The relationship between change in subjective outcome and change in disease: a potential paradox

Wietske Kievit · Jos Hendrikx · Peep F. M. Stalmeier ·
Mart A. F. J. van de Laar · Piet L. C. M. Van Riel ·
Eddy M. Adang

Abstract

Background Response shift theory suggests that improvements in health lead patients to change their internal standards and re-assess former health states as worse than initially rated when using retrospective ratings via the then-test. The predictions of response shift theory can be illustrated using prospect theory, whereby a change in current health causes a change in reference frame. Therefore, if health deteriorates, the former health state will receive a better rating, whereas if it improves, the former health state will receive a worse rating.

Objective To explore the predictions of response shift and prospect theory by relating subjective change to objective change.

Methods Baseline and 3-month follow-up data from a cohort of rheumatoid arthritis patients ($N = 197$) starting on TNFα-blocking agents were used. Objective disease change was classified according to a disease-specific clinical outcome measure (DAS28). Visual analogue scales (VAS) for general health (GH) and pain were used as self-reported measures. Three months after starting on anti-TNFα, patients used the then-test to re-rate their baseline health with regard to general health and pain. Differences between then-test value and baseline values were calculated and tested between improved, non-improved and deteriorated patients by the Student *t*-test.

Results At 3 months, 51 (25.9%) patients had good improvement in health, 83 (42.1%) had moderate improvement, and 63 (32.0%) had no improvement or deteriorated in health. All patients no matter whether they improved, did not improve, or even became worse rated their health as worse retrospectively. The difference between the then-test rating and the baseline value was similarly sized in all groups.

Conclusion More positive ratings of retrospective health are independent of disease change. This suggests that patients do not necessarily change their standards in line with their disease change, and therefore it is inappropriate to use the then-test to correct for such a change. If a then-test is used to correct for shifts in internal standards, it might lead to the paradoxical result that patients who do not improve or even deteriorate increase significantly on self-reported health and pain. An alternative explanation for differences in retrospective and prospective ratings of health is the implicit theory of change which is more successful in explaining our results than prospect theory.

Keywords Response shift · Retrospective pre-test · Then-test · Rheumatoid arthritis · Implicit theory of change

W. Kievit (✉) · J. Hendrikx · P. L. C. M. Van Riel
Department of Rheumatology, Radboud University Nijmegen
Medical Centre, PO box 9101, 6500 HB Nijmegen,
The Netherlands
e-mail: w.kievit@reuma.umcn.nl

W. Kievit · P. F. M. Stalmeier · E. M. Adang
Department of Epidemiology, Biostatistics and Health
Technology Assessment, Radboud University Nijmegen Medical
Centre, Nijmegen, The Netherlands

P. F. M. Stalmeier
Department of Radiation Oncology, Radboud University
Nijmegen Medical Centre, Nijmegen, The Netherlands

M. A. F. J. van de Laar
Department of Rheumatology, Medisch Spectrum Twente &
University Twente, Enschede, The Netherlands

List of abbreviations

CRP	C-reactive protein
DAS28	Disease activity score for 28 joint counts
DMARD	Disease modifying anti-rheumatic drug

ESR	Erythrocyte sedimentation rate
EULAR	European league against rheumatism
GH	General health
HAQ	Health assessment questionnaire
SW28	28 joint count for swelling
TEN28	28 joint count for tenderness
TNFα	Tumour necrosis factor alpha
VAS	Visual analogue scale

Introduction

When a physician sees a patient, the most frequently asked question will probably be "How do you feel today concerning your general health (or pain)?". This question can be formally measured using a Visual Analogue Scale (VAS) for general health (GH) or pain, which is considered a feasible, valid and reliable method [1–3]. In the field of rheumatology, indices based on patient reported outcomes, among which the VAS GH and VAS pain, are suggested to use for systematically monitoring patients on therapy effectiveness [4]. In general, to compare assessments over time and between individuals, it is a prerequisite of a measurement that the patient's concept of the subject of assessment is the same in all situations. However, there are many examples of patients reporting their health state as stable over time, despite improvements in their objective health status [5–12], which challenge the use of self-reporting outcomes. It has been shown in a variety of disorders that patients whose health improved re-assessed their former health state as worse than initially rated [5–13]. Furthermore, in cancer patients the relative importance of different domains of quality of life varied over time [14]. Finally, there are many studies that demonstrate that patients value a certain health state differently than individuals not in that particular health state [7, 15–18], for example that healthy people perceive the general health of patients suffering from end-stage renal disease as worse than the patients themselves [7].

In response shift theory, these findings are interpreted [19] as follows: patients do not rate their own health in reference to an absolute standard but in reference to a relative standard. In health-related subjective measures, this relative standard is tentatively taken as the patient's own current health state. However, current health, and therefore an individual's standard, may change over time [19–21]. For example, when a patient is asked whether or not he or she has improved, the fact that the patient refers to his/her current health can lead to different valuations of the same objective health state at different moments in time. Some authors suggest that self-reported outcomes should be corrected for changes in relative standards [22, 23] by, for instance, a then-test [22]. By using a then-test, patients are asked to re-rate a former health state from their current perspective. Because the former health state and the follow-up health state are then rated from the same perspective, the change in health (then-minus-post) is corrected for a possible change in internal standards.

Support for the then-test as a correction for changes in relative standards can be found in prospect theory [24, 25]. Prospect theory was originally developed in the context of gambling for money and is a theory about decisions under risk [26]. Prospect theory defines money outcomes as gains and losses relative to a reference point. Different reference frames are assumed to shift the reference point along the outcome dimension, altering the location of an S-shaped value function [25, 27]. In valuing health, the reference point of the S-shaped value functions can be described as the individual's current health, as described by Treadwell and Lenert [25]. Figure 1 shows two such S-shaped value functions for two different health states in time, A and B. The x-axis represents a person's health state, and the y-axis represents subjective values attributed to health states. When a patient improves in health (moves from A to B), his value function shifts to B. According to the shifted S-shaped value function, the former state is rated worse (A') than previously (A). When a patient deteriorates in health (moves from B to A), his value function shifts to the left. According to that function, the former health state will receive a better rating (B') than previously (B). When a health state remains stable, no shift in the value function is expected. Prospect theory therefore provides a clear illustration of the predictions of response shift theory with respect to recalibration response shifts. Hence, this study explores the predictions by prospect theory to explain response shift.

However, it should be noted that alternative predictions of outcomes of the then-test can be found in, for example, Norman's discussion of implicit theory of change [28]. This theory originates in psychological research [29] and states that individuals might be unable to accurately remember a former situation but instead extrapolate backwards from their present state invoking an implicit theory of change [28, 29]. A treatment received could trigger an implicit theory of positive change. Working back from their present health status, such an implicit positive change would lead patients to re-rate their baseline health status (then-test) as worse than their initial baseline rating independent of the actual effectiveness of the treatment.

That patients re-rate (by means of a then-test) their former health state as worse than previously after an improvement in health has been shown several times in the medical field [5–13]. However, there is very little evidence

Fig. 1 Two S-shaped value functions for two different health states in time. Modified figure from the article by Treadwell and Lenert [25]. Response shift: the difference between A'-A or B' and B. Prospective change: difference between A and B. Then-minus-post change: difference between A-B' or B-A'
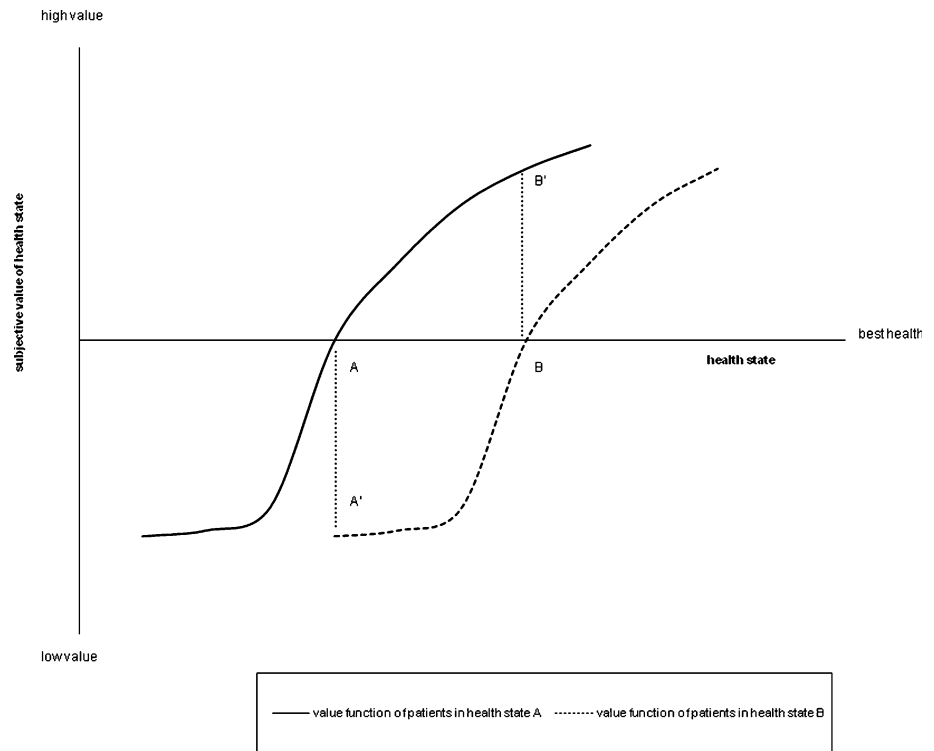
in the literature for predictions in patients who deteriorate in health compared to patients who improve in health. Response shift and prospect theory both predict that shifts in internal standards will be bi-directional: retrospective ratings of a former health state will be worse in patients who improve (health state A' worse than previously (A)) and better in patients who deteriorate (health state B' rated better than previously (B)). However, implicit theory of change would predict that both patients who objectively improve or deteriorate will rate their former health state as worse than previously rated if patients perceive there should have been an improvement.

The aim of this study was to explore the predictions from response shift and prospect theory by relating subjective change to objective change. This was done by comparing retrospective scores in patients with severe and prolonged rheumatoid arthritis who either deteriorated or improved after treatment with TNF-blocking agents.

Methods

Study site and subjects

Our hypothesis was tested in a population of chronically ill rheumatoid arthritis patients receiving anti-TNFα treatment. This patient population was chosen, because of the chronic character of the disease, which ensures the possibility of adaptation by patients to their imperfect health state, and thereby allows the reference level to shift. Anti-TNFα is a promising treatment to establish a significant decrease in disease activity [30–32]. In addition, this population was chosen because of the possibility of categorizing patients by disease activity based on a validated objective measure of disease activity [33].

Data from a prospective registry were used. Since February 2003, all rheumatoid arthritis patients, from 11 Dutch hospitals, who started on an anti-TNFα agent (either adalimumab, etanercept or infliximab) for the first time, were included in this registry [34]. The primary aim of this registry is to evaluate and monitor anti-TNFα treatment in rheumatoid arthritis patients. Since January 2004, then-tests have been included in this registry.

Study design and outcome measures

Patients' baseline characteristics were registered at the start of the first anti-TNFα treatment. These characteristics included age, gender, weight, disease duration, rheumatic factor status, the presence of one or more erosions in hand or feet and number of previous DMARDs (disease modifying anti-rheumatic drugs) used. In addition to these characteristics, the three main measures for this study, disease activity, the patient's self-perceived general health and self-perceived pain, were registered at baseline and at a 3-month follow-up assessment. The 3-month follow-up

assessment was chosen because at that point a clinical decision about the medication policy is made on the basis of the underlying disease state alteration (with the DAS28).

Disease activity was assessed using a modified Disease Activity Score 28 (DAS28), which is a statistically derived index combining the following variables: the 28 joint count for swelling (SW28), the 28 joint count for tenderness (TEN28) and the Erythrocyte Sedimentation Rate (ESR), with the ESR having the largest weight in the algorithm [35]. In order to minimize subjective interpretation of disease activity in our study, the VAS general health was left out of the original DAS28 algorithm. As a valid alternative to ESR, C-reactive Protein (CRP) scores were used in case of missing ESR values [36]. The research nurses were trained in giving a standardized amount of pressure (defined as causing whitening of the examiners nail beds) on the joints to measure tenderness and in using a standardized grip for assessing swelling [36]. For the purpose of describing the population, we also administered the disability index from the Health Assessment Questionnaire (HAQ-DI). The HAQ-DI is a self-assessed questionnaire asking about the ability of patients to perform several daily activities over the past week [37, 38]. The HAQ-DI is well validated and provides information on disease activity as well as joint damage [39, 40]. The questionnaire provides a score between zero and three, where a higher score indicates more functional impairment.

The patients' self-perceived general health and pain were assessed on a 100-mm Visual Analogue Scale (VAS GH and VAS pain), which are feasible, valid and reliable methods to measure these constructs [1–3]. Patients were asked to rate their health or pain by placing a vertical line on a horizontal line, indicating the perceived amount of health or pain. This line ranged from zero to one hundred, where a score of zero indicates "worst imaginable general health" or "extreme pain", and a score of one hundred indicates "best imaginable general health" or "no pain", respectively. Patients were asked to rate their current general health at baseline and at 3 month follow-up. At 3 months, the patients were also asked to retrospectively re-rate the baseline general health and pain by means of a then-test [22, 41]. Patients were specifically instructed not to try and remember what they scored at baseline (recall), but to re-assess their health and pain at baseline from their current perspective. We chose to use the then-test for assessing response shift, because it is easy to use and it is most frequently used by others in longitudinal studies [42]. The actual DAS28 and the response category were calculated in the dataset afterwards and were, therefore, not known to the patient at the moment that he or she completed the scales and the then-test.

Statistical analysis

Descriptives of baseline characteristics were determined. Baseline values were subtracted from then-test score. A negative difference indicated that retrospective scores are worse than baseline scores (Fig. 1: A'-A), and a positive difference indicated that retrospective scores are better than baseline score (Fig. 1: B'-B). Changes over time in VAS scores were calculated as either prospective change or then-minus-post change. The prospective change was calculated by subtracting baseline scores from 3-month follow-up scores (Fig. 1: difference between A and B). A positive prospective change was interpreted as an improvement in health and a negative prospective change as deterioration in health. Post-minus-then change was calculated by subtracting then-test scores from the 3-month follow-up scores (Fig. 1: A-B' or B-A'). One can consider the post-minus-then change as a prospective change corrected for a shift in internal standards (then-test score at 3 months referring to baseline). A positive post-minus-then change suggests a perceived improvement and a negative post-minus-then change suggests a perceived deterioration.

To test the hypothesis that patients who deteriorated retrospectively rate a former health state as better and patients who improved retrospectively rate a former have state as worse, patients were divided into three groups (non-responders, moderate responders and good responders to treatment) depending on their objectively determined European League Against Rheumatism (EULAR) response status at the 3-month follow-up moment. The EULAR criteria were used because these have shown good construct, criterion and discriminant validity [43]. The EULAR criteria are based on the modified DAS28 and combine prospective change and absolute level of attained disease activity (Table 1). By this definition, the non-response group is a mixture of patients who do not improve significantly and patients who deteriorate. Additionally,

Table 1 EULAR response criteria (good, moderate, non-response)

Present DAS28	Change in DAS28		
	>1.2	>0.6 en ≤ 1.2	≤0.6
DAS28 ≤ 3.2 (low)	Good	Moderate	Non-response
3.2 < DAS28 ≤ 5.1 (moderate)	Moderate	Moderate	Non-response
DAS28 > 5.1 (high)	Moderate	Non-response	Non-response

Disease activity score 28 (DAS28) is a continuous score ranging from 0 to 10 in which 0 represents no disease activity and 10 present high disease activity. Thresholds as used in this table are validated thresholds for low, moderate or high disease activity [43]

patients who deteriorated significantly, defined as at least one population standard deviation (0.6 DAS28) deterioration (negative prospective change), were analysed as a subgroup.

Baseline characteristics were tested for equivalency between the groups by means of Chi-square tests, independent samples *t*-tests and Mann–Whitney U tests. The mean difference between then-test scores and baseline scores and the mean changes in VAS general health and VAS pain were tested for statistical significance using one-sample *t*-tests and were compared between groups using independent samples *t*-tests. All analyses were performed using the statistical software package SPSS 14.0.2 (SPSS Inc., Chicago, IL.)

Results

Between January 2004 and December 2006, 212 patients had completed 3-month follow-up data. Fifteen patients were excluded from the analysis due to missing DAS28 and/or VAS data. Baseline characteristics of excluded patients did not differ from included patients. Thus, 197 patients remained eligible for the analyses: 51 (25.9%) patients were classified as good responders to the anti-TNFα therapy, 83 (42.1%) as moderate responders, and 63 patients (32.0%) were classified as non-responders to therapy. In this last group, 11 patients deteriorated significantly and were analysed as a subgroup. On the DAS28, responders had a mean improvement of 2.2 points (ranging from 1.2 to 4.5), moderate responders had a mean improvement of 1.5 points (ranging from 0.6 to 4.2), and non-responders had no improvement (mean −0.05, ranging from −3.0 to 1.1). Table 2 shows the baseline characteristics of clinically responding and non-responding patients. These data show that clinically classified responders did not differ from non-responders at baseline, except on the clinical outcome measures HAQ and DAS28. This was to be expected, because the EULAR response criteria combine change and an absolute level of attained DAS28 and,

Table 2 Baseline characteristics of patients grouped by response classification

	Good responder N = 51	Moderate responder N = 83	Non-responder N = 63	P-value
Age	51.5 (13.0)	54.0 (12.5)	54.9 (15.7)	0.427
HAQ	1.2 (0.6)	1.5 (0.6)	1.3 (0.7)	0.032
DAS28	4.5 (0.9)	5.5 (0.9)	4.6 (1.3)	<0.001
Disease duration*	6.0 (11.8)	6.0 (11.0)	6.0 (8.0)	0.896
Number of previous DMARDs*	3.0 (2.0)	3 (2.0)	3.0 (2.0)	0.272
% female	71.4	78.3	62.9	0.126
% with ≥one erosion	76.6	61.7	63.3	0.202
% positive rheumatoid factor	70.8	77.1	71.0	0.626

Values are presented as mean (standard deviation) unless stated otherwise. * median (interquartile range). *Abbreviations*: *HAQ* Health assessment questionnaire; *DAS*28 disease activity score of 28 joints; *DMARD* disease modifying anti-rheumatic drugs

thus, baseline DAS28. The HAQ is strongly correlated with the DAS28.

Prospective change

Table 3 and the Figs. 2 and 3 show the changes in VAS scores on both general health (Fig. 2, solid line) and pain (Fig. 3, solid line). On average, clinical objective responders (good and moderate) to therapy scored an average of 31.0 points improvement on the VAS general health, which was a statistically significant improvement and a significantly bigger improvement than the average 9.3 points improvement that was reported by non-responders. On the VAS pain, good and moderate responders scored a significant mean improvement of 31.8 points and 30.9 points, respectively, whereas non-responders reported an insignificant improvement of 6.0 points. Patients who deteriorated according to their modified DAS28 scores deteriorated also according to

Table 3 The prospective change (3 months minus baseline) and the response shift (then-test minus baseline value) for the VAS general health (GH) and pain for responders and non-responders

	Good responder (N = 51)	Moderate responder (N = 83)	Non-responder (N = 63)	P-value between groups
Prospective change VAS GH	31.0 (22.7; 39.3)	31.0 (25.7; 36.2)	9.3 (2.5;16.1)	<0.0001
Prospective change VAS Pain	31.8 (24.8; 38.8)	30.9 (24.9; 36.8)	6.0 (−0.4;12.4)	<0.0001
Response shift VAS GH	−9.7 (−16.2; −3.2)	−3.3 (−7.1; 0.6)	−6.8 (−12.6; −0.9)	0.219
Response shift VAS Pain	−10.7 (−18.3; −3.2)	−4.6 (−8.7; −0.5)	−10.9 (−16.4; −5.5)	0.148

Values are presented as mean (95% CI)

* mean difference between responders and non-responders
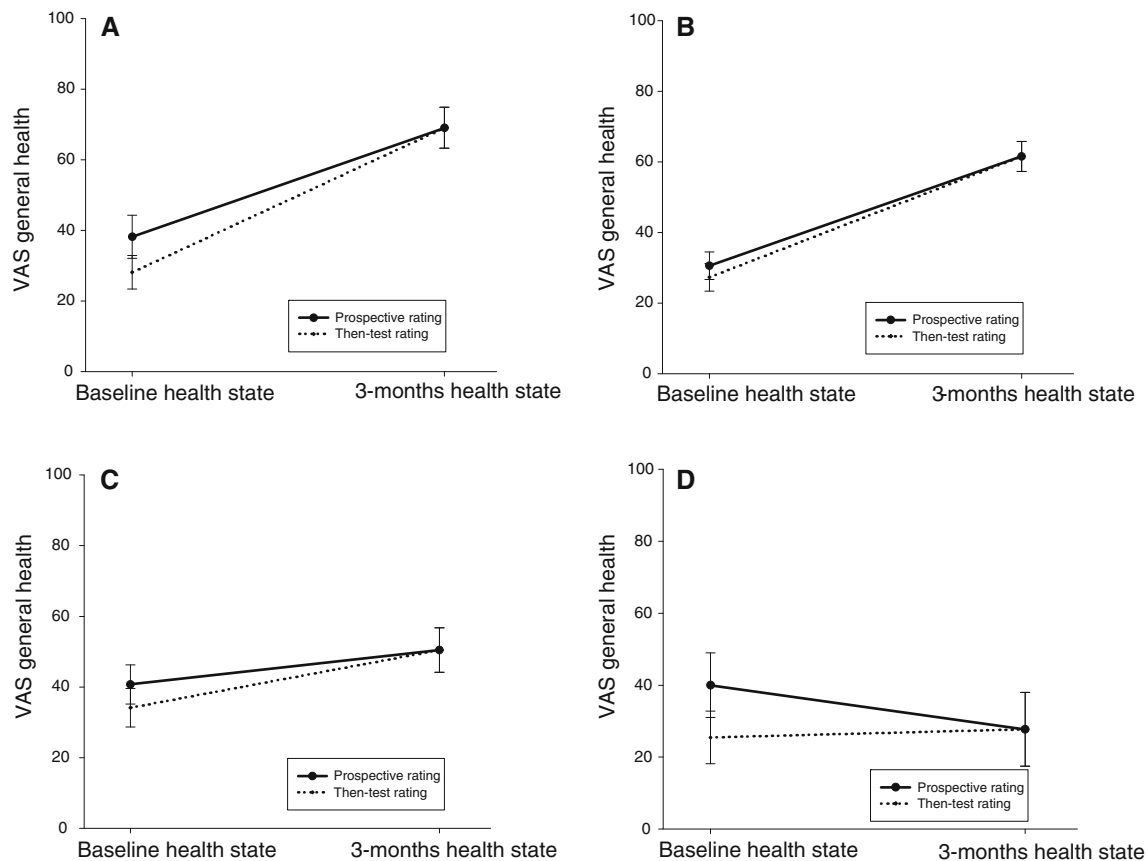
Fig. 2 Changes in VAS general health scores and then-test values of good responders (**a**), moderate responders (**b**), non-responders (**c**) and deteriorators (**d**). Note that a score of zero indicates a worst imaginable general health state, and a score of one hundred indicates best imaginable general health

their VAS GH (mean 12.3 points) and pain scores (mean 17.7 points).

Retrospective scores

Objectively classified responders and non-responders showed significant worse retrospective scores compared to baseline values on both scales (Figs. 2 and 3, dotted lines). The identified difference between the retrospective scores and the baseline scores were of equal size for responders and non-responders (Table 3). The 11 deteriorated patients showed a significant difference between retrospective score and baseline values on general health of −14.5 (95% CI −28.3; −0.8) and on pain of −19.1 (95% CI −36.9; −1.3).

Then-minus-post scores

The calculated mean then-minus-post scores are also shown in Figs. 2 and 3. When applying the then-minus-post values, clinical good and moderate responders to therapy averaged a 40.7-point and 34.2-point improvement on VAS general health, respectively. This was statistically

significant ($P < 0.0001$) and significantly more than the average 16.1-point improvement of the non-responders ($P < 0.0001$). Outcomes of the VAS Pain showed similar results. Good and moderate responders averaged a 42.6-point and 35.2-point improvement, respectively, compared to a significantly lower ($P < 0.0001$) improvement of 16.9-points for non-responders. The deteriorated patients showed stable disease activity when applying the then-minus-post score for the VAS GH (mean 2.3 points) and for the VAS pain (mean 1.4 points).

Discussion and conclusion

This study showed that both patients who improved (responders) or stayed the same/deteriorated (non-responders) rated their baseline health state worse than actually rated at baseline, when asked 3 months later with a then-test. Furthermore, paradoxical results occurred when the then-test was applied (then-minus-post) for the purpose of correcting for a shift in internal standards. For clinically stable or deteriorating patients, an improvement in health
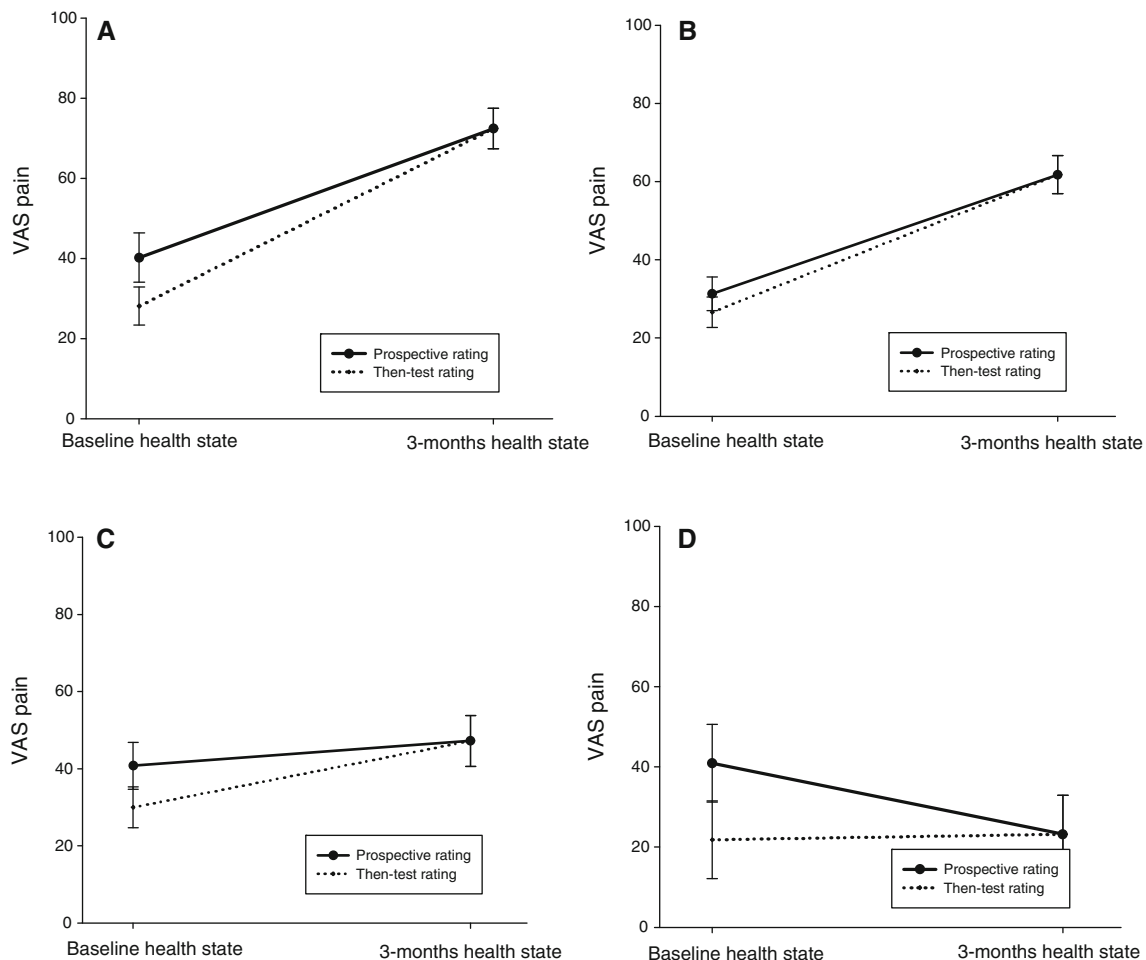
Fig. 3 Changes in VAS Pain scores and then-test values of good responders (**a**), moderate responders (**b**), non-responders (**c**) and deteriorators (**d**). Note that a score of zero indicates extreme pain, and a score of one hundred indicates no pain

would be inferred when using then-minus-post ratings when compared to using only prospective ratings (Figs. 2 and 3). The fact that patients who stayed the same or deteriorated also reported worse health retrospectively conflicted with the predictions from response shift theory and prospect theory.

The negative response shifts in patients who improved were in line with the predictions derived from prospect theory and complied with the large amount of literature reporting on negative response shifts [5–12]. However, the negative response shifts in patients who stayed the same or deteriorated, conflicted with the prediction from prospect theory. Only two previous studies assessed patients who improved, who did not improve or who deteriorated in health. Results from one study were in line with prospect theory [44], while findings from the second study were not [5]. Ahmed and colleagues [5] also used objective criteria to determine the direction of disease change. In agreement with our study, they found mean

response shifts in the same direction in 196 participants. These response shifts were independent of the direction of the disease change that had occurred. Janssen et al. [44] investigated response shift in 46 patients where the disease change was defined with a subjective change question. In contrast to Ahmed's and our findings, their response shifts were dependent on the direction of the disease changes. The explanation for these conflicting findings may be the crucial difference in the way that disease change was defined: objective (Ahmed's and our study) versus subjective (Janssen's study).

The aforementioned implicit theory of change seems to explain our results more accurately than prospect theory and also the conflicting results in the literature. In our study and the study by Ahmed et al. [5], non-responders showed no objective disease change, but the treatment received could have triggered an implicit positive change. Working back from their health status, this implicit positive change would lead patients to re-rate their baseline health status

(then-test) as worse than their initial baseline rating. In the study by Jansen et al. [44], patients self-reported a subjective deterioration in disease. Working back from their health status, they could have applied this implicit negative change to re-rate the baseline health status with the then-test as better than their initial baseline rating. In summary, an implicit positive change of patients in our and Ahmed's study and the implicit negative change of patients in Janssen's study can explain the conflicting retrospective scores more accurately than response shift or prospect theory. It has to be mentioned that an implicit positive change as described above could also have resulted from what is known as a placebo response, but either way change results in patients re-rating their baseline health status (then-test) as worse than their initial baseline rating.

The prediction of response shift theory, with respect to the fact that the internal standards by which patients rate their own health change over time, has clinical implications. This is especially true when subjective measures are used to monitor therapy effectiveness as was recently suggested in the field of rheumatology [4]. It can be suggested that physicians can try to correct for a shift in internal standards by asking patients to rerate their health state before they started therapy and rate their current health. By comparing those two ratings, therapy effectiveness can be evaluated. However, our results suggest that rather than changing internal standards, implicit theory of change is applied by patients to construct the value of a former health state. Therefore, if retrospective ratings are used, this will lead to the paradoxical result that patients who deteriorate or stay the same increase significantly on self-reported health and pain. For a clinician, this would make it more difficult to justify treatment change where from a more objective point of view or from the perspective of prospective self ratings, it would have been advisable.

Lenert et al. [45] did empirical research in health care to test predictions based on prospect theory. Patients from primary care practices with various medical illnesses and depression rated different hypothetical health states. They showed data that were consistent with the predictions from prospect theory: utility functions for health were "S" shaped and differed across levels of health [45]. The differences between their study design (cross-sectional) and our study design (longitudinal) may explain the differences in results. Their subjects rated hypothetical health states that were worse or better than their own health states, and they did this during one session. In our study, the patients had an actual change in their own health to either a worse or a better health state, and then rerated their baseline health with a then-test.

Another explanation for the fact that we did not confirm prospect theory may lie in the fact that the then-test may

not be the appropriate instrument to evaluate a former health state and subsequently to measure response shift. Although the then-test is most frequently used as an instrument to measure response shifts [42], it has the limitation that it can be subject to recall bias [23], and it is dependent on the cognitive function of the subjects. Furthermore, it has to be mentioned that the results of this study apply to visual analogue scales, and the results cannot be generalized to other quality of life measures like the Short Form 36 (SF36) or EuroQol 5d (EQ5D) without further research. It might be possible that prospect theory fits the predicted outcomes measured by the EQ-5D or SF-36. On the other hand, response shifts as a result of changes in internal standards may not occur in measures like the SF36 or EQ5D, but response shift as a result of changes in values or re-conceptualization may occur [19]. Currently, there is no evidence that prospect theory predicts response shift as a result of changing values or conceptualizations, and it was not the focus of our study to deliver such evidence.

Some shortcomings concerning this study have to be mentioned. First, by using the EULAR criteria to define disease change (Table 1), the non-response group was a mixture of patients who did not improve significantly and patients who deteriorated. A subgroup of 11 patients who deteriorated on the objective measure DAS28 was analysed. These 11 deteriorated patients showed, consistently with the other groups of patients, a significant negative response shift on general health and pain, rating their baseline health state worse than initially rated. Such a subgroup analysis might be underpowered. Furthermore, the DAS28 is used as an objective measure, whilst it can be subject to interpretation especially in the components 'joint pain' and 'joint swelling'. To minimize subjective interpretation in these components, all research nurses were trained twice per year for performing joint counts. Moreover, a modified DAS28 was calculated using the most objective components, the ESR and the joint counts, of which the ESR has the larger weight in the combined score. Another limitation is that the patients' memory or perceptions of change were not assessed, and therefore the implicit perception of change could not be assessed.

In conclusion, the similar direction and magnitude of retrospective ratings in objectively defined improved and non-improved patients suggests that patients do not necessarily change their standards in line with their disease change. If a then-test is used to correct for shifts in internal standards, it might lead to the paradoxical result that patients who do not improve or even deteriorate increase significantly on self-reported health and pain making it more difficult for the clinician to justify treatment change. An alternative explanation for differences in retrospective

and prospective ratings of health is the implicit theory of change which is more successful in explaining our results than prospect theory.

References

1. Badia, X., Monserrat, S., Roset, M., & Herdman, M. (1999). Feasibility, validity and test-retest reliability of scaling methods for health states: The visual analogue scale and the time trade-off. *Quality of Life Research, 8*, 303–310.
2. Dworkin, R. H., Turk, D. C., Revicki, D. A., Harding, G., Coyne, K. S., Peirce-Sandner, S., et al. (2009). Development and initial validation of an expanded and revised version of the Short-form McGill Pain Questionnaire (SF-MPQ-2). *Pain, 144*, 35–42.
3. Green, C., Brazier, J., & Deverill, M. (2000). Valuing health-related quality of life. A review of health state valuation techniques. *Pharmacoeconomics, 17*, 151–165.
4. Pincus, T., Yazici, Y., Bergman, M., Maclean, R., & Harrington, T. (2007). A proposed continuous quality improvement approach to assessment and management of patients with rheumatoid arthritis without formal joint counts, based on quantitative routine assessment of patient index data (RAPID) scores on a multidimensional health assessment questionnaire (MDHAQ). *Best Practice & Research. Clinical Rheumatology, 21*, 789–804.
5. 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*, 561–570.
6. Nieuwkerk, P. T., Tollenaar, M. S., Oort, F. J., & Sprangers, M. A. (2007). Are retrospective measures of change in quality of life more valid than prospective measures? *Medical Care, 45*, 199–205.
7. Postulart, D., & Adang, E. M. (2000). Response shift and adaptation in chronically ill patients. *Medical Decision Making, 20*, 186–193.
8. Razmjou, H., Yee, A., Ford, M., & Finkelstein, J. A. (2006). Response shift in outcome assessment in patients undergoing total knee arthroplasty. *Journal of Bone and Joint Surgery. American Volume, 88*, 2590–2595.
9. Rees, J., Clarke, M. G., Waldron, D., O'Boyle, C., Ewings, P., & MacDonagh, R. P. (2005). The measurement of response shift in patients with advanced prostate cancer and their partners. *Health Quality Life Outcomes, 3*, 21.
10. Rees, J., Waldron, D., O'Boyle, C., Ewings, P., & MacDonagh, R. (2003). Prospective vs retrospective assessment of lower urinary tract symptoms in patients with advanced prostate cancer: The effect of 'response shift'. *BJU International, 92*, 703–706.
11. Ring, L., Hofer, 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 Life Outcomes, 3*, 55.
12. Smith, D. M., Sherriff, R. L., Damschroder, L., Loewenstein, G., & Ubel, P. A. (2006). Misremembering colostomies? Former patients give lower utility ratings than do current patients. *Health Psychology, 25*, 688–695.
13. ten Klooster, P. M., Drossaers-Bakker, K. W., Taal, E., & van de Laar, M. A. F. J. (2006). Patient-perceived satisfactory improvement (PPSI): Interpreting meaningful change in pain from the patient's perspective. *Pain, 121*, 151–157.
14. Visser, M. R., Oort, F. J., & Sprangers, M. A. (2005). Methods to detect response shift in quality of life data: a convergent validity study. *Quality of Life Research, 14*, 629–639.
15. Sackett, D. L., & Torrance, G. W. (1978). The utility of different health states as perceived by the general public. *Journal of Chronic Diseases, 31*, 697–704.
16. Boyd, N. F., Sutherland, H. J., Heasman, K. Z., Tritchler, D. L., & Cummings, B. J. (1990). Whose utilities for decision analysis? *Medical Decision Making, 10*, 58–67.
17. Dolan, P. (1996). The effect of experience of illness on health state valuations. *Journal of Clinical Epidemiology, 49*, 551–564.
18. De Wit, G. A., Busschbach, J. J. V., & De Charro, F. T. (2000). Sensitivity and perspective in the valuation of health status: Whose values count? *Health Economics, 9*, 109–126.
19. 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*, 1507–1515.
20. Allison, P. J., Locker, D., & Feine, J. S. (1997). Quality of life: A dynamic construct. *Social Science and Medicine, 45*, 221–230.
21. Adang, E. M. (2001). A comment on "response shift and adaptation in chronically ill patients" by Postulart and Adang. *Medical Decision Making, 21*, 423–424.
22. 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*, 1531–1548.
23. Schwartz, C. E., Sprangers, M. A. G., Carey, A., & Reed, G. (2004). Exploring response shift in longitudinal data. *Psychology and Health, 19*, 51–69.
24. Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*, 453–458.
25. Treadwell, J. R., & Lenert, L. A. (1999). Health values and prospect theory. *Medical Decision Making, 19*, 344–352.
26. Kahneman, D., & Tversky, A. (1979). Prospect theory—analysis of decision under risk. *Econometrica, 47*, 263–291.

27. Stalmeier, P. F. M., & Bezembinder, T. G. G. (1999). The discrepancy between risky and riskless utilities: A matter of framing? *Medical Decision Making, 19*, 435–447.

28. Norman, G. (2003). Hi! How are you? Response shift, implicit theories and differing epistemologies. *Quality of Life Research, 12*, 239–249.

29. Ross, M. (1989). Relation of implicit theories to the construction of personal histories. *Psychological Review, 96*, 341–357.

30. Blumenauer, B., Judd, M., Cranney, A., Burls, A., Coyle, D., Hochberg, M., Tugwell, P., & Wells, G. (2003). Etanercept for the treatment of rheumatoid arthritis. *Cochrane Database System Review*, CD004525.

31. Navarro-Sarabia, F., riza-Ariza, R., Hernandez-Cruz, B., & Villanueva, I. (2006). Adalimumab for treating rheumatoid arthritis. *Journal of Rheumatology, 33*, 1075–1081.

32. Blumenauer, B., Judd, M., Wells, G., Burls, A., Cranney, A., Hochberg, M., & Tugwell, P. (2002) Infliximab for the treatment of rheumatoid arthritis. *Cochrane Database System Review*, CD003785.

33. van Gestel, A. M., Anderson, J. J., van Riel, P. L., Boers, M., Haagsma, C. J., Rich, B., et al. (1999). ACR and EULAR improvement criteria have comparable validity in rheumatoid arthritis trials. American College of Rheumatology European League of Associations for Rheumatology. *The Journal of Rheumatology, 26*, 705–711.

34. Kievit, W., Fransen, J., Oerlemans, A. J., Kuper, H. H., van der Laar, M. A., de Rooij, D. J., et al. (2007). The efficacy of anti-TNF in rheumatoid arthritis, a comparison between randomised controlled trials and clinical practice. *Annals of the Rheumatic Diseases, 66*, 1473–1478.

35. Prevoo, M. L. L., VantHof, M. A., Kuper, H. H., Vanleeuwen, M. A., VandePutte, L. B. A., & Vanriel, P. L. C. M. (1995). Modified Disease-Activity Scores That Include 28-Joint Counts - Development and Validation in A Prospective Longitudinal-Study of Patients with Rheumatoid-Arthritis. *Arthritis and Rheumatism, 38*, 44–48.

36. Van Riel, P. L. C. M., Fransen, J., & Scott, D. L. (2004) *EULAR handbook of clinical assessments in rheumatoid arthritis*. Alphen aan de Rijn, The Netherlands: Van Zuiden Communications B.V.

37. Fries, J. F., Spitz, P. W., & Young, D. Y. (1982). The dimensions of health outcomes—the health assessment questionnaire, disability and pain scales. *Journal of Rheumatology, 9*, 789–793.

38. Zandbelt, M. M., Welsing, P. M. J., van Gestel, A. M., & Van Riel, P. L. C. M. (2001). Health assessment questionnaire modifications: Is standardisation needed? *Annals of the Rheumatic Diseases, 60*, 841–845.

39. Marra, C. A., Woolcott, J. C., Kopec, J. A., Shojania, K., Offer, R., Brazier, J. E., et al. (2005). A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. *Social Science and Medicine, 60*, 1571–1582.

40. Russell, A. S., Conner-Spady, B., Mintz, A., Mallon, C., & Maksymowych, W. P. (2003). The responsiveness of generic health status measures as assessed in patients with rheumatoid arthritis receiving infliximab. *Journal of Rheumatology, 30*, 941–947.

41. Howard, G. S., Schmeck, R. R., & Bray, J. H. (1979). Internal invalidity in studies employing self-report instruments: A suggested remedy. *Journal of Educational Measurement, 16*, 129–135.

42. 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*, 1533–1550.

43. van Gestel, A. M., Prevoo, M. L., 't Hof, M. A., van Rijswijk, M. H., van de Putte, L. B., & van Riel, P. L. (1996). Development and validation of the European league against rheumatism response criteria for rheumatoid arthritis. Comparison with the preliminary American college of rheumatology and the World Health Organization/International league against Rheumatism criteria. *Arthritis and Rheumatism, 39*, 34–40.

44. 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*, 603–615.

45. Lenert, L. A., Treadwell, J. R., & Schwartz, C. E. (1999). Associations between health status and utilities implications for policy. *Medical Care, 37*, 479–489.