

Response Shift in Measuring Health-related Quality of Life: Concepts, Definitions, and Challenges

Bayan Hosseini¹, Saharnaz Nedjat^{2*}, Kazem Zendehdel³, Reza Majdzadeh²
Ali Montazeri⁴

ABSTRACT

As a major consequence of medical interventions, quality of life (QoL) is of great importance for patients suffering from refractory diseases, particularly cancer. The unbiased measurement of changes in QoL is thus crucial in such cases. A prevalent bias related to QoL research is the 'response shift' (RS) phenomenon. This review article aims to define RS and the challenges in measuring it. In addition, it addresses the methodological approaches used to measure this bias in observational and clinical studies. Response shift refers to changes in one's health condition as a result of changes in the meaning of one's self-evaluation. These changes result when the patient faces his/her new conditions and may be reflected as greater as or smaller than they actually are. The present article describes the individualized methods, the preference-based methods, the structural equation modeling and the then-test method used for evaluating RS, and discusses their applications. Finally, by comparing these methods, it concludes that the simplest and most efficient approach for evaluating RS is the then-test approach. By emphasizing that these methods should be applied in clinical studies, the present article describes the most important methods for evaluating RS. The effect of RS has been neglected in the majority of QoL studies. We therefore recommend taking into account the effect of RS in the interpretation of QoL changes in longitudinal studies.

Keywords: *Response Shift, Then test, Quality of life.*

2

1. MSc. Student of Epidemiology, Department of Epidemiology and Biostatistics, School of Public Health and Institute of Public Health Research, Tehran University of Medical Sciences, Tehran, Iran.

2. Professor, Department of Epidemiology and Biostatistics, School of Public Health, Knowledge Utilization Research Center, Tehran University of Medical Sciences, Tehran, Iran.

3. Assistant Professor, Cancer Research Center, Tehran University of Medical Sciences, Tehran, Iran.

4. Professor, Mental Health Research Department, Institute for Health Sciences Research, ACECR, Tehran, Iran.

***Corresponding Author:**

Saharnaz Nedjat,
Associate Professor, Epidemiology and Biostatistics Department, School of Public Health, Knowledge Utilization Research Center, Tehran University of Medical Sciences, Tehran, Iran.
Email: nejatsan@sina.tums.ac.ir



Introduction

Quality of life (QoL) is increasingly being considered as one of the major outcomes of medical healthcare, both in clinical decision makings and health policy makings. One of the main objectives of disease management is maintaining patients' QoL as a significant outcome of medical and relief intervention.¹ It is of even greater importance in patients whose treatment is not curative, but rather palliative.² Therefore, it is essential to ensure the accurate measurement of QoL. Precise measurement can differentiate between various subgroups of patients and make their QoL changes reliable over time.³ As defined by the World Health Organization, quality of life is a person's interpretation of one's own conditions in life with regard to the culture and the value systems one belongs to and the goals, expectations, standards and concerns that one has set for oneself.⁴ Therefore, QoL involves a vast range of concepts and is well beyond physical well-being. An important aspect of QoL is its subjective and multidimensional nature; therefore, we can evaluate it with both subjective and objective questions.^{5,6} People's values differ because they are different from each other. This difference might invalidate measurements related to the individuals' internal standards. Features such as skills, cognitions, emotions, temperament, attitude and QoL are generally measured by self-reports; The respondents filling out these self-evaluation questionnaires might consider frameworks that are different from the reference response, hence leading to bias in resultant measurements.⁷ Every day, physicians encounter patients who wish to be relieved of their unbearable pain and suffering, and demand that their lives be terminated upon disclosure about their cancer or refractory disease diagnosis. At this point in man's life, life seems unendurable with all the pain, limited motion and the proximity of death.⁸ Nevertheless, patients adapt themselves to their new conditions and define new standards for their health states after a certain period of time into the disease. For instance, when a patient is informed about the malignancy of his osteosarcoma, he claims that if his bone tumor deprives him of the ability to walk, going on with life would become meaningless and he'd rather die. But when the patient becomes wheelchair-bound he confesses that life is still precious to him. At this point the patient

claims that life will lose its meaning if he becomes bedridden, and that he'd rather die if he ever reaches this point. Nonetheless, when he moves forward to the next stage of his disease and becomes bedridden, the value of life remains intact. It is at this stage of disease that the patient has come to resent death to such a degree that he no longer prefers to die—in spite of all his pain and suffering. Therefore, the patient adjusts himself to the new conditions at all the different stages of disease progression and repeatedly aligns his internal standards with his new conditions. The changes occurring over time in the definition of QoL and the problems with their interpretation are considered challenging issues for researchers.⁹ One of the main purposes of assessing QoL changes and internal standards over time in clinical studies is to determine what amount of the reported changes are real and were caused by medical interventions, and what amount reflects errors of measurement.¹⁰ Recent progress in QoL research suggests that part of these changes could be due to changes in one's own sense of health condition as reported by self-evaluation.¹¹ Evidence suggests that patients suffering from chronic diseases change their internal standards as their disease progresses. This phenomenon reflects RS.¹²

Response shift refers to a change in the definition of QoL over time; these changes can occur as a result of any of the following:

A. Changes in an individual's internal standards of measurement (recalibration): the patient changes his/her previous calibration of QoL when faced with new conditions.

B. Changes in value prioritization, (reprioritization): i.e. if the economic domain had a higher priority compared to other QoL domains previously, the prioritization might change in the face of new conditions such as diagnosis of a disease; physical conditions might now have a higher priority in evaluating QoL compared to economic conditions.

C. Changes in the individual's concept of health status (reconceptualization):^{11, 13} For instance, an individual's concept of QoL and health status differs at the time of well-being- prior to the disease. Over time, the individual's concept is influenced by the new conditions and he might even report his QoL to be better than before, despite his current suffering.

Therefore, the differences between QoL levels reflect the differences that exist between distinct groups of patients and the changes that occur in one's QoL over time (3, 14). Response shift is defined as a particular type of bias, which might influence the results when measuring QoL changes through the common approach (mean pre-test scores – mean posttest scores). Therefore, this phenomenon should be identified and removed. This bias can also be considered an adaptive response to the changes in health states and, a positive phenomenon, given that it is part of the patient's psychological adaptation to his new health state and that it helps him better adapt. Therefore, measuring RS in QoL researches allows for the direct, valid and reliable assessment of QoL changes over time.¹⁵ Response shift is a significant phenomenon that has constantly been neglected in medical and health sciences in QoL researches.¹⁶ We did an advanced search with two key words "quality of life" and "Response shift" in Pubmed database, which was narrowed down 1990-2012. The ascending trend of QoL researches, reaching from 122 articles in 1990 to 2581 in 2012, suggests this bias more seriously. Meanwhile, among longitudinal studies, the number of articles containing QoL in their titles that mention RS in their content- was only 154 between 1990 and 2012. The publication of this article seemed important for several reasons. To our knowledge, there are few review articles on this topic. The last methodological review article on RS was written in 1999, which merely had a methodological approach. Furthermore, there are no coherent articles on the major methods applied for evaluating RS in clinical studies.

We may deduce from the aforementioned article that, firstly, the simultaneous use of several methods for evaluating RS can be beneficial because the data inaccessible through one method might be accessible through another. Secondly, if this phenomenon occurred with prior warning, the assessment methods would be different and the RS would be measured based on the external changes of the internal standards rather than through direct questioning.

Given the nature of RS and that it cannot be regarded as an objective variable, there are no methods for preventing this phenomenon except for evaluating its effects on QoL changes and indicating its magnitude and direction in the interpretation and analysis of results.⁷

The following approaches have been introduced for measuring RS.¹¹

1. Then-test approach:

This method originated from education research evaluations. Several studies have been conducted in this field. Its application in health-related quality of life (HRQoL) research is relatively limited. The then-test approach is one of the most frequently applied methods of evaluating RS and changes in the internal standards.^{17, 18} This approach entails first an evaluation at the beginning of the study, as pre-test -prior to the intervention, and post-test -after the intervention. The difference between pre-test and post-test scores measured by standard methods of evaluating QoL changes will be considered a 'traditional change'. This approach is a retrospective evaluation, looks back upon the first evaluation,^{9, 13, 19} the patient will be asked to reflect on his past and declare, based on his current standards, what score he would give to his QoL at that time.

The difference between post-test and then-test scores reflect QoL changes after the RS occurs, i.e. the QoL changes with respect to the RS effect, showing the real changes once the influence of the RS is removed.³ The advantage of this method is that the then-test is taken immediately after the post-test. Therefore, the levels of the patients' internal standards are similar on both tests. As a result, the comparison of the then-test and post-test will not be confounded by the RS recalibration, and the difference between these two can be considered as the actual change.

In the then-test evaluation, it is assumed that "response shift" creates a bias in measuring the effect of treatment on the patients' normal conditions. Meanwhile, if the patients have adapted themselves to the toxic effects of treatment or the effects of disease progression over time, they might demonstrate the effects of treatment or medical interventions or palliative care to be of a lesser or greater degree than they actually are.^{16, 20} A major criticism of the then-test approach is the probability of recall bias, which depends on the validity of the patient's accurate recalling of QoL levels in the past. Nevertheless, evidence is against the presence of a great deal of recall bias in this approach.⁹ Results of most studies indicate the occurrence of recalibration shortly after the first hospital-

ization or treatment. These changes must be taken into consideration while measuring QoL.²¹ Indeed, we can use the then-test approach for evaluating RS once the internal standards have been recalibrated. The then-test method is in fact based on measuring recalibration and it also takes a subtle hint at reconceptualization. Measuring changes is one of the most important factors required for evaluating RS; Therefore, if reconceptualization is the phenomenon that has occurred, the then-test approach is no more an appropriate method for measuring the RS, since it entails that RS is measured through the evaluation of both reconceptualization and recalibration. At this point, given the occurrence of reconceptualization and stabilization of the internal standards, RS cannot be evaluated.⁹ According to the studies and with respect to the nature of RS and the potential of this method in evaluating subjective changes, significant relationships have been observed between the then-test approach and changes in subjective variables.^{13, 22}

2. Individualized Methods:

These methods require that to determine domains that most closely associated with one's quality of life domains be specified. The functional level for each QoL domain will be determined and then the relative importance of each domain will be evaluated based on its functional level. The patient will be assessed in stages T1 and T2 and data will be collected through a semi-structured interview, allowing the patient to freely express his opinions, making data analysis a complicated and time-consuming task. The RS evaluated in the posttest is based on a semi-structured interview. The following question is asked in the posttest:

If you compare the domains you prioritized a few months ago on the basis of their significance in your QoL with those of today, can you say why they were different? Why have some of the domains remain unchanged?

The theory of individualized methods for evaluating RS is based on measuring reconceptualization and changes in the weight given to values pertaining to each domain.^{23,24} This approach is in fact a method of evaluating changes in the internal standards. According to their respective domains, changes are divided into two categories: changes in values, which occur simultaneously with changes in priorities given to each domain, and, changes

in perception, explained by changes in the content of each domain. Since the practicality of this method has not been assessed in RS studies yet, there are no empirical data available on it.¹¹

3. Preference-based Methods:

These methods evaluate the significance given to one's health state and QoL in different life scenarios, in other words, what degree of importance each individual gives to living a life of quality.²⁵ In this method, changes in values are defined by reprioritization of health states. For instance, the patient is asked to rank the value of particular health states using a visual analogue scale –a scale in which 1 represents perfect health and 0 represents death. In the time trade-off method, which is one of the preference-based methods of evaluating RS, the patient is asked to choose between a longer life in the current health state, and a shorter life spent in full health. This method refers to tangible changes in values. The disadvantages of this method are: a) it is time consuming due to the lengthy process of collecting reprioritization data, b) complicated statistical techniques are used to assess RS.¹¹

4. Successive Comparison Approaches:

These methods include judging the rank of the domains on a psychological or physical scale. These rankings are conducted by pairing the variables and asking each person to judge which option from each pair is of higher value to him. The most important feature of this approach is its 'pair-wise comparison'. In this approach the patient is asked to compare and rank QoL domains consecutively and pair-wise at least twice during the study, based on the degree of importance he gives to them. One of the methods used in this approach is the 'card sorting' method, which allows the respondent to examine the relationships among different conditions he might face. The cards are sorted upon the patient's confrontation with new conditions during the interview.¹¹

5. Structural Equation Modeling:

The structural equation modeling (SEM) is one of the statistical analytic approaches used to measure changes over time, such as in longitudinal studies.²⁶ In this approach, RS is evaluated through four stages of analysis conducted on data provided by the pretest and posttest:

Table1: Summary of Advantages of Response Shift Evaluation Methods.					
Method	Feasibility	Simplicity of Analysis	Response Shift Aspects Studied	Interpretation at Individual Level	No need for a large sample size
Then-Test Approach	✓	✓	Changes in Internal Standards Reconceptualization		✓
Structural Equation Modeling		✓ If conducted by an expert	Changes in Internal Standards Reconceptualization		
Anchor-Recalibration Approach				✓	
Individualized Methods	Is time-consuming because of the data-collecting interviews	✓	Changes in Internal Standards Reconceptualization	✓	✓
Preference-based Methods	Time-consuming	Complicated Analysis	Changes in Internal Standards		
Successive Comparison Approaches		Complicated Analysis	Changes in Internal Standards Reconceptualization		
Card Sorting	Is time-consuming because it can be confusing for the respondent		Changes in Internal Standards Reconceptualization\	✓	

During the first stage, the linear relationship between the variables under study and the RS (hidden variable) is studied, based on the confirmatory factor analysis technique. At this stage, the mean, variance, and covariance of measured scores in the pretest and posttest are fitted. In the second stage, the model 'without' RS is fitted and compared against the model 'with' RS. If the overall result of the test is significant, the recalibration, reprioritization and reconceptualization are tested in the third stage. Eventually in the fourth stage, the RS level is estimated and the real changes are evaluated.⁹

An advantage of this approach is that it allows for the distinct domains to be evaluated and compared with each other through appropriate statistical tests.²⁷ Moreover, this method requires a large sample size that is more precisely defined.²⁸ In this approach, RS is identified at group level using the mean and covariance.²⁹ It could however remain hidden in group data, owing to the various directions of change in different individuals. In other words, due to the estimated overall mean, the overall RS may not be observed in group data even if RS actually exists. Therefore, compared to individual approaches, this approach

is more sensitive.²⁴ Given that this approach does not use retrospective data, a recall bias is much less likely to occur in this method.³⁰ Hence, it is considered a promising approach for RS evaluation in clinical studies.

6. Anchor-Recalibration Approach:

First the patient rates his QoL on a scale of 1 to 10, where 1 and 10 designate the worst and best perceptions of one's QoL, respectively. The patient is then asked to describe his perceptions of the best and worst imaginable QoL and his descriptions are recorded. Since the defining process might affect recalibration, the patient is asked whether or not he would like to change his initial rating. The whole process is repeated during the second evaluation. The interviewer then asks the patient whether these descriptions have the same meaning as the recent descriptions given in the posttest, or if they are better or worse. For example, during the pretest, the patient is first asked to rate his physical functionality on a scale of 1 to 10 and then to describe the best and worst perception he has of his physical functionality. For instance, during the pretest, the patient describes 'being hospitalized' as

his worst imaginable perception of physical functionality and 'being able to run' as his best. When the description process ends, the patient is asked to declare whether the score he gave to his physical functionality has changed or not. In the posttest evaluation, the process applied in the pretest is repeated once again. However, at this point, the patient's descriptions of his best and worst perceptions of physical functionality may have changed. For example, his best perception of physical functionality may have changed from 'being able to run' to 'being able to do gardening', and his worst perception may have changed from 'being hospitalized' to 'receiving artificial respiration'. Using this method, RS is estimated by the difference between the original pretest and the transformed pretest scores, and the real changes are calculated by the difference between the posttest and the transformed pretest scores. The transformed scores are a function of the pretest scores and the worst and best imaginable cases perceived.^{9, 18}

The appropriate approach for evaluating response shift:

The following criteria may be applied to select the most appropriate method of RS evaluation:

1. Feasibility (time and cost)
2. Reliability (pretest and posttest should have reliability and be consistent with each other)
3. Validity (structure and content)
4. Empirical evidence of existing response shift.¹¹

Table 1 summarizes the criteria used in choosing the appropriate approach.

According to literature, and compared to the then-test method and the individualized methods, the SEM method is less sensitive in identifying the RS effect on QoL.²⁴ On the other hand, the then-test approach is used in the majority of clinical studies because it is better applied in chronic diseases.^{16, 20, 21} In the majority of longitudinal studies, the then-test approach is used for evaluating the RS magnitude and direction. Furthermore, the RS magnitude and direction is highly consistent between the data provided by the then-test and those provided by the SEM method, more so than the individualized approaches.⁹ Even though the then-test method is apparently the principal method applied in RS-related research, it has a number of limitations.³¹ For example, it is potentially

time consuming. And although it has been validated by education-related studies, it can still be confounded by recall bias.

A significant factor leading to the occurrence of RS is 'time', a factor that must be taken into account in longitudinal studies. Therefore, patients experiencing special conditions such as disease, pain or fatigue for a relatively long period might have had their internal standards adjusted to their conditions by the time they enter the study, and thus the patient may never report any particular changes throughout the study. Under such circumstances, the occurrence of RS is the least likely. It is therefore assumed that RS only occurs in patients who have recently been afflicted with disease.¹⁶

Conclusions

This article discusses the most important methods applied for evaluating response shift in QoL changes in longitudinal clinical studies. In addition, these methods were compared with one another and their advantages and disadvantages were briefly discussed. Response shift has recently become a growing topic in QoL health studies, yet there are several aspects of this topic that require further studies. RS may potentially conceal the effect of treatment or exaggerate its effects, thus causing an error in the estimation of QoL changes. Since it is an inevitable phenomenon, and there is a constant probability of its presence in these changes, it should be evaluated as an integral part of any study evaluating QoL changes within different sub-groups of patients to minimize the possibility of error. To this end, we propose using the traditional approach (pretest–posttest (to identify this type of bias and to allow QoL changes to be interpreted with respect to the RS effect. The common method for evaluating this bias is the 'then-test approach'. Though not flawless, the latter method's easier application and interpretation makes it more appropriate compared to the other available methods, allowing RS to be taken into consideration in the interpretation of QoL changes over time

References

1. Brow JP, McGee HM, O'Boyle CA. Conceptual approaches to the assessment of quality of life. *Psychology and Health*. 1997;12(6):737-51.

2. Pratheepawanit N, Salek M, Finlay I. The applicability of quality-of-life assessment in palliative care: comparing two quality-of-life measures. *Palliative medicine*. 1999;13(4):325-34.
3. Rees J, Clarke MG, Waldron D, O'Boyle C, Ewings P, MacDonagh RP. The measurement of response shift in patients with advanced prostate cancer and their partners. *Health and Quality of Life Outcomes*. 2005;3(21):1-8.
4. Groupt W. Study protocol for the World Health Organization project to develop a Quality of Life assessment instrument (WHOQOL). *Quality of life Research*. 1993;2(2):153-9.
5. ABDOLLAHPOUR I, SALIMI Y, NEDJAT SAHARNAZ JSZ. Quality of life and effective factors on it among governmental staff in Boukan city. *THE JOURNAL OF URMIA UNIVERSITY OF MEDICAL SCIENCES*. 2011.
6. Nedjat S, Naieni KH, Mohammad K, Majdzadeh R, Montazeri A. Quality of life among an Iranian general population sample using the World Health Organization's quality of life instrument (WHOQOL-BREF). *International journal of public health*. 2011;56(1):55-61.
7. Oort FJ, Visser MR, Sprangers MA. Formal definitions of measurement bias and explanation bias clarify measurement and conceptual perspectives on response shift. *Journal of clinical epidemiology*. 2009;62(11):1126-37.
8. Van der Maas PJ, Van Der Wal G, Haverkate I, De Graaff CLM, Kester JGC, Onwuteaka-Philipsen BD, et al. Euthanasia, physician-assisted suicide, and other medical practices involving the end of life in the Netherlands, 1990–1995. *New England Journal of Medicine*. 1996;335(22):1699-705.
9. Visser MRM, Oort FJ, Sprangers MAG. Methods to detect response shift in quality of life data: a convergent validity study. *Quality of Life Research*. 2005;14(3):629-39.
10. Ubel PA, Peeters Y, Smith D. Abandoning the language of "response shift": a plea for conceptual clarity in distinguishing scale recalibration from true changes in quality of life. *Quality of Life Research*. 2010;19(4):465-71.
11. Schwartz CE, Sprangers MAG. Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Social Science & Medicine*. 1999;48(11):1531-48.
12. Bernhard J, Lowy A, Maibach R, Hürmly C. Response shift in the perception of health for utility evaluation: an explorative investigation. *European Journal of Cancer*. 2001;37(14):1729-35.
13. Howard GS, Ralph KM, Gulanick NA, Maxwell SE, Nance DW, Gerber SK. Internal invalidity in pretest-posttest self-report evaluations and a re-evaluation of retrospective pretests. *Applied Psychological Measurement*. 1979;3(1):1-23.
14. Nedjat S, Montazeri A, Holakouie K, Mohammad K, Majdzadeh R. Psychometric properties of the Iranian interview-administered version of the World Health Organization's Quality of Life Questionnaire (WHOQOL-BREF): a population-based study. *BMC Health Services Research*. 2008;8(1):61.
15. Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. *Social science & medicine*. 1999;48(11):1507-15.
16. Sprangers MAG, Van Dam FSAM, Broersen J, Lodder L, Wever L, Visser MRM, et al. Revealing response shift in longitudinal research on fatigue: the use of the thentest approach. *Acta Oncologica*. 1999;38(6):709-18.
17. Sprangers MAG. Response-shift bias: a challenge to the assessment of patients' quality of life in cancer clinical trials. *Cancer treatment reviews*. 1996;22:55-62.
18. Visser MR, Oort FJ, Sprangers MA. Methods to detect response shift in quality of life data: a convergent validity study. *Quality of Life Research*. 2005;14(3):629-39.
19. Wiklund I. Assessment of patient-reported outcomes in clinical trials: the example of health-related quality of life. *Fundamental & clinical pharmacology*. 2004;18(3):351-63.
20. Howard GS, Dailey PR, Gulanick NA. The feasibility of informed pretests in attenuating response-shift bias. *Applied Psychological Measurement*. 1979;3(4):481-94.
21. Dabakuyo T, Guillemin F, Conroy T, Velten M, Jolly D, Mercier M, et al. Response shift effects on measuring post-operative quality of life among breast cancer patients: a multicenter cohort study. *Quality of Life Research*. 2012:1-11.
22. Bray JH, Howard GS. Methodological considerations in the evaluation of a teacher-training program. *Journal of Educational Psychology*. 1980;72(1):62.
23. Jambon B, Johnson KI. Individual quality of life and clinical trials. *Quality of Life Newsletter*. 1997:1-2.
24. Ahmed S, Mayo NE, Wood-Dauphinee S, Hanley JA, Cohen SR. The structural equation modeling technique did not show a response shift, contrary to the results of the χ^2 test and the individualized approaches. *Journal of clinical epidemiology*. 2005;58(11):1125-33.
25. Ditto PH, Druley JA, Moore KA, Danks JH, Smucker WD. Fates worse than death: the role of valued life activities in health-state evaluations. *Health Psychology*. 1996;15(5):332.
26. Millsap RE, Hartog SB. Alpha, beta, and gamma change in evaluation research: A structural equation approach. *Journal of Applied Psychology*. 1988;73(3):574.
27. Kline RB. Principles and practice of structural equation modeling: Guilford press; 2011.

28. Donaldson GW. Structural equation models for quality of life response shifts: promises and pitfalls. *Quality of life research*. 2005;14(10):2345-51.
29. Oort FJ. Using structural equation modeling to detect response shifts and true change. *Quality of Life Research*. 2005;14(3):587-98.
30. Oort FJ, Visser MR, Sprangers M. Incorporating the Thetest into

- the Structural Equation Modeling (SEM) Approach to Response-Shift Detection. *Quality of Life Research*. 2003:784-.
31. Schwartz CE, Bode R, Repucci N, Becker J, Sprangers MA, Fayers PM. The clinical significance of adaptation to changing health: a meta-analysis of response shift. *Quality of Life Research*. 2006;15(9):1533-50.