

Guidelines for secondary analysis in search of response shift

Carolyn E. Schwartz · Sara Ahmed · Richard Sawatzky · Tolulope Sajobi ·
Nancy Mayo · Joel Finkelstein · Lisa Lix · Mathilde G. E. Verdam ·
Frans J. Oort · Mirjam A. G. Sprangers

Accepted: 22 March 2013 / Published online: 10 April 2013
© Springer Science+Business Media Dordrecht 2013

Abstract

Objective Response shift methods have developed substantially in the past decade, with a notable emphasis on model-based methods for response shift detection that are appropriate for the analysis of existing data sets. These secondary data analyses have yielded useful insights and motivated the continued growth of response shift methods. However, there are also challenges inherent to the successful use of secondary analysis for response shift detection. Based on our experience with a number of secondary analyses, we propose guidelines for the optimal implementation of secondary analysis for detecting response shift.

Methods We review the definition of response shift and recent advances in response shift theory. We describe current statistical methods that have been developed for or applied to response shift detection. We then discuss lessons

learned when using these methods to test specific hypotheses about response shift in existing data and of the features of a data set that could guide early decision-making about undertaking a secondary analysis.

Results A checklist is provided that includes guidelines for secondary analyses focusing on: (1) selecting an appropriate data set to investigate response shift; (2) prerequisites of data sets and their preparation for analysis; (3) managing missing data; (4) confirming that the data fit the requirements and assumptions of the selected response shift detection technique; (5) model fit evaluation; (6) interpreting results/response shift effect sizes; and (7) comparing findings across methods.

Conclusions The guidelines-checklist has the potential to stimulate rigorous and replicable research using existing data sets and to assist investigators in assessing the

C. E. Schwartz (✉)
DeltaQuest Foundation, Inc., 31 Mitchell Road, Concord,
MA 01742, USA
e-mail: carolyn.schwartz@deltaquest.org

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

S. Ahmed · N. Mayo
Faculty of Medicine, School of Physical Therapy,
McGill University, 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

T. Sajobi
Department of Community Health Sciences,
University of Calgary, Calgary, Canada

J. Finkelstein
Division of Orthopaedics, Sunnybrook Health Sciences Center,
The University of Toronto, Toronto, ON, Canada

L. Lix
Department of Community Health Sciences, University of
Manitoba, Winnipeg, MB, Canada

M. G. E. Verdam · M. A. G. Sprangers
Department of Medical Psychology, Academic Medical Center,
University of Amsterdam, Amsterdam, The Netherlands

M. G. E. Verdam · F. J. Oort
Department of Education, University of Amsterdam,
Amsterdam, The Netherlands

appropriateness and potential of a data set and model-based methods for response shift research.

Keywords Response shift · Analytic · Methods · Guidelines

Introduction

It is increasingly acknowledged that the process of adaptation affects the measurement of evaluative outcomes. In the field of health-related quality of life (QOL), these ‘response shifts’ are thought to be triggered by health state changes (catalyst) and result in changes in internal standards (recalibration), values (reprioritization), or meaning (reconceptualization) of the construct being measured [1, 2]. The past decade has witnessed an increasing number of empirical papers on response shift in the context of QOL research. With this expansion, the field of health-related QOL and patient-reported outcomes research has benefited from insights into how individuals experience change in health over time. Response shift has been studied and recognized in patients with multiple sclerosis [3], cancer [4–7], stroke [8, 9], diabetes [10, 11], dental disorders [12] and in the fields of geriatric medicine [13–15], palliative care [16–19], and orthopedics [20, 21]. Response shift effects have been documented to range from small to medium in size [22]. Even when effect sizes are small, however, the impact of adjusting for response shifts has been to increase intervention effect sizes from small to moderate [23], rendering them clinically significant [24].

Background on response shift

Definition

The term ‘response shift’ was coined by Howard and colleagues and originates from research on educational training interventions [25]. They defined response shift in terms of changes in internal standards of measurement. Parallel in time but independently, Golembiewski et al. [26] forwarded a change typology originating from management sciences. They distinguished observable change (alpha change) from changes in internal standards (beta change) and meaning (gamma change) [26]. Changes in values were inherent to Golembiewski’s description of reconceptualization, but were not explicitly seen as a separate component. These concepts led to the *working definition* proposed by Sprangers and Schwartz [27]:

Response shift refers to 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 (scale recalibration); b) a change in the respondent’s values (reprioritization); or c) a redefinition of the target construct (reconceptualization).

Reprioritization is thus an added explicit component of this definition. This is particularly relevant to the field of health-related QOL, where many of the measurement instruments are multidimensional. Reprioritization relates to a change in the relative importance of different health-related QOL dimensions.

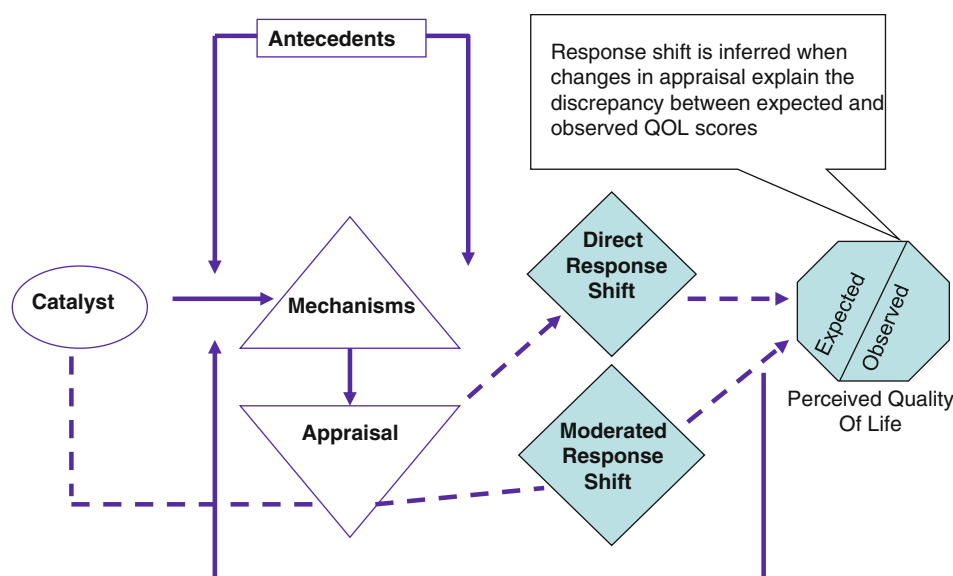
Theoretical model

This working definition was integrated into a theoretical model that relates health changes to one’s self-evaluation of health or QOL [27]. In this model, changes in health status (*catalysts*) may induce response shift via *mechanisms* to accommodate this change. Mechanisms refer to cognitive, affective, and behavioral strategies to deal with the catalyst, such as engaging in a special type of social comparison or seeking social support. The kinds of mechanisms individuals may use depend not only on the specific health change catalyst, but also on stable characteristics of the person, the so-called *antecedents*, such as gender and personality traits. Antecedents would work directly to increase or decrease the likelihood or strength of catalysts and/or mechanisms. Rather than being measured directly, response shift is inferred when changes in *appraisal* explain discrepant (e.g., residual) changes in QOL scores that cannot be explained by the ‘standard QOL model’ [28] (see Fig. 1). For more information on the theoretical models and discourse on response shift, see [27–30]. Response shift is thus conceptualized as an *effect* rather than as a *construct* one measures directly. This distinction has implications for how researchers would operationalize it and report their findings, for example, response shift is not ‘adaptation’, but rather an effect of adaptation.

Response shift methods

A number of methods for measuring response shift have been developed and studied in the past decade. These methods range from approaches that require alternative data collection methods [28, 31–33] to statistical methods that can be applied to both primary and secondary data sets. A growing evidence base supports the convergence but also divergence among response shift detection methods [34–36]. Current statistical methods represent notable improvements from the factor analytic approaches first utilized in the education- and management-research contexts of the early days of response shift research [26, 37].

Fig. 1 Theoretical model of response shift and quality of life. Adapted from Sprangers and Schwartz [1] (with permission from Elsevier) and Rapkin and Schwartz [28] models of appraisal and quality of life



Capitalizing on the recent advances in computational science, there are a number of accessible and sophisticated methods that can be used to assess response shift without requiring additional data collection. For example, clinical trial or patient registry data can now be used or made amenable for response shift research. Sometimes secondary analysis may also be useful in response shift studies when a new method allows a different way of analyzing the data, for example to assess another aspect of response shift. Table 1 summarizes the group- and individual-based analytical approaches that are currently available for response shift detection. We briefly describe each of these methods below and reference primary sources for the interested reader.

Data mining

Recent advances in computational science have led to a host of methods aimed at discovering patterns in large data

sets. One such method, *Recursive Partitioning and Regression Tree* method [38], uses a non-parametric statistical index to iteratively separate respondents into increasingly homogeneous subgroups. It has been applied in response shift research [33, 39], as well as numerous other problems of classifications [40–43]. This method can result in multiple pathways leading to a similar QOL response shift pattern and to finding that the same variable plays a different role in different contexts. The Recursive Partitioning Regression Tree produces a set of logical if-then conditions for predicting the group memberships of a study sample. The method has been applied to response shift research in the context of having QOL Appraisal Profile data [33] or not having data on appraisal and inferring differences in appraisal [44, 45]. The method seeks to maximize explained variance in identifying homogeneous subgroups of patients. Response shift may be defined in this approach to be a change score in physical functioning or disability that is accompanied by stability of

Table 1 Response shift detection methods useful for secondary data analysis

Method	Focus of analysis	Aspect(s) of response shift in focus			No. time points required
		Recalibration	Reprioritization	Reconceptualization	
Data mining					
Recursive partitioning and regression tree modeling	Individual	✓	✓	✓	2
Regression-based methods					
Relative importance analysis	Group		✓		2
Latent variable models					
Structural equation modeling	Group	✓	✓	✓	2
Latent class growth curve models	Group	✓	✓	✓	3+
Latent variable mixture models	Group	✓	✓	✓	2
Latent trajectory of residuals	Individual		✓	✓	4

mental health functioning [33]. Because this method seeks to maximize explained variance, the results of an analysis may be sample-specific and have limited external validity. It is thus necessary to implement the analysis on two independent samples—‘test’ and ‘validate’ samples—to confirm the robustness of the findings.

Regression-based methods

Response shift methods have also been developed based on changes in the relative importance of domains of a validated patient-reported outcomes measure using logistic regression and discriminant analysis models. Lix and colleagues applied the *Relative Importance Analysis* method to detect reprioritization response shift in longitudinal data. The method utilizes the relative importance weights and ranks derived from measures of relative importance based on logistic regression and discriminant analysis to quantify the importance of each domain at each occasion [46, 84]. In logistic regression model, the group-membership variable that divides the study sample into two known groups related to health state change (i.e., catalyst for response shift) is the dependent variable, while independent variables are the domain or subscale scores. In contrast, the domains are the dependent variables in discriminant analysis model, while group-membership variable is the independent variable. *Reprioritization response shift* is considered present in a domain if the change over time in signed values of the relative importance weights or ranks on the domain is statistically significant at some prespecified α level [46].

Latent variable models

Structural Equation Modeling (SEM) combines factor analysis and regression analysis [23, 47–56]. By analyzing covariances and means matrices, it is possible to first test the assumption of measurement invariance and subsequently test substantive hypotheses. This method involves iterative model-testing that systematically loosens model constraints over time to identify how and where the measurement model varies over time [23, 48]. Different parameters in the measurement model (i.e., factor structure) are indicative of different aspects of response shift. This approach extended earlier work done by Schmidt [49] and yielded more sensitive algorithms for detecting response shifts. The SEM approach can also be used to examine explanatory and biasing variables associated with response shift [50]. To date, the Oort SEM method has been applied to cancer [23], stroke [51, 52], musculoskeletal diseases [53], chronic obstructive pulmonary disease [54], multiple sclerosis [55], and primary care samples [56]. See Oort [48] for details. This method is sensitive to

response shifts only when experienced by a majority of the sample [54, 57]. Since preliminary estimates of the prevalence of response shift suggest that about one-half to one-third of respondents exhibit response shifts that are detectable by these methods [21, 58], one would have to over-sample people prone to response shifts to be able to detect such change using SEM. Over-sampling will be feasible when we are better able to predict who experiences response shifts.

Other response shift detection methods that are outgrowths of regression and SEM modeling include latent class *growth curve* models [59], *latent variable mixture* models [60, 61], and *latent trajectory of residuals* models [58]. Growth curve analysis is looking at a group-level change in latent constructs over time. Latent class models separate a growth model into different latent subgroups in the population. Latent variable mixture models are an extension of the SEM approach that allows for the examination of measurement invariance over time for different latent subgroups in the population. Latent trajectory of residuals separates subjects based on the differences between observed and expected scores, as derived from mixed effects regression modeling [58, 62]. It is noteworthy that, except for the latent variable mixture models, the above approaches require longitudinal data with more than two time points. In addition, like other complex latent variable methods, these approaches require a substantial sample size.

A general caveat Regardless of the response shift detection approach, the investigator is strongly advised to adhere to the following guiding principles: always base the response shift analysis on theory-driven hypothesis testing (e.g., by using a comparison/control group); have clearly stated hypotheses about when the response shift will occur (catalyst and change); use a combination of approaches to provide information about convergence among methods; and include an objective clinical criterion measure so that it is possible to distinguish between expected and observed change in QOL over time.

Lessons learned: guidelines for secondary analysis

Secondary data analyses related to response shift research are generally initiated because specific hypotheses are generated by theory or previous empirical work. Utilizing existing data is a cost-effective way to test such hypotheses. This practice is growing in prevalence not only because research funding is increasingly difficult to procure, but also because recent innovations in quantitative methods require large samples that would be cost-prohibitive to collect from scratch. There are, however, caveats

and challenges to such a path. We encountered such caveats and learned some important lessons that we believe will be useful to other investigators. We will describe these lessons and provide a brief checklist with recommendations to facilitate and stimulate future secondary analyses for response shift research. These recommendations are provided in the order in which a researcher will likely need them. It should be noted that some of the issues raised below are also relevant to response shift detection in the context of primary data collection.

Lesson #1: Selecting an appropriate data set to investigate response shift

There are several prerequisites to making it interesting or worthwhile to check for response shift. First, there must be changes relevant to the outcome of interest (i.e., *catalyst*). In the case of health-related QOL outcomes, the catalyst would be a health state change (e.g., onset of stroke or malignancy); or initiation of treatment for a disorder (e.g., surgical, psychosocial, or pharmaceutical interventions). Even caring for someone with an illness could be a powerful catalyst. Other catalysts could involve novel parts of the assessment procedure (test of exercise capacity, asking novel questions about mood or memory). The mere passage of time might not be. For psychological well-being outcomes, the catalyst might be a psychosocial or rehabilitation intervention. Importantly, these catalysts should be salient and induce changes that are of such magnitude and duration that most patients are required to adapt to this new state.

With secondary data, the ability of the researcher to investigate response shift is dependent on the first measurement occurring before or very closely following the catalyst and having at least one other measurement that occurs sufficiently far after the catalyst for response shift effects to be observed. The relevant observation period will depend on the nature of the catalyst and of the disease/condition in question. There are clinical research standards within clinical populations that could be used as a reference point for response shift analyses. For example, in multiple sclerosis research, a follow-up period of 3 years is recommended to detect clinically important change in imaging outcomes [63, 64]; in spine surgery, the recommended follow-up for lasting and important clinical change is 2 years [65]. Research done by members of our group has documented that more acute catalysts, such as surgery, have shorter-term response shift effects [66], which is the first documentation to our knowledge of the transience of response shift effects for acute catalysts. Such information would be useful for planning future response shift studies.

Nonetheless, there are two important caveats. First, it is possible that some people have been exposed to a catalyst

whereas others have not. A straightforward response shift analysis assumes that the catalyst is applicable to all people in the sample. If possible, it would be advisable to check this assumption, perhaps using multi-group or multi-level methods to isolate people who have not been exposed to, or who have not responded to, a specified catalyst. Second, changes in health status may have occurred before an investigator has acquired some baseline measure, so that more realistically identifiable catalytic events will be limited to post-illness onset. This will limit the types of data an investigator may examine; (e.g., the onset of chemotherapy, surgery, and other such interventions.).

A second prerequisite is that studies need to have sufficient follow-up and multiple time points to capture changes in internal standards, values, and conceptualization. For some methods, such as growth curves or latent trajectory analysis [67], having a sufficient number of time points (i.e., four [58]) is an important consideration to identify changes that differ from what is predicted. For others, where only two time points are required and/or needed (e.g., structural equation modeling [48, 68]), the window between time points is an important consideration. While these prerequisites seem straightforward, in practice, it is less easy to know if or when such a catalytic event took place. What we do know is that the mere passage of time or minor changes in patient-reported outcomes without supporting evidence from clinical indicators may not yield clinically relevant or important changes in outcome, let alone large enough for response shift effects to be detectable with group-level methods (e.g., structural equation modeling [48, 68]). In such cases, we would not advise embarking on a secondary analysis.

Third, response shift only affects and is detectable by evaluative measures [28, 29]. Such measures explicitly seek input on subjective, internal experiences. There is no ‘right’ answer. Secondary analyses will be more impactful if specific hypotheses are formulated at the outset regarding the direction of and domains in which response shift will occur. Such hypotheses are also important for subjective decisions during the process of data analysis.

Finally, the outcome must be relevant for response shift detection. Indices composed of multiple items may have response shift differentially at the item level masking total score effects. Global indicators may be highly susceptible to response shift, but may lack face validity. They also may not have a normal distribution, which may invalidate the use of methods that assume a normal distribution of responses. Other models may need to be considered. For example, a continuous scale may be optimal for some analyses but only a categorical scale may be available. Other analysis may require a grouping variable for discrimination purposes, and the distribution of this variable may not be conducive for this purpose (e.g., too few

subjects in one group). At the very least, knowing that the outcome is a continuous composite index, continuous single item, or ordinal single item would be crucial to evaluate the relevance of the data set.

Lesson #2: Prerequisites of data sets and preparation thereof

We highlight two distinct types of responsibilities on which good secondary data analysis relies. First, it relies on the primary researcher clearly documenting the data set, leaving a useful and complete trail on which the secondary analyst can build. Second, it relies on the secondary analyst being as knowledgeable as possible about the data set and fully documenting his/her analyses.

Data set documentation

Particularly for secondary analysis, it is key to have clear documentation about the data set. Specifically, the codebook should include names of variables, the full text of the items, the response options and how they are coded, and how missing data are coded. It is critical to have clear documentation of the instrument version and scoring algorithms. Whereas these caveats sound self-evident, we have had several experiences where such basic information was lacking and led to problems in model estimation. For example, a structural equation model would not converge (i.e., it was not possible to estimate the model parameters), and we learned that it was due to what looked like missing data but was actually due to using a scoring algorithm for the wrong version of the questionnaire; the codebook had not specified the version used. It is crucial to have documentation about the values or cut-points that are considered meaningful both on the outcome variable and on the potential predictors. For example, if binary indicator variables will facilitate interpretation of data, the cut-points for the different levels need to be determined prior to beginning the analysis (rather than being based on data snooping) and consistent across analyses.

Data screening

This critical step in data analysis includes the inspection of univariate descriptive statistics, such as invalid or implausible values, means and standard deviations (Are they plausible?), and checking for outliers. It is also important to examine the score distribution (Is it normal? Skewed?) The next step is to deal with non-normal variables, either by conducting linear transformations or by choosing statistical techniques that are robust to deviations from normality. These issues are important to consider before beginning inferential data analysis, because the

subjective decisions on how to handle these issues can affect the choice of response shift detection method or even its outcome. See Tabachnick and Fidell [69] for more detail.

Lesson #3: Managing missing data

Missing data can result in imprecision or bias in estimates of change over time due to loss of data or a non-random pattern of missingness (i.e., data are missing not at random [MNAR]). Both the quantity and pattern of missing data are important considerations in choosing a response shift method and should be assessed through descriptive analyses and/or inferential techniques [70–72]. For example, analyses that compare participants with complete data to participants with missing data on baseline demographic characteristics, covariates, or patient-reported outcomes may facilitate decision-making. Several statistical approaches have been suggested for handling missing data in secondary longitudinal data [70] when it is plausible to assume that the pattern of missingness is ignorable. These approaches include expectation–maximization, full-information maximum likelihood imputation, and multiple imputation methods. If the pattern of missingness is not ignorable, then more sophisticated model-based methods may be appropriate, although they are more challenging to implement in conjunction with response shift methods; some examples include pattern mixture and selection models [73]. We recommend that more than one method to address missing data be considered as part of a sensitivity analysis, to assess whether they result in consistent conclusions about response shift [74]. If not, it may be advisable to report these differences in results, to alert the reader to the potential effects of missing data.

Lesson #4: Confirming that the data fit the requirements of the selected response shift technique

The data set must include a sufficient number of respondents to support the advanced multivariate methods that can detect response shift. These sample size requirements depend on a number of parameters (e.g., size of the model, size of the factor loadings, strength of the relationship of the items, data distribution, and parameter estimation procedure), but as a rule of thumb should include 200–300 cases for relatively small structural equation models [75–78] to a minimum of 500 for classification and regression trees [38]. When complex modeling strategies are employed, it is recommended to conduct a simulation study to assess statistical power to detect response shift effects [79]. Desired subgroup analyses to perform internal validation [80], to compare respondents with different response shift trajectories [81], or to assess the impact of antecedents (e.g., disease severity)

would lead to an even larger minimum sample size requirement.

Another important step involves checking the statistical assumptions of the method that is being used for response shift detection [69]. Failure to satisfy assumptions could result in invalid conclusions about response shift.

Lesson #5: Model fit evaluation

Many response shift analyses entail a number of both small and large subjective decisions in model selection and validation throughout the analysis (e.g., model specification, sequential freeing, or constraining of model parameters). Generally speaking, subjective decisions are a necessary part of any statistical analysis and should be based on both substantive and statistical arguments. For example, when applying the Oort structural equation modeling method for response shift detection [48], it is recommended to document the theoretical and statistical reasons for specific analytic decisions [56]. Further, internal cross-validation or external validation of the model can help to ensure that the conclusions about response shift are not an artifact of the data [80]. Replicating such analyses would be facilitated if these details as well as specific statistical analysis codes were provided as transparently as possible, for example in online supplements.

Related to the issue of subjective decisions, there is a risk in secondary analyses of producing false positives—a response shift is observed but is a statistical artifact. This can occur because of the selective use of the data (e.g., selecting specific groups to study but not others) or multiple tests of significance without control of the overall (i.e., family-wise) error rate. External validation of the finding of a response shift effect(s) is recommended to avoid the potential for the literature to be filled with non-replicable results.

Lesson #6: Interpreting results/response shift effect sizes

To provide an indication of the magnitude of detected response shift effects, effect sizes would need to be reported. For some methods, calculating effect sizes is relatively straightforward (e.g., structural equation modeling [23, 48, 54, 55], growth curve analysis [82], latent trajectory analysis [58, 83]), whereas for others, it needs to be developed (e.g., recursive partitioning and regression trees [33, 39]). This reporting would also facilitate meta-analyses across studies and methods.

The effect sizes should correspond to the analytical objectives of the project, and different effect size statistics could be used to describe different forms of response shift. For example, if the objective is to evaluate measurement bias, then an effect size that compares the results of models

that accommodate and ignore response shift is appropriate [48]. In contrast, if the objective is to evaluate change in values (reprioritization) or meaning (reconceptualization), then effect sizes such as relative importance measures (e.g., the Pratt index) may be more appropriate [46, 84]. This reporting would also facilitate meta-analyses across studies and methods.

Lesson #7: Comparing findings across methods

If the data set is deemed useful for comparing findings across different response shift detection methods, the investigator needs to recognize that this comparison is not straightforward. It is hindered by the fact that the different methods measure different aspects of response shift (i.e., recalibration, reprioritization, and reconceptualization), test different hypotheses about response shift, and may also use different subsets of the data [36]. These inter-method differences lead to variations in the methods' unit of analysis (i.e., group versus individual) as well as their ability to deal with data structure deficits (e.g., missing data, low prevalence of response shift, inability to use more than two time points) [36]. Despite these hindrances, the advantages of comparing different methods on the same data set are numerous and are worth the effort. At the very least, they may provide some assurance that the results are not completely model-dependent (if different methods produce similar results).

Another related issue to comparing findings both across and within methods is heterogeneity in response shift effects in study populations. That is, response shift may not affect all sub-populations equally. Further, response shifts can have different effects on the measurement of outcome (e.g., 'positive' and 'negative' response shift subgroups [81]). In group-level analysis, these different effects can 'cancel' each other out. This heterogeneity may mask the detection of response shift [67].

Discussion

It is our intention that this brief discussion will help future researchers in their usage of secondary analyses to provide better insight when investigating response shift phenomena in various settings, populations, and practices. These guidelines will not only improve the quality of future research, but will also likely work to improve our understanding of the prevalence of response shift. Table 2 provides a brief checklist based on the above lessons learned (Table 2).

How to use this checklist

Generally speaking, one would like to be able to check off all of the boxes in the checklist prior to engaging in a

Table 2 Checklist for secondary data analysis to detect response shift**Lesson #1: Selecting an appropriate dataset to investigate response shift**

- ☐ Is the catalyst salient enough to induce response shift?
- ☐ Is the first assessment administered prior to occurrence of catalyst?
- ☐ Is the observation period of sufficient length to reveal response shift?
- ☐ Is there a subgroup who has not been exposed to or has not responded to the catalyst?
- ☐ Is there a sufficient number of time points for the intended analytical technique?
- ☐ Are the outcomes evaluative in nature?
- ☐ Can you formulate hypotheses regarding direction of and domains in which response shift will occur?
- ☐ Does the data set contain the types of variables required for the desired analysis (e.g., grouping variable, continuous index, original items, etc.).

Lesson #2: Prerequisites of data sets and preparation thereof

Responsibility of primary researcher: Does the initial codebook include:

- ☐ Names of variables?
- ☐ The full text of the items?
- ☐ The response options and how they are coded?
- ☐ How are missing data coded?
- ☐ Instrument(s) version and scoring algorithms are specified?
- ☐ Does the data set have documentation about the values or cut-points that are considered meaningful both on the outcome variable and on the potential predictors?

Responsibility of secondary researcher: Have you implemented standard data screening steps?

- ☐ Have you inspected univariate descriptive statistics for accuracy of input (e.g., out-of-range values, plausible means and standard deviations, univariate outliers)?
- ☐ Have you checked pairwise plots for nonlinearity and heteroscedasticity?
- ☐ Have you identified and dealt with non-normal variables and univariate outliers?
 - a. Check skewness and kurtosis, probability plots
 - b. Transform variables (if desirable)

Lesson #3: Managing missing data

- ☐ Are there missing data?
 - ☐ What proportion of the dataset is missing?
- What are the reasons for missing data: are they ignorable or non-ignorable?
- ☐ Can you determine the pattern or mechanism of missing data from this dataset?
 - ☐ Do you use missing data methods (imputation techniques) consistent with pattern and reasons of missingness?
 - ☐ Have you performed a sensitivity analysis to evaluate and compare the results of different missing data methods?

Lesson #4: Confirming that the data fit the requirements of the response shift detection technique

- ☐ Is the sample size sufficiently large for the planned analytic method?
- ☐ Have you checked the data regarding assumptions of the selected method?

Lesson #5: Model fit evaluation

- ☐ Do you provide an online supplement with the following details:
 - ☐ Subjective decisions made (e.g., model specifications decision about sequential freeing or constraining of model parameters in structural equation model? Pruning thresholds in recursive partitioning and regression tree? Minimal important difference for characterizing patient groups in latent trajectory analysis?
- ☐ Does internal cross-validation of model fit lead to consistent results across subgroups?
- ☐ Software code for programming response shift detection analyses?

Lesson #6: Interpreting results/response shift effect sizes

- ☐ Do you provide effect size(s) in table(s)?
- ☐ Have you selected effect sizes consistent with the analytical objective?

Lesson #7: Comparing findings across methods

- ☐ Can you compare different response shift methods in the same dataset?
- ☐ Do you mention caveats in comparing findings across methods (e.g., does response shift affect subgroups differently in magnitude and direction)?

secondary analysis for response shift research. Given ‘real-world’ limitations, however, it might be worthwhile to utilize the checklist as a proxy for a ‘quality score’, similar to what is used in meta-analyses [22]. The summary score reflects the caliber of the data set to yield meaningful and interpretive response shift findings. However, not all items in the checklist will carry equal weight. For example, if a data set does not ‘score’ well on the first lesson’s dimension (i.e., appropriateness to detect a response shift), then it is not advisable to proceed, regardless of the ‘score’ on the subsequent lessons’ dimensions. In contrast, failing to meet the criteria for lessons 2–7 weakens but does not invalidate the potential of the secondary analysis. Finally, in situations where some of the recommendations are not feasible, the checklist can be used to identify specific limitations when reporting on the results of a response shift analysis.

Utilizing a guidelines-checklist has the potential to stimulate rigorous and replicable research using secondary analysis, assist investigators in assessing the appropriateness and potential of a data set, and facilitate replication of model-based methods for use in response shift research. More consistent standards for implementing and reporting methods and findings will eventually yield a better understanding of the prevalence of response shift.

Acknowledgments Ideas from this manuscript were previously presented as part of a symposium presentation at the International Society for Quality of Life (ISOQOL) in October 2012, in Budapest, Hungary. This work grew out of collaborations among members of the ISOQOL Response Shift Special Interest Group and 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. Drs. Sawatzky, Sajobi, Mayo, and Lix are supported by an operating grant from the Canadian Institutes of Health Research. Dr. Lisa Lix is supported by a Manitoba Research Chair. We are grateful for assistance with manuscript preparation from Brian R. Quaranto, B.S.

References

1. 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.
2. 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.
3. Evers, K. J., & Karnilowicz, W. (1996). Patient attitude as a function of disease state in multiple sclerosis. *Social Science and Medicine*, 43(8), 1245–1251.
4. Jansen, S. J., Stiggelbout, A. M., Nooij, M. A., Noordijk, E. M., & Kievit, J. (2000). Response shift in quality of life measurement in early-stage breast cancer patients undergoing radiotherapy. *Quality of Life Research*, 9(6), 603–615.
5. Bernhard, J., Hurry, C., Maibach, R., Herrmann, R., & Laffer, U. (1999). Quality of life as subjective experience: Reframing of perception in patients with colon cancer undergoing radical resection with or without adjuvant chemotherapy. Swiss Group for Clinical Cancer Research (SAKK). *Annals of Oncology*, 10(7), 775–782.
6. Chapman, G. B., Elstein, A. S., Kuzel, T. M., Sharifi, R., Nadler, R. B., Andrews, A., et al. (1998). Prostate cancer patients’ utilities for health states: how it looks depends on where you stand. *Medical Decision Making*, 18(3), 278–286.
7. Hagedoorn, M., Sneeuw, K. C., & Aaronson, N. K. (2002). Changes in physical functioning and quality of life in patients with cancer: Response shift and relative evaluation of one’s condition. *Journal of Clinical Epidemiology*, 55(2), 176–183.
8. Ahmed, S., Mayo, N. E., Wood-Dauphinee, S., Hanley, J. A., & Cohen, S. R. (2004). Response shift influenced estimates of change in health-related quality of life poststroke. *Journal of Clinical Epidemiology*, 57(6), 561–570.
9. Ahmed, S., Mayo, N. E., Wood-Dauphinee, S., Hanley, J. A., & Cohen, S. R. (2005). The structural equation modeling technique did not show a response shift, contrary to the results of the then test and the individualized approaches. *Journal of Clinical Epidemiology*, 58(11), 1125–1133.
10. Wikby, A., Stenstrom, U., Hornquist, J. O., & Andersson, P. O. (1993). Coping behaviour and degree of discrepancy between retrospective and prospective self-ratings of change in quality of life in type 1 diabetes mellitus. *Diabetic Medicine*, 10(9), 851–854.
11. Postulart, D., & Adang, E. M. (2000). Response shift and adaptation in chronically ill patients. *Medical Decision Making*, 20(2), 186–193.
12. Ring, L. H. S., Heuston, F., Harris, D., & O’Boyle, C. A. (2005). Response shift masks the treatment impact on patient reported outcomes (PROs): The example of individual quality of life in edentulous patients. *Health & Quality of Life Outcomes*, 3, 55.
13. Daltroy, L. H., Larson, M. G., Eaton, H. M., Phillips, C. B., & Liang, M. H. (1999). Discrepancies between self-reported and observed physical function in the elderly: The influence of response shift and other factors. *Social Science and Medicine*, 48(11), 1549–1561.
14. Heidrich, S. M., & Ryff, C. D. (1993). The role of social comparisons processes in the psychological adaptation of elderly adults. *Journal of Gerontology*, 48(3), 127–136.
15. Rijken, M., Komproe, I. H., Ros, W. J. G., Winnubst, J. A. M., & van Heesch, N. C. A. (1995). Subjective well-being of elderly women: Conceptual differences between cancer patients, women suffering from chronic ailments and healthy women. *British Journal of Clinical Psychology*, 34, 289–300.
16. Rees, J., MacDonagh, R., Waldron, D., & O’Boyle, C. (2004). Measuring quality of life in patients with advanced cancer. *European Journal of Palliative Care*, 11(3), 104–106.
17. Schwartz, C. E., Merriman, M., Reed, G., & Hammes, B. (2004). Measuring patient treatment preferences in end-of-life care research: applications for advance care planning interventions and response shift research. *Journal of Palliative Medicine*, 7(2), 233–245.
18. Schwartz, C. E., Wheeler, H. B., Hammes, B., Basque, N., Edmunds, J., Reed, G., et al. (2002). Early intervention in planning end-of-life care with ambulatory geriatric patients: Results of a pilot trial. *Archives of Internal Medicine*, 162(14), 1611–1618.
19. Schwartz, C. E., Merriman, M. P., Reed, G., & Byock, I. (2005). Evaluation of the Missoula-VITAS Quality of Life Index - Revised: Research tool or clinical tool? *Journal of Palliative Medicine*, 8(1), 121–135.
20. Razmjou, H., Yee, A., Ford, M., & Finkelstein, J. A. (2006). Response shift in outcome assessment in patients undergoing total knee arthroplasty. *The Journal of Bone and Joint Surgery American*, 88(12), 2590–2595.
21. Finkelstein, J. A., Razmjou, H., & Schwartz, C. E. (2009). Response shift and outcome assessment in orthopedic surgery: Is

- there is a difference between complete vs. partial treatment? *Journal of Clinical Epidemiology*, 82, 1189–1190.
22. 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. doi:[10.1007/s11136-006-0025-9](https://doi.org/10.1007/s11136-006-0025-9).
 23. Oort, F. J., Visser, M. R., & Sprangers, M. A. G. (2005). An application of structural equation modeling to detect response shifts and true change in quality of life data from cancer patients undergoing invasive surgery. *Quality of Life Research*, 14, 599–609.
 24. Norman, G. R., Sloan, J. A., & Wywich, K. W. (2003). Interpretation of changes in health-related quality of life: The remarkable universality of half a standard deviation. *Medical Care*, 41, 582–592.
 25. Howard, G. S., Ralph, K. M., Gulanick, N. A., Maxwell, S. E., Nance, D. W., & Gerber, S. K. (1979). Internal invalidity in pretest-posttest self-report evaluations and a re-evaluation of retrospective pretests. *Applied Psychology Measurement*, 3(1), 1–23.
 26. Golembiewski, R. T., Billingsley, K., & Yeager, S. (1976). Measuring change and persistence in human affairs: types of change generated by OD designs. *J. Applied Behav. Sci.*, 12, 133–157.
 27. 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.
 28. 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.
 29. 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.
 30. Oort, F. J., Visser, M. R. M., & Sprangers, M. A. G. (2009). Measurement and conceptual perspectives on response shift: Formal definitions of measurement bias, explanation bias, and response shift. *Journal of Clinical Epidemiology*, 62, 1126–1137.
 31. Korfage, I. J., de Koning, H. J., & Essink-Bot, M. L. (2007). Response shift due to diagnosis and primary treatment of localized prostate cancer: a then-test and a vignette study. *Quality of Life Research*, 16, 1627–1634.
 32. Sprangers, M., & Hoogstraten, J. (1989). Pretesting effects in retrospective pretest-posttest designs. *Journal of Applied Psychology*, 74(2), 265–272.
 33. Li, Y., & Rapkin, B. (2009). Classification and regression tree analysis to identify complex cognitive paths underlying quality of life response shifts: A study of individuals living with HIV/AIDS. *Journal of Clinical Epidemiology*, 62, 1138–1147.
 34. Visser, M. R. M., Oort, F. J., & Sprangers, M. A. G. (2005). Methods to detect response shift in quality of life data: A convergent validity study. *Quality of Life Research*, 14, 629–639.
 35. Schwartz, C. E., & Rapkin, B. D. (2011). Understanding appraisal processes underlying the then-test: A mixed methods investigation. *Quality of Life Research*, 21(3), 381–388. doi:[10.1007/s11136-011-0023-4](https://doi.org/10.1007/s11136-011-0023-4).
 36. 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. doi:[10.1007/s11136-011-0056-8](https://doi.org/10.1007/s11136-011-0056-8).
 37. Howard, G. S., & Dailey, P. R. (1979). Response shift bias: a source of contamination of self-report measures. *Journal of Applied Psychology*, 64(2), 144–150.
 38. Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1993). *Classification and regression trees*. New York: Chapman & Hall/CRC.
 39. Li, Y., & Schwartz, C. E. (2011). Data mining for response shift patterns using recursive partitioning tree analysis. *Quality of Life Research*, 20(10), 1543–1553. doi:[10.1007/s11136-011-0004-7](https://doi.org/10.1007/s11136-011-0004-7).
 40. Martin, M. A., Meyricke, R., O'Neill, T., & Roberts, S. (2006). Mastectomy or breast conserving surgery? Factors affecting type of surgical treatment for breast cancer: a classification tree approach. *BMC Cancer*, 6, 98.
 41. Gruenewald, T. L., Mroczek, D. K., Ryff, C. D., & Singer, B. H. (2008). Diverse pathways to positive and negative affect in adulthood and later life: an integrative approach using recursive partitioning. *Developmental Psychology*, 44, 330–343.
 42. Radespiel-Troger, M., Rabenstein, T., Schneider, H. T., & Lausen, B. (2003). Comparison of tree-based methods for prognostic stratification of survival data. *Artificial Intelligence in Medicine*, 28, 323–341.
 43. Sedrakyan, A., Zhang, H., Treasure, T., & Krumholz, H. M. (2006). Recursive partitioning-based preoperative risk stratification for atrial fibrillation after coronary artery bypass surgery. *American Heart Journal*, 151, 720–724.
 44. Li, Y., & Schwartz, C. E. (2011). Using classification and regression tree to examine evidence of quality of life response shift in patients with multiple sclerosis. *Quality of Life Research*, 20(10), 1543–1553.
 45. Boucekine, M., L'Mouaci, R., Flores, P. M., Butzkueven, H., Baumstarck, K., Ghattas, B., et al. (2012). Understanding response shift in multiple sclerosis patients: application of the Random Forest method. [Abstract]. *Quality of Life Research*, 21(S1), 32–33.
 46. Lix, L. M., Sajobi, T. T., Sawatzky, R., Liu, J., Mayo, N. E., Huang, Y., et al. (2012). Relative importance measures for reprioritization response shift. *Quality of Life Research*. doi:[10.1007/s11136-012-0198-3](https://doi.org/10.1007/s11136-012-0198-3).
 47. Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
 48. Oort, F. J. (2005). Using structural equation modeling to detect response shifts and true change. *Quality of Life Research*, 14, 587–598.
 49. Schmitt, N. (1982). The use of analysis of covariance structures to assess beta and gamma change. *Multivariate Behavioral Research*, 17, 343–358.
 50. King-Kallimanis, B. L., Oort, F. J., Visser, M. R., & Sprangers, M. A. (2009). Structural equation modeling of health-related quality-of-life data illustrates the measurement and conceptual perspectives on response shift. *Journal of Clinical Epidemiology*, 62(11), 1157–1164.
 51. Ahmed, S., Mayo, N. E., Corbiere, M., Wood-Dauphinee, S., Hanley, J., & Cohen, R. (2005). Change in quality of life in people with stroke over time: true change or response shift? *Quality of Life Research*, 14, 611–627.
 52. Barclay-Goddard, R., Lix, L. M., Tatec, R., Weinberg, L., & Mayo, N. E. (2009). Response shift was identified over multiple occasions with a structural equation modeling framework. *Journal of Clinical Epidemiology*, 62, 1181–1188.
 53. Nolte, S., Elsworth, G. R., Sinclair, A. J., & Osborne, R. H. (2009). A test of measurement invariance fails to support the application of then-test questions as a remedy to response shift bias. *Journal of Clinical Epidemiology*, 62, 1173–1180.
 54. 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.

55. King-Kallimanis, B. L., Oort, F. J., Nolte, S., Schwartz, C. E., & Sprangers, M. A. (2011). Using structural equation modeling to detect response shift in performance and health-related quality of life scores of multiple sclerosis patients. *Quality of Life Research*, 20(10), 1527–1540. doi:[10.1007/s11136-010-9844-9](https://doi.org/10.1007/s11136-010-9844-9).
56. Ahmed, S., Sawatzky, R., Levesque, J. F., Ehrmann-Feldman, D., & Schwartz, C. E. (2012). Minimal Evidence of Response Shift in the Absence of a Catalyst. *Quality of Life Research*, 21(Supplement 1), 2–3.
57. Ahmed, S., Schwartz, C., Ring, L., & Sprangers, M. A. (2009). Applications of health-related quality of life for guiding health care: Advances in response shift research. *Journal of Clinical Epidemiology*, 62(11), 1115–1117. doi:[10.1016/j.jclinepi.2009.04.006](https://doi.org/10.1016/j.jclinepi.2009.04.006).
58. Mayo, N., Scott, C., & Ahmed, S. (2009). Case management post-stroke did not induce response shift: The value of residuals. *Journal of Clinical Epidemiology*, 62, 1148–1156.
59. Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, 2(1), 302–317. doi:[10.1111/j.1751-9004.2007.00054.x](https://doi.org/10.1111/j.1751-9004.2007.00054.x).
60. Sawatzky, R., Ratner, P. A., Kopec, J. A., & Zumbo, B. D. (2011). Latent variable mixture models: a promising approach for the validation of patient reported outcomes. *Quality of Life Research*. doi:[10.1007/s11136-011-9976-6](https://doi.org/10.1007/s11136-011-9976-6).
61. Hancock, G. R., & Samuelsen, K. M. (2008). *Advances in latent variable mixture models*. Charlotte, NC: Information Age Pub.
62. Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage Publications.
63. Goodkin, D. E., Reingold, S., Sibley, W., Wolinsky, J., McFarland, H., Cookfair, D., et al. (1999). Guidelines for clinical trials of new therapeutic agents in multiple sclerosis: Reporting extended results from phase III clinical trials. National Multiple Sclerosis Society Advisory Committee on Clinical Trials of New Agents in Multiple Sclerosis. [Practice Guideline]. *Annals of Neurology*, 46(1), 132–134.
64. Goodkin, D. E., Ross, J. S., Medendorp, S. V., Konecsni, J., & Rudick, R. A. (1992). Magnetic resonance imaging lesion enlargement in multiple sclerosis. Disease-related activity, chance occurrence, or measurement artifact? [Research Support, Non-U.S. Gov't]. *Archives of Neurology*, 49(3), 261–263.
65. Deyo, R. A., Battie, M., Beurskens, A. J., Bombardier, C., Croft, P., Koes, B., et al. (1998). Outcome measures for low back pain research. A proposal for standardized use. *SPINE*, 23(18), 2003–2013.
66. Finkelstein, J. A., Quaranto, B. R., & Schwartz, C. E. (2013). Threats to the internal validity of spinal surgery outcome assessment: Recalibration response shift or implicit theories of change? *Applied Quality of Life Research*. (in press).
67. Sawatzky, R., Gadermann, A., Ratner, P. A., Zumbo, B. D., & Lix, L. (2012). Identifying individuals with inflammatory bowel disease who experienced response shift: A latent class analysis? *Quality of Life Research*, 21(Supplement 1), 33.
68. Gandhi, P. K., Ried, L. D., Huang, I. C., Kimberlin, C. L., & Kauf, T. L. (2012). Assessment of response shift using two structural equation modeling techniques. *Quality of life research: an international journal of quality of life aspects of treatment, care and rehabilitation*. doi:[10.1007/s11136-012-0171-1](https://doi.org/10.1007/s11136-012-0171-1).
69. Tabachnick, B. G., & Fidell, L. S. (2013). Assumptions and limitations of multivariate statistical methods. In *Using multivariate statistics* (6th ed.). Boston: Allyn and Bacon.
70. Fairclough, D. L. (2010). *Design and analysis of quality of life studies in clinical trials* (2 ed.). Chapman & Hall/CRC Interdisciplinary Statistics Series). New York: CRC Press, Taylor & Francis Group.
71. Fielding, S., Fayers, P. M., & Ramsay, C. R. (2009). Investigating the missing data mechanism in quality of life outcomes: A comparison of approaches. *Health and Quality of Life Outcomes*, 7, 57–66.
72. Little, R. J. A., & Rubin, B. D. (2002). *Statistical analysis with missing data* (2nd ed.). New Jersey: Wiley.
73. Ibrahim, J. G., & Molenbergh, G. (2009). Missing data methods in longitudinal studies: A review. *Test*, 18, 1–43.
74. Little, R. J., D'Agostino, R., Cohen, M. L., Dickersin, K., Emerson, S. S., Farrar, J. T., et al. (2012). The prevention and treatment of missing data in clinical trials. *The New England journal of medicine*, 367(14), 1355–1360. doi:[10.1056/NEJMsr1203730](https://doi.org/10.1056/NEJMsr1203730).
75. Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling*, 6(1), 56–83. doi:[10.1080/10705519909540119](https://doi.org/10.1080/10705519909540119).
76. Jackson, D. L. (2003). Revisiting Sample Size and Number of Parameter Estimates: Some Support for the N:q Hypothesis. *Structural Equation Modeling: A Multidisciplinary Journal*, 10(1), 128–141.
77. Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences*, 42, 815–824.
78. Shah, R., & Goldstein, S. M. (2006). Use of structural equation modeling in operations management research: Looking back and forward. *Journal of Operations Management*, 24, 148–169.
79. Muthén, L., & Muthén, B. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 9, 599–620.
80. Steyerberg, E. W., Harrell, F. E. J., Borsboom, G. J. J. M., Eijkemans, M. J. C., Vergouwe, Y., & Habbema, J. D. F. (2001). Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *Journal of Clinical Epidemiology*, 54, 774–781.
81. Mayo, N. E., Scott, S. C., & Lix, L. (2012). Caring for a spouse with stroke: True change in QOL and response shift. *Quality of Life Research*, 21(Supplement 1), 2–3.
82. Brossart, D. F., Clay, D. L., & Willson, V. L. (2002). Methodological and statistical considerations for threats to internal validity in pediatric outcome data: Response shift in self-report outcomes. *Journal of Pediatric Psychology*, 27(1), 97–107.
83. Ahmed, S., Mayo, N., Scott, S., Kuspinar, A., & Schwartz, C. (2011). Using latent trajectory analysis of residuals to detect response shift in general health among patients with multiple sclerosis article. *Quality of Life Research*, 20(10), 1555–1560. doi:[10.1007/s11136-011-0005-6](https://doi.org/10.1007/s11136-011-0005-6).
84. Schwartz, C. E., Sajobi, T., Lix, L., Quaranto, B. R., & Finkelstein, J. A. (2013). Changing values, changing outcomes: The influence of reprioritization response shift on outcome assessment after spine surgery. *Quality of Life Research*, (in press). doi:[10.1007/s11136-013-0377-x](https://doi.org/10.1007/s11136-013-0377-x).