Related Quality of Life Research. Soc Sci Med 1999;48:1531-48.

6. Wilson IB. Clinical Understanding and Clinical Implications of Response Shift. Soc Sci Med 1999;48:1577-88.

7. Pepine CJ, Handberg-Thurmond E, Marks RG, Conlon M, Cooper- DeHoff R, 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. J Am Coll Cardiol 1998;32:1228 –37.

8. 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. JAMA 2003;290:2805–16.

9. Korfage IJ, de Koning HJ, Essink-Bot ML. Response shift due to diagnosis and primary treatment of localized prostate cancer: a Then test and a vignette study. Qual Life Res 2007;10,1627-34.

10. Oort FJ. Using structural equation modeling to detect response shifts and true change. Qual Life Res 2005;14:587-98.

11. Ried LD, Tueth MJ, Taylor MD, Sauer BC, Lopez LM, Pepine CJ. Depressive symptoms in coronary artery disease patients after hypertensive treatment. Ann Pharmacother 2006;40:597-604.

12. Aggarwal A, Ades PA. Exercise rehabilitation of older patients with cardiovascular disease. Cardiol Clin 2001;19:525-36.

13. Croog SH, Levine S, Testa MA, Brwon B, Bulpittt CJ, Jenkins CD, et al. The effects of antihypertensive therapy on the quality of life. N Engl J Med 1986;314(26):1657-64.

14. Mitchell RA, Imperial E, Kelleher P, Brunker P, Gass G. Perceived health problems in subjects with varying cardiovascular diagnoses. J Behav Med 1991;14:505-12.

15. Bar-On D, Lazar A, Amir M. Quantitative assessment of response shift in QOL research. Soc Indic Res 2000;49:37-49.

16. 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. JAMA 2003;289:2560-72.

17. Ried LD, Tueth MJ, Handberg E, Kupfer S, Pepine CJ, and 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). Psychosom Med 2005;67:398-406.

18. Ried LD, Johnson RE, McFarland BH, Brody KK. Antihypertensive drug use and the risk of depression among older hypertensives in an HMO. J Pharmacoepi 2000;8:1-25.

19. Ried LD, McFarland BH, Johnson RE, Brody KK. Betablockers and depression: the more the murkier? Ann Pharmacother 1998;32:699-708.

20. Ko DT, Hebert PR, Coffey CS, Sedrakyan A, Curtis JP, Krumholz HM. Beta-blocker therapy and symptoms of depression, fatigue, and sexual dysfunction. JAMA 2002;288:351-57.

21. Lacy CF, Armstrong LL, Goldman MP, Lance LL. Drug Information Handbook. 8th ed. Hudson: OH; 2000-01.

22. Avorn J, Everitt DE, Weiss S. Increased antidepressant use in patients prescribed beta-blockers. JAMA 1986; 255:357-60.

23. Thiessen BQ, Wallace SM, Blackburn JL, Wilson TW, Bergman U. Increased prescribing of antidepressants subsequent to beta-blocker therapy. Arch Intern Med 1990;150:2286-90.

24. Bright RA, Everitt, DE. Beta-blockers and depression. Evidence against an association. JAMA 1992;267:1783-87.

25. Gerstman BB, Jolson HM, Bauer M, Cho P, Livingston JM, Platt R. The incidence of depression in new users of beta-blockers and selected antihypertensives. J Clin Epidemiol 1996;49:809-15.

26. Lindberg G, Bingefors K, Ranstam J, Rastam L, Melander A. Use of calcium channel blockers and risk of suicide: ecological findings confirmed in population based cohort study. Br Med J 1998;316:741-5.

27. 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? Qual Life Res 2005;14:611-27.

28. Schmitt N. The use of analysis of covariance structures to assess beta and gamma change. Multivariate Behav Res 1982;17:343-58.

29. Schmitt N, Pulakos ED, Lieblein A. Comparison of three techniques to assess group-level beta and gamma change. Appl Psychol Meas 1984; 8: 249-60.

30. Schaubroeck J, Green SG. Confirmatory factor analytic procedures for assessing change during organizational entry. J Appl Psychol 1989;74(6):892-900.

31. Oort FJ, Visser MRM, Sprangers MAG. Formal definitions of measurement bias and explanation bias clarify measurement and conceptual perspectives on response shift. J Clin Epidemiol 2009;62:1126-37.

32. King-Kallimanis BL, 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. J Clin Epidemiol 2009;62:1157-64.

33. Bush DE, Ziegelstein RC, Tayback M, Richter D, Stevens S, Zahalsky H, et al. Even minimal symptoms of depression increase mortality risk after acute myocardial infarction. Am J Cardiol 2001;88:337-41.

34. Barclay-Goddard R, Epstein JD, Mayo NE. Response shift: a brief overview and proposed research priorities. Qual Life Res 2009;18:335-46.

35. 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. J Clin Epidemiol 2009;62:1165-72.

36. 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. Qual Life Res 2005;14:599-609.

37. Visser M, Oort F, Sprangers M. Methods to Detect Response Shift in Quality of Life Data: A Convergent Validity Study. Qual Life Res 2005;14:629-39.

38. Donaldson GW. Structural equation models for quality of life response shifts: Promises and pitfalls. Qual Life Res 2005;14:2345-51.

39. Barclay-Goddard R, Lix LM, Tate R, Weinberg L, Mayo NE. Response shift was identified over multiple occasions with a structural equation modeling framework. J Clin Epidemiol 2009;62:1181-88.

40. Howard GS, Ralph KM, Gulanick NA, Maxwell SE, Nance SW, Gerber SK. Internal invalidity in pretest-posttest self-report evaluations and a reevaluation of retrospective pretests. Appl Psychol Meas 1979;3:1-23.

41. Golembiewski R, Billingsley K, Yeager S. Measuring Change and Persistence in Human Affairs: Types of Change Generated by OD Designs. J Appl Behav Sci 1976;12:133-57.

42. Rapkin BD, Schwartz CE. Toward a theoretical model of quality of life appraisal: Implications of findings from studies of response shift. Health Qual Life Outcomes 2004;2:14.

43. Breetvelt IS, van Dam FSAM. Underreporting by cancer patients: the case of response shift. Soc Sci Med 1991;32:981-87.

44. Groenvold M, Fayers PM, Sprangers MAG, Bjorner JB, Klee MC, Aaronson NK, et al. Anxiety and depression in breast cancer patients at low risk of recurrence compared with the general population – unexpected findings. J Clin Epidemiol 1999,52(6):523-30.

45. Padilla GV, Mishel MH, Grant MM. Uncertainty, appraisal and quality of life. Qual Life Res 1992;1:155-65.

46. Wilson IB, Cleary PD. Linking clinical variables with health related quality of life: a conceptual model of patient outcomes. JAMA 1995;273:59-65.

47. Sprangers MAG, Van Dam FSAM, Broersen J, Lodder L, Wever L, Oosterveld P, et al. Revealing response shift in longitudinal research on fatigue: The use of the then-test approach. Acta Oncol 1999;38:709-18.

48. Jansen SJ, Stiggelbout AM, Nooij MA, Noordijk EM, Kievit J. Response shift in quality of life measurement in early-stage breast cancer patients undergoing radiotherapy. Qual Life Res 2000;9:603-15.

49. Adang EMM, Kootstra G, Engel GL, van Hooff JP, Merckelback HLGJ. Do retrospective and prospective quality of life assessments differ for pancreas-kidney transplant recipients? Transpl Int 1998;11:11-5.

50. Schwartz CE, Coulthard-Morris L, Cole B, Vollmer T. The quality of life effects of Interferon-Beta-1b in multiple sclerosis: An Extended Q-TWiST analysis. Arch Neurol 1997;54:1475-80.

51. Hagedoorn M, Sneeuw KC, Aaronson NK. Changes in physical functioning and quality of life in patients with cancer: Response shift and relative evaluation of one's condition. J Clin Epidemiol 2002;55:176-83.

52. Schwartz CE, Feinberg RG, Jilinskaia E, Applegate JC. An evaluation of a psychosocial intervention for survivors of childhood cancer: Paradoxical effects of response shift over time. Psychooncology 1999;8:344-54.

53. Rees JE, Waldron D, O'Boyle CA, MacDonagh RP. Response shift in individualized quality of life in patients with advanced prostate cancer [abstract]. Clin Ther 2002; 24 Suppl B:33-4.

54. Howard GS. Response-shift bias: A problem in evaluating interventions with pre/post self-reports. Eval Rev 1980;4(1):93-106.

55. Howard GS, Dailey PR. Response-shift bias: A source of contamination of self-report measures. J Appl Psychol 1979;64:144-50.

56. Howard GS, Ralph KM, Gulanick NA, Maxwell SE, Nance DW, Gerber SK. Internal invalidity in pretest-posttest self-report evaluations and a reevaluation of retrospective pretests. Appl Psychol Meas 1979;3:1-23.

57. Howard GS, Schmeck RR, Bray JH. Internal invalidity in studies employing self-report instruments: A suggested remedy. JEM 1979;16:129-35.

58. Terborg JR, Howard GS, Maxwell SE. Evaluating planned organizational change: A method for assessing alpha, beta, and gamma change. Acad Manage Rev 1980;5:109-21.

59. King RB. Quality of life after stroke. Stroke 1996;27(9):1467-72.

60. engs TO, Yu M, Luistro E. Health-related quality of life after stroke – A comprehensive review. Stroke 2001;32(4):964-71.

61. Breetvelt IS, Van Dam FS. Underreporting by cancer patients: The case of response shift. Soc Sci Med 1991;32(9):981-87.

62. Decker SD, Schulz R. Correlates of life satisfaction and depression in middle-aged and elderly spinal cord-injured persons. Am J Occup Ther 1985;39(11):740-5.

63. World Health Organization [Internet]. International Classification of Functioning, Disability, and Health. [updated 2001; cited 2010 Oct 5]. Available from: http://books.google.com/books?id=pwb9ywSVKxwC&dq=ICF,+WHO&printsec=frontcover&source=bn&hl=en&ei=2wD9SaTHNMWEtweZ05nFCg&sa=X&oi=book_result&ct=result&resnum=4.

64. Wang R, Zhao Y, He X, Ma X, Yan X, et al. Impact of hypertension on health-related quality of life in a population-based study in Shanghai, China. Public Health 2009;13(8):534-9.

65. Boini S, Briancon S, Guillemin F, Galan P, Hercberg S. Occurrence of coronary artery disease has an adverse impact on health-related quality of life: A longitudinal controlled study. Int J Cardiol 2006;113(2):215-22.

66. Testa MA, Anderson RB, Nackley JF, Hollenberg NK. Quality of life and antihypertensive therapy in men. A comparison of captopril and enalapril. The quality of life hypertension study group. N Engl J Med 1993;328(13):907-13.

67. Wald A, Scarpignato C, Kamm MA, Mueller-Lissner S, Helfrich I, Schuijt C, et al. The burden of constipation on quality of life: results of a multinational survey. Aliment Pharmacol Ther 2007;26:227-36.

68. Ried LD, Gandhi PK. Predictors of health status change among patients treated with a calcium antagonist or an atenolol-led hypertension strategy in the Internation Verapamil SR-Trandolapril Study (INVEST) [abstract]. Value Health 2008;11(6):A411.

69. Ahmed S, Mayo N, Corbiere M, Wood-Dauphinee S, Hanley J, Cohen R. Using the Patient Generated Index to Evaluate Response Shift Post Stroke. Qual Life Res 2005;14:2247-57.

70. Nolte S, Elsworth GR, Sinclair AJ, Osborne RH. Tests of measurement invariance failed to support the application of the “then-test”. J Clin Epidemiol 2009;62:1173-80.

71. Osborne RH, Elsworth GR, Sprangers MAG, Oort FJ, Hopper JL. The value of the Hospital Anxiety and Depression Scale (HADS) for comparing women with early onset breast cancer with population-based reference women. Qual Life Res 2004;13:191-206.

72. Joint National Committee on Prevention Detection Evaluation and Treatment of High Blood Pressure. The sixth report of the Joint National Committee on prevention, detection, evaluation, and treatment of high blood pressure. Arch Intern Med 1997;157:2413-46.

73. Saris-Baglama RN, Dewey CJ, Chisholm GB, Kosinski M, Bjorner JB, Ware JE. SF-36 Health Outcomes Scoring Software. 2004:1-157.

74. 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.

75. Brown TA. Confirmatory factor analysis for applied research. Guilford Press. New York: London, Guilford Press; 2006.

76. Jöreskog K, Sörbom D. LISREL 8.8: Structural Equation Modeling with the SIMPLIS Command Language. Lincolnwood, Il: Scientific Software Internatonal Inc.; 1993.

77. Ludbrook J. Statistics in Physiology and Pharmacology: A Slow and Erratic Learning Curve. Clin Exp Pharmacol Physiol 2001;28:488-92.

78. MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. Psychol Methods 1996;1(2):130-49

79. Hochberg Y, Benjamini Y. More Powerful Procedures for Multiple Significance Testing. Stat Med 1990;9:811-8.

80. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. Appl Psychol Meas 1977;1:385-401.

81. Sachdev M, Sun JL, Tsiatis AA, Nelson CL, Mark DB, Jollis JG. The prognostic importance of comorbidity for mortality in patients with stable coronary artery disease. J Am Coll Cardiol 2004;l43:576–82.

82. Chirinos JA, Veerani A, Zambrano JP, Schob A, Perez G, Mendez AJ, et al. Evaluation of comorbidity scores to predict all-cause mortality in patients with established coronary artery disease. Int J Cardiol 2007;117:97-102.

83. Meyer-Rosberg K, Burckhardt CS, Huizar K, Kvarnstrom A, Nordfors L, Kristofferson A. A comparison of the SF-36 and Nottingham Health Profile in patients with chronic neuropathic pain. Eur J Pain 2001;5:391-403.

84. 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. Ann Vasc Surg 2002;16:495-500.

85. Sevinc S, Akyol AD. Cardiac risk factors and quality of life in patients with coronary artery disease. J Clin Nurs 2010;19:1315-25.

86. 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. J Psychosom Res 1987;31:521-9.

87. Norris CM, Ghali WA, Galbraith PD, Graham MM, Jensen LA, Knudtson ML for the APPROACH investigators. Women with coronary artery disease report worse health-related quality of life outcomes compared to men. Health Qual Life Outcomes 2004;2:21.

88. 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. Soc Sci Med 1999;48:89-98.

89. 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.

90. Schwartz CE, Rapkin BD. Reconsidering the psychometrics of quality of life assessment in light of response shift and appraisal. Health Qual Life Outcomes 2004;2(1):16.

91. Rosenberg MJ. The Logic of Survey Analysis. New York:Basic Books; 1968.

92. Lauridsen HH, Hartvigsen J, Manniche C, Korsholm L, Grunnet-Nilsson N. Responsiveness and minimal clinically important difference for pain and disability instruments in low back pain patients. BMC Musculoskelet Disord 2006;7:83.

93. Myint PK, Surtees PG, Wainwright NWJ, Luben RN, Welch AA, Bingham SA, et al. Physical health-related quality of life predicts stroke in the EPIC-Norfolk. Neurology 2007;69(24):2243-8.

94. Wyrwich KW, Spertus JA, Kroenke K. Tierney WM, Babu AN, Wolinsky FD. Clinically important differences in health status for patients with heart disease: An expert consensus panel report. Am Heart J 2004;147(4):615-22.

95. Mayo NE, Scott SC, Ahmed S. Case management poststroke did not induce response shift: the value of residuals. J Clin Epidemiol 2009;62:1148-56.

96. Mayo NE, Scott SC, Dendukuri N, Ahmed S, Wood-Dauphinee S. Identifying response shift statistically at the individual level. Qual Life Res 2008;17:627-39.

97. 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. J Pediatr Psychol 2002;27:97-107.

98. Bloem EF, van Zuuren FJ, Koeneman MA, Rapkin BD, Visser MRM, Koning CCE, et al. Clarifying quality of life assessment: do theoretical models capture the underlying cognitive processes? Qual Life Res 2008;17:1093-1102.

99. Lenert LA, Treadwell JR, Schwartz CE. Associations between health status and utilities implications for policy. Med Care 1999;37:479-489.

100. Gibbon FX. Social comparison as a mediator of response shift. Soc Sci Med 1999;48:1517-30.

101. Mayo Clinic.com [Internet]. Mayo Foundation for Medical Education and Research. [updated 2010 July 2; cited 2010 Oct 5]. Available from: http://www.mayoclinic.com/health/coronary-artery-disease/DS00064/DSECTION=lifestyle-and-home-remedies.

BIOGRAPHICAL SKETCH

Pranav Kirit Gandhi was born in 1980 at Mumbai, India to Kirit and Kalpana Gandhi. He grew up there with his parents. His younger sister was born in 1983. Pranav did all his schooling in Mumbai. He attended the Mumbai Educational Trust Institute of Pharmacy, Mumbai, India and earned a bachelor's in pharmaceutical sciences in 2002. He then attended graduate school at University of the Science in Philadelphia and earned a master's in pharmacy administration in 2006. He then attended graduate school at the University of Florida (UF), Gainesville, Florida and earned a doctorate in pharmaceutical sciences – pharmaceutical outcomes and policy in 2010. While at UF, he got married to his love interest, Poonam Sanjanwala in 2009.