

Data mining for response shift patterns in multiple sclerosis patients using recursive partitioning tree analysis

Yuelin Li · Carolyn E. Schwartz

Accepted: 29 August 2011 / Published online: 11 September 2011
© Springer Science+Business Media B.V. 2011

Abstract

Aims To examine evidence of QOL response shift in patients with multiple sclerosis (MS) using recursive partitioning tree analysis (RPART) technique.

Methods Subjects: MS patients from the NARCOMS registry assessed an average of 6 times at a median interval of 6 months. Outcomes: SF-12v2 Physical & Mental Component Scores (PCS, MCS). Covariates: Patient-determined disease steps, Performance Scales, and symptomatic therapies. RPART trees were fitted separately by 3 disease-trajectory groups: (1) relapsing ($n = 1,582$); (2) stable ($n = 787$); and (3) progressive ($n = 639$). The resulting trees were interpreted by identifying salient terminal nodes that showed the unexpected *quantitative* patterns of contrasting MCS and PCS scores (e.g., PCS deteriorates but MCS is stable or improves), using a minimally important difference of at least 5 points on the SF-12v2. *Qualitative* indicators of response shift were different thresholds (recalibration), content (reconceptualization), and order (reprioritization) of disability domains in predicting PCS change by group.

Results Overall, 20% of patients demonstrated response shift quantitatively, with 10% in the “progressive” cohort, 8% in the “relapsing” cohort, and 2% in the “stable” cohort. RPART trees differed qualitatively across disease-trajectory groups in patterns suggestive of recalibration, reprioritization, and reconceptualization. Disability subscales, but not symptom management, distinguished homogenous groups.

Conclusions PCS and MCS change scores are obfuscated by response shifts. The contingent true scores for PCS change scores are not comparable across patient groups.

Keywords Response shift · Statistical analysis · Recursive and partitioning tree analysis · Methods · Multiple sclerosis · Appraisal

Introduction

The recent expansion of data mining methods and technology has opened the door to new tools for studying patterns in data. These tools have been applied in a number of fields, including marketing research [1], economic prediction [2], epidemiologic research [3], and health services research [4], but their application to quality-of-life (QOL) research is only recent [5–9]. Such methods present intriguing tools for hypothesis generation in evaluating the role of qualitative processes in quantitative assessment. As QOL measurement matures and methods become increasingly sophisticated, the onus for mixed methods that integrate qualitative and quantitative data becomes clear [10, 11]. The present work applies a data mining technique to a large registry of multiple sclerosis (MS) patients to examine evidence of qualitative differences in patient-

The authors Yuelin Li and Carolyn E. Schwartz contributed equally to this work.

Y. Li
Behavioral Science, Memorial Sloan Kettering Cancer Center,
New York, NY, USA

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

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

reported outcomes (PROs) as a function of disease trajectory.

Background on MS

MS is the most prevalent chronic progressive neurological disease among young adults worldwide and in the United States. The National Multiple Sclerosis Society [12] has estimated that there are approximately 400,000 cases of MS in the United States with an incidence of nearly 200 new cases each week. The majority of people with MS are diagnosed between 20 and 50 years of age, and women are affected two to three times as often as men [12].

The clinical course of MS is generally characterized by reversible periods of neurological impairment and disability (relapsing-remitting disease) that may be followed by continuous irreversible impairment and disability (secondary progressive disease), or progression from clinical onset with occasional plateaus or temporary minor improvements (primary progressive disease) [13]. MS is associated with immune dysregulation and neurodegenerative processes [14]. The past decade has witnessed vast development in disease-modifying agents for MS [15–17].

A large number of MS-specific measures have also been developed in this period of time [18–23]. Although these measurement advances are substantial, meta-analyses of the impact of disease-modifying agents suggest that their impact is small [24–26]. It is possible that the magnitude of detected effects is obfuscated by adaptive processes or “response shifts”. Such “response shifts” represent health-related changes in the meaning of measured concepts, due to changes in the individual’s internal standards, values, and conceptualization of the concept(s) being measured [27].

Response shifts would be expected in a patient population that has changeable and unpredictable disease trajectories that affect a broad range of symptoms. For example, an ambulation difficulty such as stumbling in the period after diagnosis might be experienced and reported as “badly off” by an individual with MS, whereas later in the disease experience an ambulation difficulty that requires a walker might be reported as satisfactory functioning. Similarly, the meaning of “severe fatigue” may depend on whether the patient just experienced an exacerbation. This experience or health state can affect the person’s internal standards of what “severe” fatigue means, their values in terms of what activities take higher priority in thinking about their quality of life, and even their conceptualization of key health-related QOL concepts such as role performance (e.g., what is “role performance” if one is in the middle of an exacerbation as compared to in remission? How do expectations of one’s functioning change during an exacerbation as compared to before or much later following remission?).

Response shift as an epiphenomenon

Response shift is an epiphenomenon that is inferred when changes in appraisal explain the discrepancy between expected and observed indicators of QOL [28]. Li and Rapkin [5] recently tested this model using recursive partitioning tree (RPART) analysis in longitudinal data of people with HIV. Their analysis integrated qualitative data from the QOL Appraisal Profile [28] and quantitative data from the MOS-36 [29], revealed distinct patterns in appraisal change that substantially increased the amount of explained variance in mental health outcomes, and revealed complex and non-linear patterns. These distinct appraisal patterns supported the idea of the “contingent true score” [30], where an observed PRO score is comparable across people or across time only if the appraisal processes are similar across comparisons.

It is possible that RPART analysis could be used to investigate appraisal processes even when they are not measured directly. As an extension of this epiphenomenological approach, one could investigate indicators of appraisal changes via emergent patterns of disease-specific PRO subscales in explaining noticeable incongruent generic physical and mental health composite scores. One would expect that disease-specific measures would be less susceptible to response shifts if they query more specific symptoms or functional limitations than generic measures, although this issue has not been addressed empirically to date to our knowledge. Response shifts are more detectable in evaluative measures [30] that are less specific, such as generic measures.

These incongruent patterns of PRO subscales would be reminiscent of the paradoxical discrepancies in QOL valuations on chronic diseases between the general public and the patients, e.g., the general public gives a QOL value of 0.39 to dialysis, whereas dialysis patients give a 0.56 to their own QOL (e.g., [31, 32]). Some MS patients in our sample may report improved mental health status despite severe impairments in physical functioning. Conceivably, if these patients were asked to evaluate their own QOL on a scale where 0 represents death and 1 represents perfect health, they might assign a number nearer 1 than 0. We may therefore reasonably expect response shift to manifest itself in a large registry data as patients reporting high mental health scores despite severe physical limitations or vice versa. Even though the psychometric properties may dictate a low correlation between the PRO subscales we study, we would nevertheless be surprised to find improved mental functioning despite severe physical limitations.

The present work attempts a data mining approach in a large registry sample of MS patients. As this data base did not measure appraisal processes or other indicators of response shift directly, response shift is inferred as an

epiphenomenon in the interpretation of the apparent discrepancies between mental health and physical functioning. This analytic approach has the potential to be more informative than other response-shift detection methods because it examines all possible relationships, that is both linear and non-linear predictors, in partitioning patients into homogeneous groups. It thus has the potential to show more complex patterns of relationships that are qualitatively distinct and meaningful. For example, response shift theory predicts that response shift is inferred when changes in appraisal explain the discrepancy between expected and observed QOL. In the present work which does not have direct measures of appraisal processes, we operationalize response shift as unexpected patterns of contrasting MCS and PCS scores (e.g., PCS deteriorates but MCS is stable or improves).

Methods

Sample

The North American Research Committee on Multiple Sclerosis (NARCOMS) Registry is a self-report Registry developed in 1993 by the Consortium of Multiple Sclerosis Centers and includes over 34,000 individuals who have multiple sclerosis, with over 10,000 updating their data every 6 months. Patients participate by completing either paper or secure web-based survey forms capturing data on demographics, disease characteristics, disability and handicap, treatments, and access to health providers. For the purpose of this project, a NARCOMS sample was drawn of patients who had enrolled in the NARCOMS registry within 1 year of diagnosis and who provided at least annual updates to the Registry ($N = 3,839$).

Measures

Standardized PRO questionnaires include: (1) the *performance scales (PS)* measure of disability [18], which includes items for mobility, hand function, fatigue, cognition, bladder/bowel, sensory, spasticity, vision, depression, tremor, and pain. Subscale scores range from 0 to 5, with the exception of mobility, which ranges from 0 to 6; (2) the 9-level categorical *patient-determined disease steps (PDDS)* measure of disease progression that was adapted from the clinician-reported disease steps [19] and the gold-standard clinical neurological exam, the Expanded Disability Status Scale (EDSS) [33]. PDDS scores range from 0 (normal) to 8 (bedridden) and correlate highly with the EDSS [33]; and (3) the *Short-Form 12v2 (SF-12)* [34], a generic health measure that yields composite scores for mental health (MCS) and physical functioning (PCS). In

addition to the above PROs, patient responses to items on symptom change and relapse experience were used to classify patients' disease course (i.e., relapsing, stable, or progressive disease; see below).

Statistical analysis

Creating the “relapsing,” “stable,” and “progressive” patient cohorts

We focused on the 2005–2009 data because the more recent data were associated with more complete assessments. Each patient was deemed “relapsing,” “stable,” or “progressive” by the following algorithm on the basis of defining parameters that are consistent with clinical definitions of these subgroups, and as discussed and agreed upon with a senior neurologist involved in the project (TV). We first identified each patient's latest assessment and went back in time in a retrospective review of symptoms for up to 2 years. Patients were identified as part of the “*actively relapsing*” if they reported any relapse within 2 years ($n = 1,582$, 53% of the sample). Patients were identified as “*progressive*” if they reported no relapse within 2 years and reported worsening symptoms at least once in the past 2 years ($n = 639$, 21% of the sample). Patients were identified as “*stable*” if they reported no worsening of symptoms and no relapse for all consecutive assessments over the duration of up to 2 years ($n = 787$, 26% of the sample). The stable group would be used as a comparison group in the analysis under the assumption that they would not be prone to response shifts in perceived QOL.

RPART model specifications

We used the recursive partitioning and regression trees method [35, 36] in R software [37] to model the changes in the SF-12 PCS and MCS scores. We fitted RPART trees for the three cohorts separately, using PCS scores as the dependent variable, and a set of 42 clinically relevant covariates including use of 29 treatment or symptom relief medications (e.g., baclofen, ditropan, etc.), 11 items from the performance scales reflecting physical and psychosocial function (e.g., mobility, fatigue, cognition, etc.), the PDDS disability score, and time since diagnosis. These selected covariates were included on the basis of clinical logic and previous knowledge that justifies the inclusion of this set of covariates. Because patients provided multiple SF-12 assessments (median = 5), an analysis based on the full dataset would violate the assumption of independent assessments. We did not use the longitudinal RPART approach (“long RPART” software program [38]) that is designed to classify cases by growth curves (e.g., by each

person's growth curve intercept and slope fitted to all available assessments over the 5-year period). Preliminary graphing of the longitudinal PCS scores revealed that the individual growth curves over time were highly variable. We were not able to identify visibly linear nor quadratic growth patterns over time, suggesting that the "long RPART" approach was unlikely to yield insightful classifications. Thus, to avoid double counting the patients, we randomly sampled two consecutive assessments per patient and calculated each patient's change scores on the PCS scores as well as changes in functional indicators.

The RPART [35] model fitted the changes in PCS component scores as a function of the abovementioned 42 predictors, representing the changes in physical and psychosocial functional indicators between randomly selected consecutive assessments. We included each patient's initial PCS scores to control for ceiling and floor effects and to simplify the interpretation of the magnitude of changes.

The full sample of 3,839 patients in the 2005–2009 registry was first analyzed by a preliminary model with all default parameters of the RPART statistical procedure (found in the RPART software documentations as part of the RPART download). This served to identify cases that deemed by RPART as containing no usable information, cases in which the outcome variable is missing and/or all predictors in the preliminary RPART model are missing. Therefore, these are cases that contain insufficient information for RPART to make a classification. A sample of 831 patients was excluded thusly, leaving an analytic sample of 3,008 patients.

Separate RPART trees were fitted for each patient cohort. We followed the general approach in RPART analysis: (1) stopping rule for a terminal node (20 observations) (2) tenfold cross-validation automatically carried out, (3) true-pruning by the result of the tenfold cross-classification and the one-standard-deviation rule [39] in the cost-complexity criterion, (4) specification of priors (proportional to data counts), and (5) missing data are handled by surrogate splits. The following algorithm provides an intuitive explanation on surrogate splits. For example, a person has a missing value in a model that partitions an outcome variable *Y* by predictor variables *A*, *B*, and *C*. Assume also that this person's missing value is found in predictor *A* so that his or her branching by *A* cannot be determined. The model has to rely instead on this person's non-missing values in *B* and *C* as surrogate splits. An agreement is calculated between the classifications based on the primary split of *A* and the surrogate splits *B* and *C* (on cases with non-missing *A*). Whichever surrogate split with the highest agreement wins and the case is classified accordingly.

RPART performs a tenfold cross-classification by default to help evaluate the reliability of the tree model. The full sample is randomly divided into 10 sub-samples. Internally, the full RPART tree is carried out with 90% of

the full sample, and the remaining 10% of the sample is used as a validation dataset to calculate a cross-classification error rate. This procedure is repeated 10 times, each time with 9 subsets as the modeling dataset and the remaining 1 subset as the validation dataset. We used the results of this tenfold cross-validation to prune the full tree back down to a more parsimonious model by the 1-SD rule. Additional technical details can be found in [35]. These considerations were similar to our prior work [5].

Interpretation of RPART findings

We report the full RPART trees, rather than the pruned trees by the 1-standard error rule in cross-validation error [35, 36]; because pruning is likely to omit small groups of patients with subtle QOL changes. We wanted to better examine the sizes of these potentially small clusters to inform the relative scale of response shift within each patient cohort. The interpretation of changes in physical and mental functioning was generally based on the scales of population norms for these summary scores, with a mean of 50 and a standard deviation of 10 [40]. To the best of our knowledge, there is no consensus yet on the minimally important difference in the SF-12 component scores for MS patients. Thus, in evaluating the RPART findings, we generally considered a 5-point change or greater a medium change (i.e., one-half standard deviation), and a change of 10 points or greater a large change (i.e., one-standard deviation). These reference points provide useful guidance for clinically meaningful change to identify clusters of patients who report considerable changes in physical functioning despite their mental health scores remaining unchanged. We primarily focus on these salient patterns of PCS and MCS scores and the corresponding sizes of these patient clusters. Response shift was inferred by *qualitative* differences in thresholds, content, and order of disability domains that were retained by the RPART analysis. We operationalize response shift *quantitatively* as unexpected patterns of contrasting MCS and PCS scores (e.g., PCS deteriorates but MCS is stable or improves). We typically offer no remarks on patient clusters with change scores less than this minimally important difference (MID) of 5 points.

Results

Sample characteristics

Table 1 shows the demographic characteristics of the study sample. The sample consisted of 3,008 patients with a mean age of 42.4, of whom 83% were female. Time since diagnosis was less than 1 year in 42 percent of the sample, and greater than 1 year in 58 percent of the sample.

Table 1 Sample demographics

	Progressive, <i>N</i> = 639 (201%)		Relapsing, <i>N</i> = 1,582 (53%)		Stable, <i>N</i> = 787 (26%)	
	Mean	SD	Mean	SD	Mean	SD
Current age ^b	44.9	9.3	40.9	9.5	40.9	9.5
Age @ diagnosis	43.57	9.29	40.25	9.06	39.90	9.47
Age @ symptom onset	35.65	9.72	32.70	9.81	33.87	9.91
	<i>N</i> ^a	%	<i>N</i>	%	<i>N</i>	%
Gender: female	477	77	1,273	84	623	86
Gender: male	141	23	250	16	103	14
Race: white	572	93	1,385	91	672	93
Race: nonwhite	43	7	129	9	51	7
Education: ≤ HS	211	35	558	37	214	30
Education: > HS	403	65	958	63	500	70
SF-12 physical component score (PCS)						
Baseline mean (SD)	41.77	10.15	41.28	10.74	50.43	8.24
Change score mean (SD)	−0.44	6.88	0.07	7.47	0.12	5.05
SF-12 mental component score (MCS)						
Baseline mean (SD)	48.08	7.81	46.14	7.90	50.07	6.52
Change score mean (SD)	−0.05	6.97	0.10	7.65	−0.47	6.0
Patient-determined disease steps (PDDS)						
Baseline mean (SD)	2.83	2.11	2.59	2.00	1.06	1.59
Change score mean (SD)	−0.10	0.87	−0.12	0.95	0.01	0.71
<i>Performance scales (PS)</i>						
Baseline mean (SD)						
Mobility	2.15	1.78	1.96	1.64	0.79	1.23
Hand function	1.29	1.17	1.44	1.19	0.69	0.94
Fatigue	2.59	1.26	2.82	1.32	1.59	1.20
Cognition	1.56	1.17	1.85	1.24	0.99	0.96
Bladder/Bowel	1.40	1.13	1.48	1.22	0.79	0.95
Sensory	1.67	1.20	1.97	1.27	1.10	0.90
Spasticity	1.61	1.30	1.77	1.31	0.75	0.97
Vision	0.96	1.07	1.35	1.15	0.75	0.89
Pain	1.72	1.39	2.08	1.44	0.91	1.10

^a Numbers may not add up to total *N* because of missing information, which is omitted in the denominators of the percentages

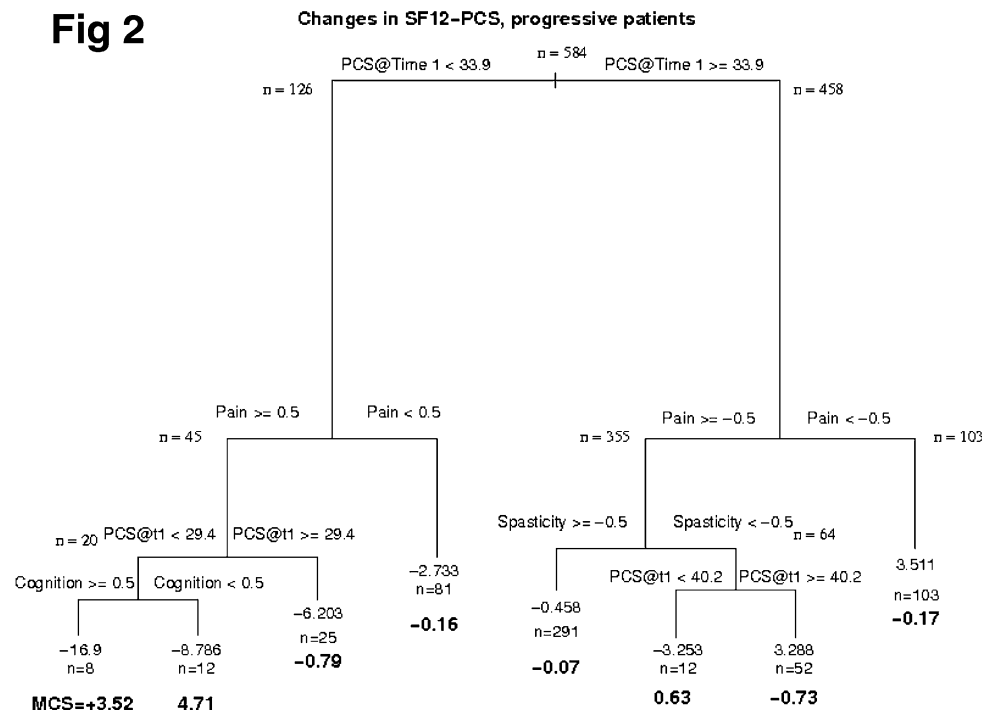
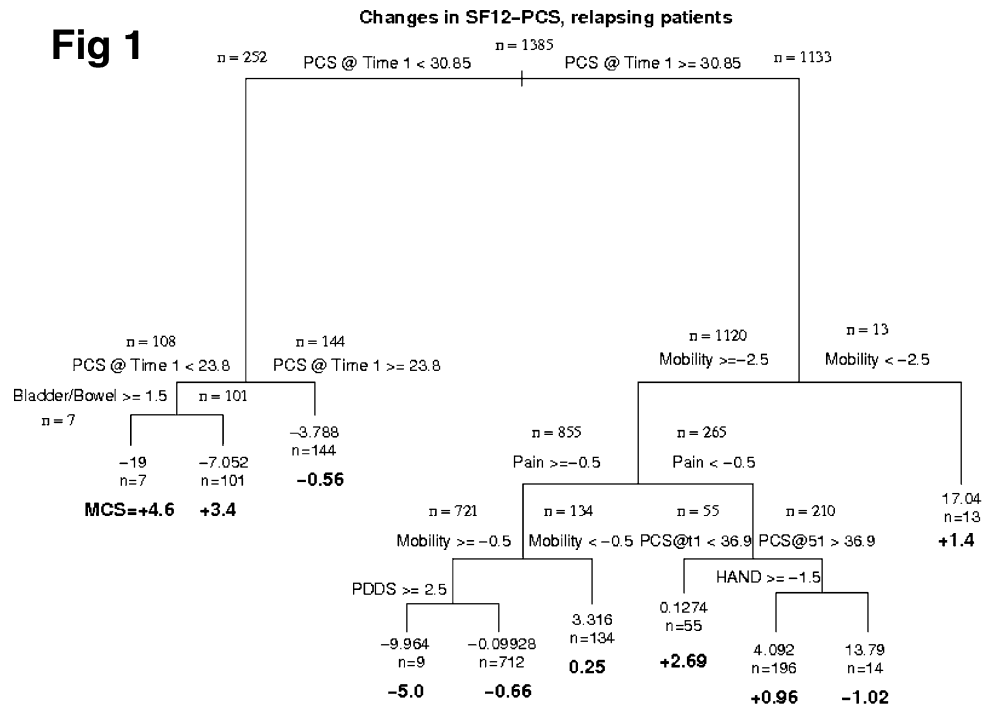
^b Registry data only contained year of birth. Thus, current age was calculated by the difference in year of birth and year of the date of survey

Generally, we interpret the RPART trees by first identifying salient terminal nodes that show the greatest changes in PCS and MCS component scores. The size of the identified terminal nodes is noted. We then focus on tracing the binary splits in the RPART tree for changes in disease management and functional status that contribute to such extreme changes in quality of life assessments.

PCS trees

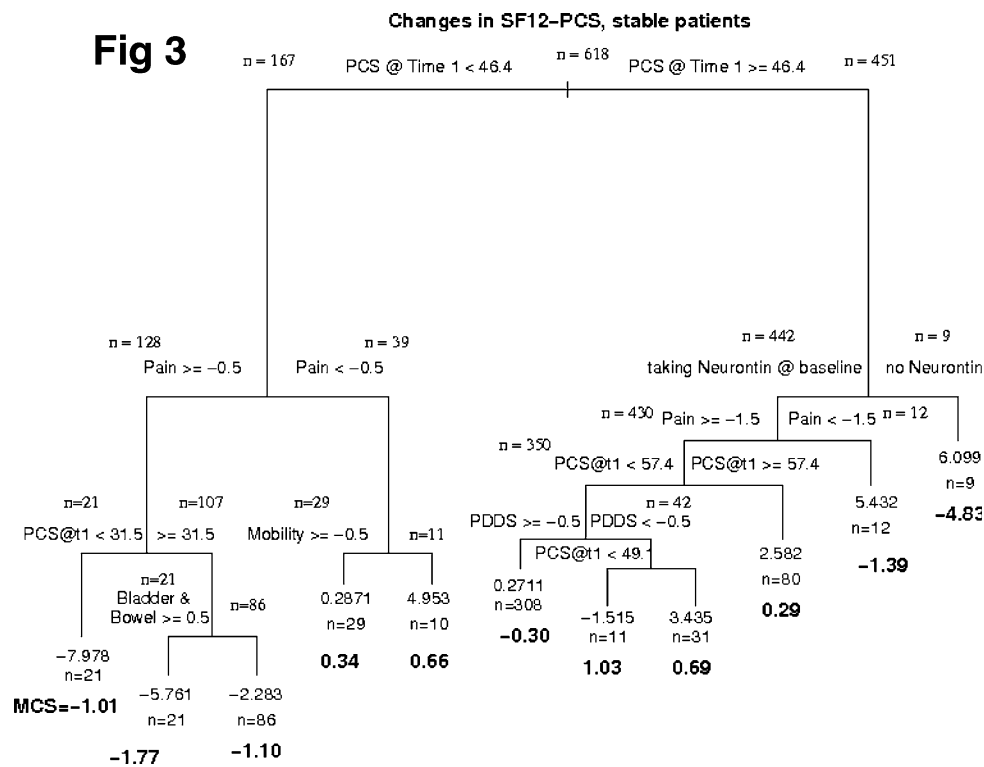
Figures 1–3 shows the full RPART tree modeling the PCS changes over time among “relapsing” patients. All splits including the initial split are data driven. None of the treatment and symptom-management covariates were

deemed predictive of PCS change. The greatest PCS change was observed in the leftmost terminal node of 7 “relapsing” patients, who reported an average *reduction* of 19 points between the two assessments. These patients’ baseline PCS scores were already extremely low (less than 23.84, after the first two consecutive splits to the left) and their self-reported bladder/bowel functioning deteriorated considerably (an increased disability of greater than 1.5; e.g., from 2 = “mild bladder/bowl disability” to 4 = “severe bladder/bowl disability”). A group size of 7 is small relative to the full sample. However, a PCS score near zero means that these patients reported severe limitations in all aspects of the PCS domain. Also plotted in bold font at the bottom are the average MCS change scores



Figs. 1–3 Figures 1 through 3 present the pruned RPART trees for changes in SF-12 PCS scores over time for relapsing, progressive and stable patients, respectively. The initial cut-point represents the most important interaction term for distinguishing homogenous patient groups within the disease-trajectory grouping and the branches that follow indicate the interaction terms that create increasingly homogenous patient groupings. Note that if a statement is true (e.g., PCS at time 1 < 30.85 in Fig. 1, top branch), the group for whom that

statement is true falls on the left side of the tree; if false, on the right side of the tree. The final groupings reflect patient groups who share cut-points, content and number of domains, and order of domains in predicting their change in PCS change scores. MCS change scores are also added for reference in *bold*. These tree branches thus reflect latent appraisal processes, and thus contingent true scores on PCS



Figs. 1–3 continued

for the patients in the corresponding terminal nodes. The 7 patients in the leftmost node reported a small *improvement* of 4.6 points in MCS. The neighboring end cluster, the 101 patients who reported a reduction of 7.05 points in PCS from a score of less than 23.8 at time 1, also reported a small 3.4 points improvement in MCS.

Another great change was an *increase* in PCS of 17.04 points among 13 patients on the other initial node whose initial PCS was above 30.85, and who had maintained mobility (no more than a 2.5-point reduction, e.g., from 0 = “Normal” to 2 = “Mild gait disability” but no more than 3 = “Occasional use of cane or unilateral support”). These patients’ MCS scores improved by 1.4 points—smaller than the 4.6 MCS improvement among the 7 patients described above. Another noteworthy group is the second terminal node from the right of 14 patients whose average PCS scores increased by 13.79 points and average MCS scores decreased by 1.02 points. Although these patients did not maintain mobility, their PCS scores were noticeably similar to the change of 17.04. However, the increase of 13.79 points in PCS among these 14 patients arose from a different configuration of contributing factors: lesser degrees of pain, a baseline PCS greater than 36.88, and lesser hand disability. Something similar occurs for the subgroups $n = 196$ /PCS change = 4.092 and $n = 134$ /PCS change = 3.316. They also result from different pathways but seem to have similar PCS and MCS change scores. These examples

illustrated by the tree indicate that patients may reach similar physical and mental QOL changes by way of disparate pathways of physical symptoms and limitations.

Overall, PCS change among “relapsing” patients was strongly affected by baseline PCS, bladder/bowel disability, mobility, pain, and hand disability, as well as PDDS scores (PCS plus 5 domains). The terminal node groups in the middle of the tree showed change scores below the MID and thus no change in this sampling of time points. These groups included the largest numbers of “relapsing” patients ($\Delta = -0.099$, $n = 710$ and $\Delta = 3.316$, $n = 134$). Using a PCS change of 5 points as a crude guide, we found unexpected patterns of contrasting MCS and PCS scores in 135 patients, in two patient clusters of decreased PCS scores accompanied by increased MCS scores ($n = 7$ and 101), and in two patient clusters of increased PCS scores accompanied largely unchanged MCS scores ($n = 13$ and 14). Thus, whereas the abovementioned patients showed discrepancies in their PCS and MCS change scores (i.e., response shifts), the remaining 1250 patients in the “relapsing” patient cohort (90%) show no remarkable patterns of response shift.

Figures 1–3 shows the full RPART tree modeling the PCS changes over time among “progressive” patients. None of the treatment and symptom-management covariates or PDDS was predictive of PCS change. Further, the greatest PCS change was observed in the leftmost terminal

node of 8 “progressive” patients, who reported an average *reduction* of 16.9 points between the two assessments. These patients’ baseline PCS scores were already extremely low (less than 29.4, after the first splits to the left) and their self-reported pain deteriorated a small amount (an increased pain of greater than 0.5; e.g., from 2 = “mild pain” to less than 3 = “moderate pain”) and slight increase in cognitive limitations (e.g., from 1 = “minimal cognitive disabilities” to 2 = “mild cognitive disabilities”). Although this is also a small group relative to the full sample, their PCS score nearing zero reflects severe limitations in all aspects of the PCS domain. Despite their severe physical limitations, these patients reported a 3.5-point improvement in MCS scores, reflecting a 20.42 point discrepancy in PCS and MCS scores. To the right of this group were 12 patients who reported an 8.8 point PCS reduction and stable cognitive symptoms, yet they reported a 4.7 points improvement in MCS. The greatest positive change was an *increase* in PCS of 3.5 points among 103 patients whose initial PCS was above 33.9, and who had maintained low pain disability (no more than a 0.5-point reduction, e.g., from 3 = “Moderate pain disability” to 2 = “Mild pain disability”). Their MCS change score was near zero.

Overall, PCS change was strongly affected by baseline PCS, pain, and cognitive symptoms. Several terminal node groups totaling over 500 patients showed unremarkable PCS change scores, with the most prominent being the 291 patients with a mean PCS change score of -0.46 points. Again, using a PCS change of 5 points as a crude guide, we found unexpected patterns of MCS and PCS scores in 45 patients: in three patient clusters of decreased PCS scores accompanied by increased MCS scores ($n = 8$ and 12), and in one patient cluster of decreased PCS scores accompanied by unchanged MCS scores ($MCS \Delta = -0.79$, representing a difference of 5.4 from $PCS \Delta$ of -6.2 ; $n = 25$). The remaining 539 patients in this cohort (92%) showed no remarkable patterns of response shift.

Figures 1–3 shows the full RPART tree modeling the PCS changes over time among “stable” patients. Unlike the other patients, one symptom-management covariate was deemed predictive of PCS change: patients who reported not using Neurontin and whose baseline PCS score was higher than 46.4 (i.e., normal relative to the general population) had the highest observed PCS change indicating improved function ($\Delta = 6.099$, $n = 9$). However, their average MCS score reduced by 4.83 points. “stable” patients who reported the greatest PCS decrease ($n = 21$) were in the leftmost terminal node, and an average *reduction* of 7.98 points between the two assessments. These patients’ baseline PCS scores were substantially below the general population norms (less than 31.5 points) and reported a small amount of deterioration on

pain and bladder/bowel disability (less than 0.5 change in both subscales). Another noteworthy group is the second terminal node from the right of 12 patients whose average PCS scores increased 5.43 points, but who used Neurontin and who maintained low pain disability (more than a 1.5-point reduction, e.g., from 3 = “Moderate pain disability” to 1 = “Minimal pain disability”). These 12 patients reported a slight reduction of 1.39 in MCS scores.

Overall, PCS change was strongly affected by baseline PCS, pain, mobility, and PDDS score (PCS plus 3 domains). Several terminal node groups totaling 469 patients showed unremarkable change scores, with the most prominent being the 308 patients with a mean PCS change score of 0.27 points. Patterns of response shift among “stable” patients are subtle. Again, using unexpected patterns of contrasting MCS and PCS scores as the primary method to detect response shift, we found only one cluster of 9 patients (endpoint to the furthest right) whose physical functioning had improved by 6.099 points while their mental health scores had decreased by nearly 5 points. The remaining 609 patients (98%) showed no remarkable patterns of response shift.

Possible evidence of response shift

Table 2 summarizes the operationalizations of the three aspects of response shift as well as the findings that may support response shift hypotheses. There appeared to be differences in RPART trees across disease-trajectory groups with regard to patterns of PCS and MCS suggestive of recalibration, reprioritization, and reconceptualization.

Recalibration response shift was inferred by different group-specific thresholds for cut-points in the RPART trees. For example, the first branching for “stable” patients occurs at a baseline PCS score of 46.4, a value near the general population norm, while the first branching for “relapsing” patients occurs at a much lower baseline PCS score of 30.85. What constitutes a reliable split by baseline PCS in one patient group’s configuration of QOL changes may be considerably lower or higher than another group’s QOL changes, above and beyond what may be expected due solely to measurement error.

Reconceptualization response shift was inferred by changes in the content and/or number of domains by group in the tree over time. For PCS, the statistically relevant disability domains differed by disease-trajectory group, with “relapsing” patients’ trees showing the greatest number of relevant domains (4 domains, PDDS), followed by “stable” patients (2 domains, PDDS, symptomatic therapy), and then “progressive” patients (2 domains).

Reprioritization response shift was inferred by changes in the order of domains in tree pathways over time. For PCS, the disability domains that were relevant across

Table 2 Summary of qualitative indicators of response shifts in RPART analysis

Response shift aspect	Operationalization	PCS		
		Relapsing	Progressive	Stable
Recalibration	Trees indicating relationships between predictors and outcome scores utilize different group-specific thresholds or cut-points for selected predictor-variable interaction terms to identify homogenous groupings over time	Baseline cut-points differ and are > MID Mobility and pain cut-points differ between and within groups Relevant domains differ across disease-trajectory groups so cut-points not applicable		
Reprioritization	Changes in the order of domains in tree pathways over time	Baseline-bladder/ bowel-mobility- pain-hand-PDDS	Baseline- pain- spasticity	Baseline-pain- Neurontin- mobility-PDDS
Reconceptualization	Changes in the content and/or number of domains by group in a pruned tree over time	5 Disability domains: Mobility Bladder/bowl Pain Hand function PDDS Baseline PCS	3 Disability domains: Spasticity Pain Baseline PCS	4 Disability domains: Pain Mobility PDDS Baseline PCS 1 symptomatic therapy: Neurontin

groups entered the RPART tree branches at different levels, supporting a reprioritization response shift. For example, changes in pain disability affected strongly the PCS change scores for “stable” as well as “relapsing” patients, suggesting that pain was high on the priority of change in physical disability. However, the effects of pain were only relevant to “relapsing” patients’ whose mobility was not severely impaired over time.

Discussion

This investigation applied the RPART data mining technique to identify plausible patterns of response shifts in physical health change scores in MS patients distinguished by disease trajectory. Based on the extensive quantitative data analysis, response shift was inferred by qualitative differences in thresholds, content, and order of disability domains that were retained by the RPART analysis. We conclude that there are observable patterns of emergent response shift in unanticipated PCS and MCS scores attributable to different appraisal processes among all three patient cohorts. This work demonstrates that the magnitude of detected effects is obfuscated by adaptive processes or response shifts. The tree analysis shows intriguing evidence that changes in pain disability contribute importantly to the physical functioning of all three cohorts of patients. Idiosyncratic patterns of physical functioning changes are also observable. For example, both “relapsing” and “stable” patients’ changes in physical functioning are also affected by bladder and bowel symptoms and the PDDS.

“progressive” patients’ PCS change scores seem to be affected strongly by limitations in spasticity and cognitive limitations and less so by bladder/bowel symptoms and the PDDS stage. Further, only stable patients evidence an effect of a symptomatic treatment, whereas the trees for both relapsing and progressive patients do not suggest such an effect. This lack of treatment effect may imply that in MS, treatment benefits are shown by lack of change (i.e., stability) rather than improvement, since the disease is chronic and progressive.

These findings suggest that the change scores evidenced by this sample on the PCS and MCS measures are being obfuscated by response shifts and that the contingent true scores for PCS change are not comparable across patient groups. If we accept the use of unexpected PCS and MCS score patterns as indicators of response shift, and the use of terminal node sizes as practical measures of the magnitude of response shift within a specific patient cohort, then we can determine that overall 20% of patients demonstrated response shift using an MID of at least 5 points on the SF-12v2, with 10% in the “progressive” cohort, 8% in the “relapsing” cohort, and 2% in the “stable” cohort. This pattern seems consistent with our intuitive notion of the quality-of-life differences across these three groups. For example, we would expect that “stable” patients show the lowest level of response shift as compared to the other two groups, mainly because they have to show no worsening of symptoms and no relapse for all consecutive assessments over the duration of up to 2 years. This method may be more sensitive to response shift detection than the other response shift detection methods used on this same patient

sample [41, 42]. Future research should triangulate this response-shift detection approach with direct measures of (changes in) appraisal to confirm that this approach of inferring changes in appraisal truly reflects measured changes in appraisal when these data are available.

The application of RPART to response shift research is relatively recent [5] but shows promise for identifying patterns that underlie paradoxically small change scores over time in PROs. At this stage in the application of this method, and in the context of not having direct measures of appraisal, we infer differences in appraisal as a function of differences in thresholds, content, and order of domains included in the trees. Such an analytic exercise is highly exploratory and is prone to classification errors when no direct measures of appraisal are available. It should be noted that although the RPART method has numerous cross-validation steps that would minimize chance findings, the analytic method is exploratory in nature, and findings would be best confirmed in independent samples. Future research should evaluate whether this inference is supported in data sets that include the QOL Appraisal Profile [28] over time, to assess whether RPART analyses that include or exclude appraisal generate similar conclusions about the aspects of response shift reflected in the data. Future research could also codify the types of response shift as distinguished from other types of change, similar to Oort's seminal work codifying the application of SEM to response shift detection [43]. Such codification is critical for a firm scientific foundation for identifying response shift by testing alternative explanations, as well as for providing unbiased estimates of longitudinal changes in PROs.

Acknowledgments This work was funded in part by a Visiting Scientist Fellowship from the Consortium of Multiple Sclerosis Centers to Dr. Schwartz. Dr. Li was supported in part by a Weill Cornell Medical College Clinical and Translational Science Award (NIH UL1-RR024996) (PI: Julianne Imperato-McGinley MD). We are grateful to Timothy Vollmer, M.D., for helpful discussions related to classifying patients by disease course, and to Brian Quaranto for assistance in generating figures for manuscript preparation.

References

- Shawa, M. J., et al. (2001). Knowledge management and data mining for marketing. *Decision Support Systems*, 31(1), 127–137.
- Sung, T. K., Chang, N., & Lee, G. (1999). Dynamics of modeling in data mining: Interpretive approach to bankruptcy prediction. *Journal of Management Information Systems*, 16(1), 63–85.
- Koh, H. C., & Tan, G. (2005). Data mining applications in healthcare. *Journal of Healthcare Information Management*, 19(2), 64–72.
- Koh, H. C., & Leong, S. K. (2001). Data mining applications in the context of casemix. *Annals of the Academy of Medicine, Singapore*, 30(4 Suppl), 41–49.
- 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.
- Martin, M. A., et al. (2006). Mastectomy or breast conserving surgery? Factors affecting type of surgical treatment for breast cancer classification tree approach. *BMC Cancer*, 6, p. 98.
- Gruenewald, T. L., et al. (2008). Diverse pathways to positive and negative affect in adulthood and later life: An integrative approach using recursive partitioning. *Developmental Psychology*, 44, pp. 330–343.
- Radespiel-Troger, M., et al. (2003). Comparison of tree-based methods for prognostic stratification of survival data. *Artificial Intelligence in Medicine*, 28, 323–341.
- Sedrakyan, A., et al. (2006). Recursive partitioning-based preoperative risk stratification for atrial fibrillation after coronary artery bypass surgery. *American Heart Journal*, 151, pp. 720–724.
- Ring, L., Gross, C. R., & McColl, E. (2010). Putting the text back into context: Toward increased use of mixed methods for qualitative research. *Quality of Life Research*, 19(5), 613–615.
- Brod, M., Tesler, L. E., & Christensen, T. L. (2009). Qualitative research and content validity: Developing best practices based on science and experience. *Quality of Life Research*, 18(9), 1263–1278.
- NMSS. (2005). *Multiple sclerosis information sourcebook*. New York, NY: Information Resource Center and Library of the National Multiple Sclerosis Society.
- Tremlett, H., et al. (2010). New perspectives in the natural history of multiple sclerosis. *Neurology*, 74, 2004–2015.
- Trapp, B. D., & Nave, K.-A. (2008). Multiple sclerosis: An immune or neurodegenerative disorder? *Annual Review of Neuroscience*, 31, pp. 247–269.
- Tintore, M. (2009). New options for early treatment of multiple sclerosis. *Journal of the Neurological Sciences*, 277(Suppl 1), S9–S11.
- Brown, M. G., et al. (2007). How effective are disease-modifying drugs in delaying progression in relapsing-onset MS? *Neurology*, 69(15), 1498–1507.
- Torjano, M., et al. (2007). New natural history of interferon-beta-treated relapsing multiple sclerosis. *Annals of Neurology*, 61(4), 300–306.
- Schwartz, C. E., Vollmer, T., & Lee, H. (1999). Reliability and validity of two self-report measures of impairment and disability for MS. North American Research Consortium on Multiple Sclerosis Outcomes Study Group. *Neurology*, 52(1), 63–70.
- Hohol, M. J., Orav, E. J., & Weiner, H. L. (1995). Disease steps in multiple sclerosis: A simple approach to evaluate disease progression. *Neurology*, 45, pp. 251–255.
- Vickrey, B. G., et al. (1995). A health-related quality of life measure for multiple sclerosis. *Quality of Life Research*, 4, pp. 187–206.
- Cella, D. F., et al. (1996). Validation of the functional assessment of multiple sclerosis quality of life instrument. *Neurology*, 47, pp. 129–139.
- Hamilton, B. B., et al. (1987). A uniform data system for medical rehabilitation. In M. J. Fuhrer (Ed.), *Rehabilitation outcomes: An analysis and measurements*. Baltimore: Brooks, pp. 137–147.
- Fisher, J. S., et al. (1999). Recent developments in the assessment of quality of life in multiple sclerosis. *Multiple Sclerosis*, 5, pp. 251–259.
- Filippini, G., et al. (2003). Interferons in relapsing remitting multiple sclerosis: A systematic review. *Lancet*, 361, 545–552.
- Rice, G. P., et al. (2001). Interferon in relapsing-remitting multiple sclerosis. *Cochrane Database of Systematic Reviews*, 4, p. CD002002.

26. Snook, E. M., & Motl, R. W. (2009). Effect of exercise training on walking mobility in multiple sclerosis: A meta-analysis. *Neurorehabilitation and Neural Repair* (in press).
27. Sprangers, M. A. G., & 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), p. 14.
29. 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.
30. 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*, p. 16.
31. Sackett, D. L., & Torrance, G. W. (1978). The utility of different health states as perceived by the general public. *Journal of Chronic Disease*, *31*, pp. 697–704.
32. Ubel, P. A., Loewenstein, G. & Jepson, C. (2003). Whose quality of life? A commentary on exploring discrepancies between health state evaluations of patients and the general public. *Quality of Life Research*, *12*, pp. 599–607.
33. Kurtzke, J. F. (1983). Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). *Neurology*, *33*(11), 1444–1452.
34. Ware, J. E., Jr., Kosinski, M., & Keller, S. D. (1996). A 12-item short-form health survey. *Medical Care*, *34*(3), 220–233.
35. Therneau, T. M., & Atkinson, E. J. (1997). *An introduction to recursive partitioning using the Rpart routines*. Rochester, MN: Mayo Foundation.
36. Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S*. New York: Springer Science & Business Media.
37. Team, R. D. C. (2008). *R: a language, environment for statistical computing*. Vienna, Austria: The R Foundation for Statistical Computing.
38. Abdolell, M., et al. (2002). Binary partitioning for continuous longitudinal data: Categorizing a prognostic variable. *Statistics in Medicine*, *21*, 3395–3409.
39. Venables, W. N., & Ripley, B. D. (1999). *Modern applied statistics with S*, 4th ed. Statistics and Computing. New York: Springer.
40. Ware, J. E., Jr. (1995). The status of health assessment 1994. *Annual Review of Public Health*, *16*, pp. 327–354.
41. King-Kallimanis, B. L., et al. (2011). Using structural equation modeling to detect response shift in performance and quality of life scores of multiple sclerosis patients. *Quality of Life Research* (in press).
42. Ahmed, S., et al. (2011). Using latent trajectory analysis to detect response shift in general health among multiple sclerosis patients. *Quality of Life Research* (in review).
43. Oort, F. J. (2005). Using structural equation modeling to detect response shifts and true change. *Quality of Life Research*, *14*, pp. 587–598.