

Minimal evidence of response shift in the absence of a catalyst

Sara Ahmed · Richard Sawatzky · Jean-Frédéric Levesque ·
Deborah Ehrmann-Feldman · Carolyn E. Schwartz

Accepted: 15 April 2014 / Published online: 5 June 2014
© Springer International Publishing Switzerland 2014

Abstract

Background Individuals with chronic conditions experience fluctuations in health status and thus may experience response shift. We sought to test the hypothesis that response shift effects would be non-significant among individuals with chronic disease who experienced relatively small changes in their health status over a 1-year period.

Methods This secondary analysis utilized longitudinal cohort data on a community-based sample ($n = 776$) representing four chronic diseases (arthritis, heart failure, diabetes, or chronic obstructive pulmonary disease). Information on health-care utilization was obtained from

the provincial health insurance database. Participants completed the SF-36 twice annually. Parameter invariance over 1 year in a second-order SF-36 factor structure was evaluated by adapting Oort's approach by fitting a second-order measurement structure with first-order factors for the SF-36 subscales and second-order factors for physical and mental health status while accommodating ordinal data.

Results Over 80 % of participants had no hospitalizations or emergency room visits over follow-up. The model had an acceptable fit when all measurement model parameters were constrained at both time points ($RMSEA = .035$, $CFI = .97$). There was no substantial difference in fit when measurement model parameters (item thresholds, first-order factor intercepts, and factor loadings) were allowed to vary over time.

Conclusion Among chronically ill individuals with stable health, substantial response shift effects were not detected.

Electronic supplementary material The online version of this article (doi:10.1007/s11136-014-0699-3) contains supplementary material, which is available to authorized users.

S. Ahmed (✉)
Faculty of Medicine, School of Physical and Occupational
Therapy, McGill University, 3654 Prom Sir-William-Osler,
Montreal, QC H3G 1Y5, Canada
e-mail: sara.ahmed@mcgill.ca

S. Ahmed
Clinical Epidemiology, McGill University Health Center,
Montreal, QC, Canada

S. Ahmed
Centre de recherche interdisciplinaire en réadaptation (CRIR),
Montreal, QC, Canada

R. Sawatzky
Trinity Western University School of Nursing, Langley, BC,
Canada

R. Sawatzky
Centre for Health Evaluation and Outcome Sciences, Vancouver,
BC, Canada

J.-F. Levesque
Centre de Recherche du Centre Hospitalier de l'Université de
Montréal, Montreal, QC, Canada

J.-F. Levesque · D. Ehrmann-Feldman
Université de Montréal, Montreal, QC, Canada

J.-F. Levesque
Institut National de Santé Publique du Québec, Montreal, QC,
Canada

D. Ehrmann-Feldman
Direction de Santé Publique de l'ASSS de Montréal, Montreal,
Canada

C. E. Schwartz
DeltaQuest Foundation, Inc., Concord, MA, USA

C. E. Schwartz
Departments of Medicine and Orthopaedic Surgery, Tufts
University Medical School, Boston, MA, USA

These results support the theoretical proposition that response shift is not expected to occur in patients with relatively stable conditions.

Keywords Structural equation modeling · Chronic disease · Response shift · Health-related quality of life

Introduction

The comparison of patient-reported outcome scores is based on a belief that the meaning of concepts and measurement scales remains stable in individuals' minds over time and is similar between groups [1–4]. This assumption may be challenged if there are differences between or within people over time in internal standards (recalibration), values (reprioritization), or conceptualizations (reconceptualization). These *response shift effects* reflect adaptation to changing health [5, 6] and can affect the psychometric characteristics of the outcome measurement tool [7].

Response shift theory posits that health-state changes are the requisite catalyst for response shift effects to take place [8, 9], since individuals with chronic disease are expected to experience changes in their physical health due to the natural vicissitudes of their diagnosis, as well as by effect of effective or ineffective intervention. They are thus expected to change their internal standards, values, or conceptualization of health-related quality of life because they are living day-to-day with the chronic illness. It is possible, however, that as they acquire strategies to help them cope and manage changes in health over time, they experience periods of stability when there is little change in health. In these conditions, there is no-response shift catalyst, and one would not expect to find detectable response shift effects. To date, response shift research has primarily documented conditions where response shift effects are expected and detected [10]. Our purpose is to evaluate whether response shift effects are detected in conditions where no-response shift is expected.

Recent advances in statistical methodology provide sensitive and useful tools for detecting and adjusting for response shift effects when they occur. One such approach, developed by Oort [11], builds on structural equation modeling (SEM) [12] to detect response shift effects. The Oort method has been successfully applied to patients with cancer [13], multiple sclerosis [14, 15], stroke [16], chronic obstructive pulmonary disease [17], and musculoskeletal diseases [18–21]. The method applies and systematically relaxes parameter constraints in the measurement models to detect evidence of recalibration, reprioritization, and reconceptualization response shifts [11, 22].

Several studies have evaluated response shift using SEM to assess changes in the generic Short-Form 36's (SF-36) measurement model over time by applying Oort's modeling approach [11, 12, 22]. These studies used a first-order model of the SF-36 and assessed response shift in the mental and physical component summary scores with respect to the SF-36 subscale scores. In most cases, the eight SF-36 subscales have not been represented as latent factors when evaluating response shift, which may be due to accommodating relatively small sample sizes and to the complexity of having to evaluate invariance of a large number of parameters (including all thresholds and first-order factor loadings) [23]. Consequentially, response shift has not been examined with respect to the original item responses but rather with respect to the computed domain scores following the SF-36 scoring algorithm. The limitation of this approach is that the possibility of response shift with respect to the first-order measurement structure (items within subscales) is ignored. A second-order factor model [24] that represents the relationships between the SF-36 summary scores, the SF-36 subscales, and the individual items, to our knowledge, has never been applied in a response shift evaluation. Such a model may yield different conclusions about the presence of response shift effects. That is, response shift can occur both with respect to the two second-order factors (mental and physical health components) as well as the eight first-order factors that represent the SF-36 subscales.

Given the importance of evaluating response shift to monitor changes in health for individuals with chronic disease in primary care [25], we aimed to assess response shift among a cohort of individuals with chronic disease over a 1-year period. On average, these individuals experienced relatively small changes in health over time, and, therefore, we hypothesized that very little or no-response shift effects would be detected when assessing response. Our aim was to test our hypothesis by examining whether response shift would be detected when evaluating changes over time using a second-order factor model of item responses with both the mental and physical health component summary scores and SF-36 subscale scores as latent factors.

Methods

Study sample and data collection procedures

This study is a secondary analyses of data from the Project Maladies Chroniques en Première Ligne (MACHRO-1) ([26, 27] (primary investigator Deborah Ehrmann-Feldman)). The data for this secondary analysis include 776 patients 18 years of age or older with arthritis, heart failure,

Table 1 Participant characteristics and baseline and follow-up scores on the SF-36

Characteristics	Participants	
Gender (%)		
Male	44.70 %	
Female	55.30 %	
Age, M(SD)	67.1 (SD: 11.6)	
Marital status (%)		
Married/spouse	56.70 %	
Single	15.90 %	
Divorced	10.80 %	
Widowed	15.50 %	
Diagnosis, N (%)*		
Arthritis	211 (27.2 %)	
Heart failure	150 (19.3 %)	
Diabetes	258 (33.3 %)	
COPD	157 (20.2 %)	
Baseline scores, mean (SD)	Baseline	Follow-up
PF	51.74 (29.97)	53.27 (28.86)
RP	53.65 (43.46)	52.68 (42.89)
BP	56.94 (30.05)	59.65 (29.85)
GH	55.46 (23.40)	56.98 (22.40)
VI	53.84 (22.65)	56.16 (21.21)
SF	72.87 (27.46)	73.88 (25.16)
RE	77.45 (37.15)	73.92 (37.74)
MH	70.99 (20.01)	73.74 (18.61)
PCS	39.39 (11.62)	40.07 (11.31)
MCS	50.99 (11.08)	51.38 (10.60)

**Arthritis*: All inflammatory and chronic non-inflammatory arthritis except juvenile arthritis and infectious arthritis. This included rheumatoid arthritis, ankylosing spondylitis, psoriatic arthritis, inflammatory polyarthropathies, arthrosis, or chronic osteoarthritic

Diabetes: Both types of insulin-dependent or non-insulin-dependent diabetes (type 1/type 2) with as well as without diabetes-related complications, but excluding juvenile diabetes

Heart Failure: Diagnoses of congestive, left or right heart failure, systolic or diastolic dysfunction, pulmonary edema and cardiac asthma, with or without atherosclerosis

Chronic Obstructive Pulmonary Disease (COPD): This included chronic bronchitis, emphysema, and chronic bronchial asthma

diabetes, or chronic obstructive pulmonary disease as a primary diagnosis who were recruited from 33 primary health-care clinics. The 33 PHC practices were part of different types of organizations and representative of a range of urban, suburban, and rural areas. A definition of the inclusion diagnoses is presented in Table 1. Details of the distribution of participants' characteristics for the full sample were presented in an earlier publication of the MACHRO study [26]. Participants provided informed consent for the MACHRO study, and University of

Montreal's Research and Ethics Committee approved the study (IRB Study Number A06-E40-10B). Patients were interviewed face-to-face at study entry and by phone 12 months later.

Measures

Participants completed the SF-36 Version 1 (SF-36) [28], a generic patient-reported outcome measure with documented reliability and validity across a broad range of patient populations and used in more than 6,000 published studies [29]. The tool includes 36 items for the measurement of eight subscales that are summarized in two composite scores reflecting physical and mental health status. The eight subscales include: physical functioning (Physical Function; 10 items), role limitations due to physical health (Role Physical; 3 items), bodily pain (Bodily Pain; 2 items), general health (General Health; 5 items); vitality (Vitality; 4 items), social functioning (Social Function; 2 items), role limitations due to emotional problems (Role Emotional; 3 items), and mental health (Mental Health; 5 items). Items are measured on a binary, 5-point, or 6-point Likert scales. Though some items are positively worded and others are negatively worded, all items are rescaled such that higher scores on the SF-36 indicate higher functioning. Demographic and medical co-morbidity were collected at study entry, and health-care utilization data at baseline and 12 months from the Quebec provincial health data base (RAMQ) medical service claims, physician, and beneficiary files.

Approaches to response shift detection

Correlated first-order factor model

Prior to evaluating response shift with respect to the second-order factor model, we first examined response with respect to the correlated SF-36 subscales (we refer to this as the correlated factors model of SF-36 subscales). Response shift in the second-order factor model pertains to the relationships among the SF-36 subscales. In the second-order factor model (depicted in Fig. 1), the SF-36 subscales *mediate* the relationships between the items and the corresponding physical and mental health factors. Consequently, response shift in the second-order factor model is confounded with response shift in each of the measurement structures of the SF-36 subscales. It is therefore important to first examine and accommodate any response shift with respect to the SF-36 subscales. To do so, we first examined a latent variable model with all SF-36 subscales specified as correlated latent factors. A no-response shift model was specified with measurement model parameters of each subscale (item thresholds and factor loadings) constrained



Fig. 1 The measurement model used in response shift detection. *Notes:* Circles represent latent variables (common and residual factors). All latent variable variances are fixed as one. Abbreviations: *PF* physical functioning, *RP* role physical, *BP* bodily pain, *GH*

general health, *VT* vitality, *SF* social functioning, *RE* role emotional, *MH* mental health. *Red numbers* Time 1 parameter estimates. *Blue numbers* Time 2 parameter estimates

to be equal over time. For this model, means are fixed at zero at time 1 and free at time 2, the variance of each factor was set to 1 at time 1 and time 2, and error terms of items at time 1 were correlated with error terms of items at time 2. A response shift model was subsequently specified by allowing the measurement model parameters (i.e.,

thresholds and loadings) to vary over time. One threshold of one item per latent factor was held to be equivalent over time for purposes of model identification [30]. Response shift (reprioritization and recalibration) was examined separately for each SF-36 subscale by allowing all loadings and thresholds of that subscale to vary across occasions

while holding the measurement model parameters of the other SF-36 subscales constant over time in the correlated latent factors model. Cross-loadings between the observed items and first-order latent factors were not considered because of the intent to remain consistent with the SF-36 measurement structure (i.e., retain the configuration of the subscales). Therefore, reconceptualization response shift (which would require an exploration of all possible cross-loadings) was not considered in the first-order factor model. A final model was specified to accommodate recalibration and reprioritization response shift in the measurement structures of the SF-36 subscales by allowing the item thresholds and SF-36 subscale latent factor loadings with statistically significant response shift to vary over time (i.e., they were not constrained to be equal over time). This is the comparison model for the second-order factor model.

Second-order factor model

Building on Oort's approach, response shift with respect to the second-order factor model (Fig. 1) was examined by evaluating parameter invariance over time with respect to the relationships between the first-order latent factors (the 8 subscales of the SF-36 modeled as latent factors) and second-order latent factors (the SF-36 physical and mental health components) that correspond with the SF-36 measurement structure (see [31] for specification of second-order factor models). For the no-response shift model, the second-order factors were correlated with variances of 1 at both time points. The first-order factor means were fixed at 0 at time 1 and estimated at time 2, except for one invariant first-order factor mean per second-order factor, which was constrained to be equal over time (for identification purposes). The means of the second-order factors were fixed at 0 at time 1 and estimated at time 2, and the factor loadings were constrained to be equal over time. For the second-order response shift model, the second-order factor loadings were then constrained sequentially to be invariant over time for examination of reprioritization response shift, and cross-loadings were subsequently specified sequentially to examine reconceptualisation response shift. Recalibration response shift was represented by a change in the first-order latent factor means over time. These are the intercepts for the second-order factor model.

Statistical analyses

Data screening, evaluation of variable distributions, and reasons for missing values were assessed using SAS 9.1 software [32]. Mean scores or frequencies were calculated for the demographic, medical co-morbidity, health-care

utilization data, and the SF-36 subscale scores. All latent variable models were estimated using Mplus and weighted least squares mean and variance adjusted (WLSMV) estimation [33, 34]. Pairwise deletion was used to accommodate missing data (11.7 % missing values for SF-36 items). Polychoric and polyserial correlations and probit-link functions were used to accommodate the ordinal and binary distributions of the SF-36 items.

The appropriateness of the latent variable measurement models was assessed using a variety of fit indices. Fit of the models was evaluated using the comparative fit index (CFI), which ranges from 0 to 1, with higher scores (e.g., greater than .95) indicating better fit. We also used the root mean square error of approximation (RMSEA) [35]. A RMSEA value less than .06 indicates a close fit of the model in relation to the degrees of freedom, although this figure is based on subjective judgment [36]. The chi-square statistic is also reported but not relied upon as a criterion for a well-fitting model due to its sensitivity to sample size [23]. Because the chi-square differences based on WLSMV estimation are not distributed as a chi-square, the DIFF-TEST procedure in MPlus was used to assess whether the differences in fit between the nested models were statistically significant (WLSMV χ^2 [37, 38]). Further, to assess the impact of response shift, the latent factor mean scores between the response shift models and no-response shift models were compared. The statistical significance of mean difference scores was examined using the chi-square difference test.

Results

Table 1 summarizes the characteristics of the study population. The average age was 67 and 55 % were female. Overall, 25 % of the sample had more than one diagnosis, 11.8 % had 2 co-morbidities, 2.2 % had 3 co-morbidities, and 1 % had three co-morbidities. Twenty-four percent of individuals had a hospitalization, and approximately 34 % had one emergency room visit over the one-year follow-up period. Table 2 summarizes health-care utilization across each chronic disease group. Among all chronic conditions, individuals with arthritis had the highest frequency of visits to the general practitioner. All other health-care utilization was relatively similar across chronic conditions.

As shown in Figure 2, the SF-36 item distributions are similar at baseline and 12-month follow-up. Descriptive statistics for the SF-36 subscale scores are presented in Table 1 for baseline (study entry) and 12-month follow-up. There was very little missing data in the SF-36 scores at study entry (i.e., data were available on 98–100 % of patients). At 12 months, attrition resulted in approximately 25 % of the sample providing less data on the SF-36.

Table 2 Frequency of health-care utilization by patient group

	No. times in past year	Heart failure	COPD	Arthritis	Diabetes	Likelihood-ratio χ^2 (<i>P</i> value)
Number of consultations with a GP or specialist	0	29	26	45	20	70.73 (.000)
	1	65	74	163	90	
	2	113	96	234	117	
	3	84	78	114	101	
	≥ 4	288	262	293	231	
	Do not Know	2	0	1	3	
	Refused	0	0	0	0	
	Missing	1	1	0	1	
Number of visits to hospital emergency departments	0	393	371	682	434	55.58 (.000)
	1	102	93	112	78	
	2	41	32	32	25	
	3	19	20	14	7	
	≥ 4	24	19	9	18	
	Do not Know	2	1	1	0	
	Refused	0	0	0	0	
	Missing	1	1	0	1	
Number of hospitalizations	0	430	418	758	488	99.95 (.000)
	1	92	78	69	60	
	2	34	26	12	9	
	3	11	8	3	2	
	≥ 4	13	6	4	1	
	Do not Know	1	0	4	1	
	Refused	0	0	0	1	
	Missing	1	1	0	1	

Bolded font shows the cells that were responsible for the significant chi-squared statistic. Despite the larger number of health-care utilization episodes for the arthritis subgroup, we were underpowered to do the SEM analyses just for arthritis or comparing the groups

Assessment of response shift

Correlated first-order factor model of the SF-36 subscales

The unconstrained correlated factors model fit the data well (see model 1 in Table 3). Specification of no-response shift for all latent factors (model 2) led to a statistically significant reduction in model fit, which suggests evidence of response shift. Examination of separate models that accommodate response shift for each SF-36 subscale revealed statistically significant reprioritization or recalibration response shift only for the Physical Health, Vitality, and Mental Health scales. The final model accommodating response shift with respect to these scales and holding all measurement parameters for the other subscales to be invariant (Model 3) fit the data well (RMSEA = .035) and led to a statistically significant improvement in fit relative to the no-response shift model of the correlated first-order factors. While significance was reached, the differences in magnitudes of the standardized parameter estimates that

were allowed to vary over time were very small (see Fig. 1). Thus, there was no evidence of substantial response shift.

Second-order factor model

We evaluated response shift with respect to the physical and mental health components using a second-order factor model (see Fig. 1). The unconstrained second-order factor model fit the data reasonably well (model 4 in Table 3). There was not a statistically significant change in model fit when invariance was imposed on the second-order factor loadings and thresholds (model 5). To examine response shift with respect to each SF-36 subscale, we tested the difference in fit between the no-response shift model and separate models that sequentially free one of the factor loadings (reprioritization) or a cross-loading (reconceptualisation). Statistically significant improvement of fit was observed when the Social Function loading on Physical Health was allowed to vary over time (reprioritization response shift) and when cross-loadings were specified for

Table 3 Relative fit of the response shift analyses models

	Model	WLSMV χ^2 (df)	RMSEA (90 % CI)	CFI	Comparison model	WLSMV χ^2 dif (df, p)
Correlated first-order factor model of SF-36 subscales						
Model 1	Unconstrained Model	4,401.5 (2,191)	.036 (.035–.038)	.97	–	n/a
Model 2	No RS Model (constrained)	4,342.4 (2,327)	.033 (.032–.035)	.97	Model 1	222.8 (136, .000)
Model 3a	RS in Bodily pain ¹	4,332.6 (2,317)	.034 (.032–.035)	.97	Model 1	15.9 (10, .102)
Model 3b	RS in General Health ¹	4,370.8 (2,303)	.034 (.032–.036)	.97	Model 1	35.1 (24, .067)
Model 3c	RS in Physical Function ¹	4,347.8 (2,298)	.034 (.032–.035)	.97	Model 1	52 (29, .005)
Model 3d	RS in Role Emotional ¹	4,347.8 (2,322)	.034 (.032–.035)	.97	Model 1	7.0 (5, .224)
Model 3e	RS in Role Physical ¹	4,353.4 (2,320)	.034 (.032–.035)	.97	Model 1	2.9 (7, .891)
Model 3f	RS in Social Function ¹	4,339.4 (2,318)	.034 (.032–.035)	.97	Model 1	14.2 (9, .116)
Model 3 g	RS in Vitality ¹	4,309.7 (2,304)	.034 (.032–.035)	.97	Model 1	57.9 (23, .000)
Model 3 h	RS in Mental Health ¹	4,352.3 (2,298)	.034 (.032–.036)	.97	Model 1	45.7 (29, .025)
Model 3i	Final RS Model ²	4,330.5 (2,244)	.035 (.033–.036)	.97	Model 2	156.6 (83, .001)
Second-order factor models						
Model 4	Unconstrained Model	4,662.9 (2,328)	.036 (.034–.037)	.96	–	n/a
Model 5	No RS Model (constrained)	4,510.1 (2,339)	.035 (.033–.036)	.97	Model 4	13.1 (11, .29)
Model 6	Final RS Model ³	4,505.9 (2,336)	.035 (.033–.036)	.97	Model 5	14.7 (3, .00)

All analyses are at the item level

df is based on WLSMV estimation

¹ Invariance tests of thresholds and first-order factor loadings in the corresponding subscale

² Accommodates response shift in the latent factors for the physical function, vitality, and mental health subscales

³ Accommodates reprioritization response shift in the loading of social function on physical health and reconceptualization response shift in cross-loadings for the mental health and role emotional subscales (see Fig. 1)

the role emotion subscale at time 1 and the mental health subscale at time 2 (Model 6 in Table 3). However, the difference in parameter estimates for this loading at times 1 and 2 was very small; the largest difference in standardized parameter estimates was for the cross-loading with the emotional health subscale at time 2, which was .27 (Fig. 1; unstandardized parameter estimates for the first- and second-order factors are reported in Table 4). All other second-order loading and cross-loadings were invariant. Thus, there was no evidence of substantial response shift at the second-order factor level, except for a small cross-loading with the mental health subscale, which suggests some evidence of reconceptualization response shift. There was no statistically significant differences in the means of the physical and mental health components across both time points.

Discussion

A small but statistically significant response shift effect was identified in the measurement structures of the SF-36 subscales, whereas no substantial response shift was detected in the second-order factor model. The recalibration response shift effect in the Physical Health, Vitality,

and Mental Health SF-36 subscales did not, however, have much consequence on the interpretation of the change in the factor scores over time. This effect suggests that the magnitude of the loadings of the items on the subscales changed over time, but not enough to change the interpretation of change estimates of these subscales over time. While the impact of response shift was not substantial in this sample, future research should examine the potential item-level and composite-level response shift in a population where response shift is expected to occur. Given that individuals provide responses at an item level, we believe that this more thorough examination of response shift is more precise from a conceptual perspective (i.e., both the first- and second-order factor structures are evaluated). Simulation studies are needed to evaluate the added response shift detection power that can be derived from evaluating invariance in a first-order as well as a second-order factor model.

The Oort approach, initially published in 2005, has now been applied in several patient populations, and the Oort method was further expanded in 2009. The main difference between the modeling approach used in this study as compared to the Oort approach is that in this study, the SF-36 subscales were represented as latent factors. Considering the resulting large number of possible measurement

model parameters, it was not feasible to examine any possible response shift with respect to all of the items in relation to the eight latent factors (the number of possible cross-loadings would have required testing an exorbitant number of models). We therefore assumed that the items would consistently correspond to their respective subscales (i.e., we did not consider any cross-loadings at the item level), and we examined response shift for all items corresponding to one of the SF-36 subscales at the same time. This was repeated for each SF-36 subscale. Thus, using this approach, we accommodated recalibration and reprioritization response shift at the item level with respect to each of the SF-36 subscales. The resulting correlated factors model of the SF-36 subscales, including the tenable constraints, formed the basis for our subsequent analyses of the second-order factor model. The modeling analysis we used in this study can be improved in future work that further develops the Oort approach to apply it to the item-level model that we used in the current study.

Several considerations need to be examined before applying the Oort approach to a second-order model. Because a substantially larger number of parameters are being estimated, consideration is needed as to whether there is a large enough sample. The standard errors obtained in our analyses suggest that the sample was large enough (as evidenced by statistically significant parameter estimates).

This secondary analysis has several strengths as well as limitations. The strengths include a substantial sample size representing four common chronic conditions, many of whom had co-morbidities. Additionally, the sample represents a grouping of 33 primary health-care clinics, so the study findings are likely to be generalizable to the primary care population.

The limitations of the study are that there was a 25 % attrition rate in the data on this generic patient-reported outcome measure and that there were mode differences between study entry (face-to-face interview) and the 12-month follow-up (phone interview). This mode difference could lead to differences in item response, which might look like response shift but not be response shift in the SEM models. Both of these limitations could lead to measurement variance and random error, which could affect the measurement, and structural models being tested by SEM.

There may also be some unmeasured catalysts, which may have acted differently across patients with different chronic disease conditions, such as medication use and exposure to other psychosocial interventions. In this regard, it may not be possible to suggest the presence of a minimal catalyst over the follow-up period of the study. It is also possible that different effects of response shift may have occurred among different chronic diseases and that

sociodemographic factors may interact with catalysts to influence response shift. Thus, the finding of “minimal” response shift might be due to the inclusion of a mixed group of subjects in the analyses. Specifically, some subjects had no hospitalization/emergency room visit while other subjects had hospitalization/emergency room visits. Ideally, we would have examined response shift in each individual chronic disease subgroup; however, in this study, we did not have sufficient power for the model to conduct subgroup analyses.

Future directions

Future work is needed to assess whether response shift effects evaluated among individual chronic disease subgroups yields similar results as when all subgroups with a common catalyst (i.e., a change in health) are combined. Further, additional research is needed to evaluate the influence of duration of disease on response shift effects in the presence of a catalyst.

The results of this study suggest only minimal response shift effects in this primary care sample. We based our hypothesis of minimal response shift on the relatively small number proportion of individuals who required hospitalized or who had a high frequency of health-care visits. It is possible, however, that our ability to detect response shift is limited by the method used. We know that SEM will only detect response shift effects when a substantial proportion of the sample evidences such effects (e.g., due to a common catalyst). In addition, although we did not detect substantial response shift in the overall sample, it is possible that there are subgroups of individuals in the sample who may have experienced response shift, and that there may be potential measured or unmeasured catalysts other than health-care utilization and co-morbidities that could have resulted in a response shift. Future work might use simulation modeling to clarify the proportion of individuals required to have undergone a response shift for such an effect to be detectable at a group level by SEM. Studies that can determine the response shift effect sizes that are clinically relevant will also be an important area of exploration.

We evaluated a two-factor second-order model for the SF-36 based on previous validation of the model. A more targeted model of health-related quality of life based on the Wilson and Cleary model [39], for example, may be more sensitive to response shift detection. Further, we evaluated response shift over 1 year. A longer time period may be needed to detect a clinically meaningful response shift. Future work is needed to identify other potential catalysts and to measure the time period needed to detect a response shift among individuals with chronic conditions.

Acknowledgments This work was funded in part by a Catalyst grant award from the Canadian Institute of Health Research (Grant #103630), and a Career Award (Grant #13870) from the Fond de Recherche en Sante du Quebec to Dr. Ahmed. We thank Brian Quaranto, B.Sc., for assistance with data management and manuscript preparation.

References

- Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 54(4), 297–312.
- Cronbach, L. J., & Furby, L. (1970). How we should measure “change”: Or should we? *Psychological Bulletin*, 74, 68–80.
- Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research*, 13(2), 127–143.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525–543.
- Sprangers, M. A., & Schwartz, C. E. (1999). Integrating response shift into health-related quality of life research: A theoretical model. *Social Science and Medicine*, 48(11), 1507–1515.
- Schwartz, C. E., & Sprangers, M. A. (1999). Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Social Science and Medicine*, 48(11), 1531–1548.
- Schwartz, C. E., & Rapkin, B. D. (2004). Reconsidering the psychometrics of quality of life assessment in light of response shift and appraisal. *Health and Quality of Life Outcomes*, 2, 16.
- Sprangers, M. A., & Schwartz, C. E. (1999). Integrating response shift into health-related quality of life research: A theoretical model. *Social Science and Medicine*, 48(11), 1507–1515.
- Rapkin, B. D., & Schwartz, C. E. (2004). Toward a theoretical model of quality-of-life appraisal: Implications of findings from studies of response shift. *Health and Quality of Life Outcomes*, 2(1), 14.
- Schwartz, C. E., Bode, R., Repucci, N., Becker, J., Sprangers, M. A., & Fayers, P. M. (2006). The clinical significance of adaptation to changing health: A meta-analysis of response shift. *Quality of Life Research*, 15(9), 1533–1550.
- Oort, F. J. (2005). Using structural equation modeling to detect response shifts and true change. *Quality of Life Research*, 14(3), 587–598.
- King-Kallimanis, B., Oort, F., & Garst, G. (2010). Using structural equation modelling to detect measurement bias and response shift in longitudinal data. *ASSTA Advances in Statistical Analysis*, 94(2), 139–156.
- Visser, M. R., Oort, F. J., & Sprangers, M. A. (2005). Methods to detect response shift in quality of life data: A convergent validity study. *Quality of Life Research*, 14(3), 629–639.
- Schwartz, C. E., Sprangers, M. A., Oort, F. J., Ahmed, S., Bode, R., Li, Y., et al. (2011). Response shift in patients with multiple sclerosis: An application of three statistical techniques. *Quality of Life Research*, 20(10), 1561–1572.
- Schwartz, C. E., Sprangers, M. A. G., Carey, A., & Reed, G. (2004). Exploring response shift in longitudinal data. *Psychology and Health*, 19(1), 51–69.
- Hakkennes, S. J., Brock, K., & Hill, K. D. (2011). Selection for inpatient rehabilitation after acute stroke: A systematic review of the literature. *Archives of Physical Medicine and Rehabilitation*, 92(12), 2057–2070.
- Ahmed, S., Bourbeau, J., Maltais, F., & Mansour, A. (2009). The Oort structural equation modeling approach detected a response shift after a COPD self-management program not detected by the Schmitt technique. *Journal of Clinical Epidemiology*, 62, 1165–1172.
- Razmjou, H., Schwartz, C. E., & Holtby, R. (2010). Recalibration response shift had an independent impact on perceived disability two years following rotator cuff surgery. *Journal of Bone and Joint Surgery*, 92, 2178–2186.
- Razmjou, H., Schwartz, C. E., Yee, A., & Finkelstein, J. A. (2009). Traditional assessment of health outcome following total knee arthroplasty was confounded by response shift phenomenon. *Journal of Clinical Epidemiology*, 62, 91–96.
- Razmjou, H., Yee, A., Ford, M., & Finkelstein, J. A. (2006). Response shift in outcome assessment in patients undergoing total knee arthroplasty. *Journal of Bone and Joint Surgery*, 88(12), 2590–2595.
- Finkelstein, J. A., Razmjou, H., & Schwartz, C. E. (2009). Response shift and outcome assessment in orthopedic surgery: Is there a difference between complete vs. partial treatment? *Journal of Clinical Epidemiology*, 62, 1189–1190.
- Ahmed, S., Mayo, N. E., Corbiere, M., Wood-Dauphinee, S., Hanley, J., & Cohen, R. (2005). Change in quality of life of people with stroke over time: True change or response shift? *Quality of Life Research*, 14(3), 611–627.
- Schumacker, R. E., & Lomax, R. G. (2004). *A Beginner's Guide to Structural Equation Modeling, Second Edition* (2nd ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates Inc.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4–70.
- Sholom, G. (2012). PROMs: A Critical Step. *But Only One of Many. HealthcarePapers*, 11(4), 29–33.
- Lemieux, V., Levesque, J. F., & Ehrmann-Feldman, D. (2011). Are primary healthcare organizational attributes associated with patient self-efficacy for managing chronic disease? *Health Policy*, 6(4), e89–e105.
- Feldman, D. E., Bernatsky, S., Levesque, J. F., Van, M. T., Houde, M., & April, K. T. (2010). Access and perceived need for physical and occupational therapy in chronic arthritis. *Disability and Rehabilitation*, 32(22), 1827–1832.
- Ware, J. E., Jr, Kosinski, M., & Keller, S. D. (1994). *SF-36 physical & mental scales: A user's manual*. Boston, Massachusetts: The Health Institute, New England Medical Center.
- McHorney, C. A., Ware, J. E., Jr, & Raczek, A. E. (1993). The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Medical Care*, 31(3), 247–263.
- Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research*, 39(3), 479–515.
- Byrne, B. M. (2012). *Structural equation modeling with Mplus : basic concepts, applications, and programming*. New York: Routledge Academic.
- Inc, S. A. S. I. (2009). *SAS/STAT 9.2 User's Guide*. Cary, NC: SAS Institute Inc.
- Muthén, L. K., & Muthén, B. O. (1998–2010). *Mplus user's guide* (Sixth Edition ed). Los Angeles, CA: Muthén & Muthén.
- Muthén, B., & Muthén, L. (2011). *MPlus (version 6.2)*. Los Angeles, CA: Statmodel.
- Finney, S. J., & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. In I. G. R. H. R. O. Mueller (Ed.), *Structural equation modeling: A second course* (pp. 269–314). Greenwich, CT: Information Age Publishing.
- Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling*, 13, 186–203.

37. Nussbeck, F. W., Eid, M., & Lischetzke, T. (2006). Analysing multitrait–multimethod data with structural equation models for ordinal variables applying the WLSMV estimator: What sample size is needed for valid results? *British Journal of Mathematical and Statistical Psychology*, 59(1), 195–213.
38. Asparouhov, T., & Muthén, B. (2006). Robust Chi Square difference testing with mean and variance adjusted test statistics. Mplus Web Notes: No. 10. <http://www.statmodel.com/download/webnotes/webnote10.pdf>.
39. Wilson, I. B., & Cleary, P. D. (1995). Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA*, 273(1), 59–65.