Sensitivity to Change of Generic Quality of Life Instruments in Patients with Rheumatoid Arthritis: Preliminary Findings in the Generic Health OMERACT Study

GEORGE WELLS, MAARTEN BOERS, BEVERLEY SHEA, PETER TUGWELL, RENE WESTHOVENS, MARIA SAUREZ-ALMAZOR, and RACHELLE BUCHBINDER for the OMERACT/ILAR Task Force on Generic Quality of Life

ABSTRACT. This is the initial report of the generic health OMERACT study concerned with the sensitivity to change of generic quality of life (QOL) measures. Our objective was to determine which QOL instrument is best able to show a statistically significant improvement in patients with rheumatoid arthritis (RA) demonstrating relevant improvement in a core set of disease activity and disease-specific disability measures. A multicenter controlled trial of a single group with repeated measurements at 0 (baseline), 3, and 6 months was conducted. All participating centers recruited 10 patients with RA who were about to start methotrexate therapy for the first time because of active disease. Assessments included disease activity measures, disease-specific disability measures, and generic QOL measures. To date, 40 patients have been recruited from 4 centers for the study. After 6 months of treatment many of the generic QOL measures showed a 20% improvement from baseline and medium standardized response means around 0.5. In particular, the Nottingham Health Profile (NHP) and the Rheumatoid Arthritis Quality of Life (RAQOL) measures had the largest percentage improvement (22 and 29%, respectively) and standardized response means (both with 0.54). Early results on the sensitivity of generic health QOL measures are promising, in particular for the NHP and RAQOL measures. (J Rheumatol 1999;26:217–21)

Key Indexing Terms:
RHEUMATOID ARTHRITIS
QUALITY OF LIFE
OUTCOME MEASURES
SENSITIVITY

Quality of life (QOL) refers to a variety of aspects including performance status, symptom reduction, and improved mood and sense of well being. It is an important component of health and a necessary measure of medical care. Generic QOL measures are increasingly being suggested as instruments to compare the effectiveness of interventions for different diseases and so provide data that can be used directly to allocate health care resources. At the OMERACT II Conference in Ottawa, health professionals were asked to consider whether such measures were appropriate in the study of musculoskeletal disease. Concern was expressed over the validity of these measures in their current form and the suggestion was made to extensively evaluate them for musculoskeletal disorders before recommending their use. As a result, an International League of Associations for Rheumatology (ILAR) Task Force was created to assess this evidence, and make recommendations for the necessary studies[1,2]. The study reported here focused on the issue of the sensitivity to change of these QOL measures.

Ideally, to assess sensitivity to relevant change, an external definition of what constitutes relevant is required. Randomized controlled trials have shown that the majority of patients with active rheumatoid arthritis (RA) starting methotrexate (MTX) for the first time show improvement in pain, joint count, global disease activity scores, and disease-specific disability. This study took the expected improvements in these important aspects of quality of life as the criterion against which to gauge improvement performance of generic QOL instruments. To be sensitive to change, a generic measure would have to parallel this improvement. The study question posed was which QOL instrument is best able to show a statistically significant improvement in

From the University of Ottawa and Ottawa Hospital, Ottawa, Canada; VU University Hospital, Amsterdam, The Netherlands; University of Leuven, Leuven, Belgium; University of Alberta, Edmonton, Canada; and Monash University, Melbourne, Australia.

G. Wells, PhD, Professor, Department of Medicine, and Associate Director, Clinical Epidemiology Unit, University of Ottawa and Ottawa Hospital, Civic Site, Ottawa, Canada; M. Boers, MD, Professor, Department of Clinical Epidemiology, VU University Hospital, Amsterdam, The Netherlands; B. Shea, BScN, Research Associate, Clinical Epidemiology Unit, University of Ottawa, Ottawa Hospital, Civic Site; P. Tugwell, MD, Professor and Chairman, Department of Medicine, University of Ottawa and Ottawa Hospital, General Site, Ottawa, Canada; R. Westhovens, MD, Assistant Professor, Department of Rheumatology, University of Leuven, Leuven, Belgium; M. Saurez-Almazor, MD, Associate Professor, Public Health Sciences, University of Alberta, Edmonton, Canada; R. Buchbinder, MD, Senior Lecturer, Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Australia.

Address reprint requests to Dr. G. Wells, Associate Director, Clinical Epidemiology Unit, F6, Ottawa Hospital, Civic Site, 1053 Carling Avenue, Ottawa, Canada K1Y 4E9.

patients with RA that show relevant improvement in both a more disease-specific measure of disability as represented by the Health Assessment Questionnaire (HAQ) and a core set of disease activity measures.

MATERIALS AND METHODS

A multicenter controlled trial of a single group of patients with repeated measurements at 0 (baseline), 3, and 6 months was conducted. All participating centers recruited 10 patients with RA about to start MTX therapy for the first time because of active disease. The patients' eligibility criteria included: RA according to the 1988 American College of Rheumatology (ACR) criteria[3]; no previous MTX treatment; MTX treatment for active disease now required according to the treating physician; able to complete the questionnaires according to the treating physician; and no clinically important changes in antirheumatic therapy just before or during the study that would result in improvement of disease prior to the study or flaring of disease activity during the study.

After baseline assessment, the patient started MTX at a minimum advised dose of 7.5 mg weekly, in one oral dose. The dose of MTX could be increased if the patient had shown insufficient improvement, according to the treating physician. The initial dose could be lowered, and temporary dose decreases below 7.5 mg/week were allowed. However, the patient could only be evaluated after completing a 3 month period of uninterrupted treatment with at least 7.5 mg/week. All patients were prescribed folic acid at an advised dose of 5 mg/week, but preferably at 1 mg/day. Corticosteroid injections could be given. Concurrent acceptable antirheumatic therapy included nonsteroidal antiinflammatory drugs, corticosteroids, and other disease modifying antirheumatic drugs. Preferably dosages would be kept stable. Corticosteroid dose decreases were not permitted from 2 weeks prior to study up to the final examination. If a patient was unable to complete the 3 months of treatment due to inefficacy or toxicity, the patient could be replaced.

Patients were assessed at baseline and after 3 and 6 months of treatment. The disease-specific assessments included the HAQ as a measure of disability and the ILAR/World Health Organization core set[4] for disease activity, which consisted of pain (10 cm visual analog scale, VAS), patient global assessment (10 cm VAS), physician global assessment (10 cm VAS), tender joint count (28 joint count), swollen joint count (28 joint count), and erythrocyte sedimentation rate (ESR) (mm/h). The Disease Activity Measure (DAS)[5] and the European League of Associations for Rheumatology (EULAR)[6] and ACR improvement criteria (ACR20)[7] were calculated from these assessments to derive responder and nonresponder groups. The generic QOL questionnaires used included: the Health Utilities Index (HUI)[8]; the Nottingham Health Profile (NHP)[9]; the Rheumatoid Arthritis Quality of Life (RAQOL)[10]; the European Quality of Life (EQ5D)[11]; and the Medical Outcome Survey SF-36 Health Survey (SF-36)[12]. The mental component score (SF-36 MCS) and physical component score (SF-36 PCS) based on the SF-36 were calculated[13]. Although the RAQOL is grouped with the generic instruments, it is in fact specific in the sense that it was developed to detect the QOL changes specific to patients with RA. The order of administration of these 5 "generic" QOL questionnaires varied according to a Latin square design of order 5 for each set of 5 patients recruited. The order at time 3 and 6 months was the same and both were orthogonal to the Latin square design at baseline. This ordering of the generic QOL questionnaires was fixed and strictly controlled. Each center participating in the study was to recruit 2 sets of 5 patients.

Chi-squared and Kruskal-Wallis tests were used to compare study centers on baseline characteristics and the various study measures at 3 and 6 months. Analysis of variance of the Latin square design for the generic QOL measures was conducted to determine if there were any effects due to the order in which the 5 measures were administered. The primary analysis involved the ranking of the generic QOL measures according to percentage improvement and standardized response mean (SRM, mean change/standard deviation of change). A large SRM indicated a high sensitivity to change. SRM were classified according to Cohen's effect size[14] as small (0.2), medium (0.5), or large (0.8) effects. A 20% improvement is considered by many to be the minimum percentage improvement that should be obtained. Patients were classified as responders and nonresponders using the ACR20 criteria, and the percentage of responder patients that showed a 20% improvement in a generic QOL measure (positive agreement) and the percentage of nonresponders with less than a 20% improvement (negative agreement) were calculated. The corresponding EULAR criteria, which consist of 3 categories, were not used due to the small numbers of available patients.

RESULTS

Four centers have completed their enrolment of 10 patients and a further 6 centers are actively recruiting. This report is based on the 40 patients from the completed centers, with one center from The Netherlands and one from Belgium (both Dutch language), as well as one center each from Canada and Australia (both English language). No differences were found across the study centers in baseline characteristics except for the tender joint count (center averages 8.1, 12.9, 13.8, 18.1; $p = 0.067$) and swollen joint count (center averages 8.7, 11.3, 12.4, 17.8; $p = 0.042$). There were no significant differences across the study centers in the percentage improvement and effect size of the disease activity, disease-specific disability, and generic QOL measures at the 3 and 6 month assessments. The analysis of the Latin square design for the generic QOL measures indicated no order effect.

The baseline demographic, clinical, and QOL measures for the study patients are given in Table 1. Fifty-nine percent of the patients were women, the average age was 58 years, with average disease duration of 3.3 years.

Table 1. Baseline characteristics of study patients (n = 40).

Sex (% female)	59
Age (yrs)	57.8 (14.9)*
Years since diagnosis	3.3 (3.1)
Physician global	6.0 (2.1)
Patient global	5.5 (2.0)
Pain	5.4 (2.2)
Tender joint count	13.1 (8.3)
Swollen joint count	12.4 (7.2)
ESR	38.5 (24.7)
HAQ	1.21 (0.78)
DAS	4.17 (1.09)
HUI	0.71 (0.19)
NHP	0.67 (0.20)
RAQOL	0.60 (0.23)
EQ5D	0.53 (0.27)
SF-36 MCS	47.7 (10.5)
SF-36 PCS	33.4 (9.8)

*Mean (standard deviation).
ESR: erythrocyte sedimentation rate; HAQ: Health Assessment Questionnaire; DAS: Disease Activity Index; HUI: Health Utilities Index. NHP: Nottingham Health Profile; RAQOL: Rheumatoid Arthritis Quality of Life; EQ5D: European Quality of Life; SF-36 MCS/PCS: Mental Component Score/Physical Component Score of the Medical Outcome Survey.

In Figure 1, the mean and 95% confidence interval (CI) of the percentage improvement in the disease activity and QOL measures after 3 months of MTX treatment are shown. All disease activity measures improved, with the average improvement near 20% for all but ESR. Pain, tender joint count, and the DAS and HAQ yielded statistically significant improvement as indicated by the exclusion of zero improvement from the 95% CI, but no measures were statistically larger than 20%. Also, the generic QOL measures showed improvement for all but the SF-36 MCS. After 6 months of treatment, the effect of MTX is clearly reflected in the disease activity measures (Figure 2). All measures exceeded 20% improvement on average and the tender and swollen joint count and the DAS and HAQ were significantly larger than the 20% improvement mark. Among the generic measures, the RAQOL and NHP showed the greatest improvement, with both exceeding 20% on average and being significantly larger than zero improvement. The variability of the EQ5D was found to be large, leading to a wide 95% CI that included zero improvement. Both the average MCS and PCS of the SF-36, as well as the HUI, failed to achieve the 20% improvement mark, but all were significantly greater than zero improvement.

The standardized response means (SRM) of the study measures at 6 months are shown in Figure 3. The SRM corresponding to small (0.2), medium (0.5), and large (0.8) effects are marked on the figure. All the disease activity measures exceeded the medium effect, with the tender joint count (0.98) and the DAS (1.29) far exceeding the large effect mark. Of the generic QOL measures only the RAQOL and NHP exceeded the medium effect. The HUI performed the worst, with a SRM of only 0.32. SRM were based on a sample of size 36, except for ESR and DAS, which were based on 28. Including only the 28 patients for which the DAS could be calculated, the size and ranking of the SRM did not change substantively.

The agreement of the 20% improvement in the QOL measures with responders and nonresponders according to the ACR20 criteria is provided in Table 2. The NHP had the largest positive agreement (71%) in reflecting this 20% improvement for the ACR20 responders. The negative agreement of the generic QOL measures for the ACR20 nonresponders was better than the corresponding positive agreement, with all measures exceeding 77% in reflecting no 20% improvement for the ACR20 nonresponders. In particular there was a significant relationship of the NHP ($p = 0.0034$) and RAQOL ($p = 0.0461$) agreement classification with the ACR20 responder criteria.

DISCUSSION

Economic evaluations in assessing the efficiency of health care delivery are an important addition for the proper analysis of health care. In this assessment, not only is the number of life years gained an important concept, but also the quality of these life years. These generic QOL measures can be used to compare the effectiveness of interventions for

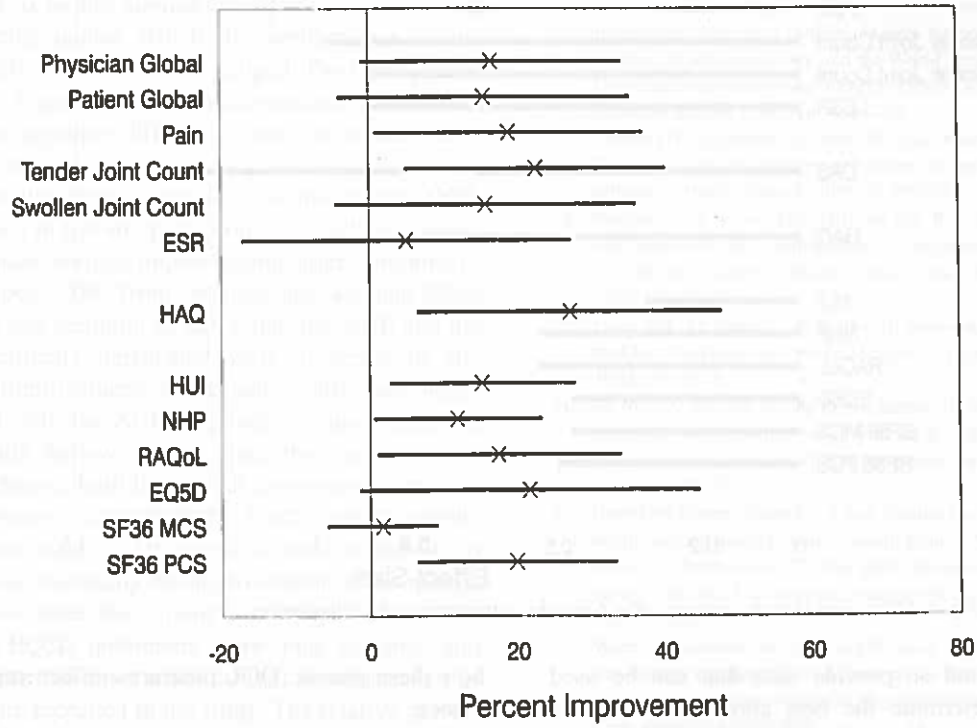

Figure 1. Percentage improvement of study measures after 3 months of treatment with methotrexate: mean and 95% CI.

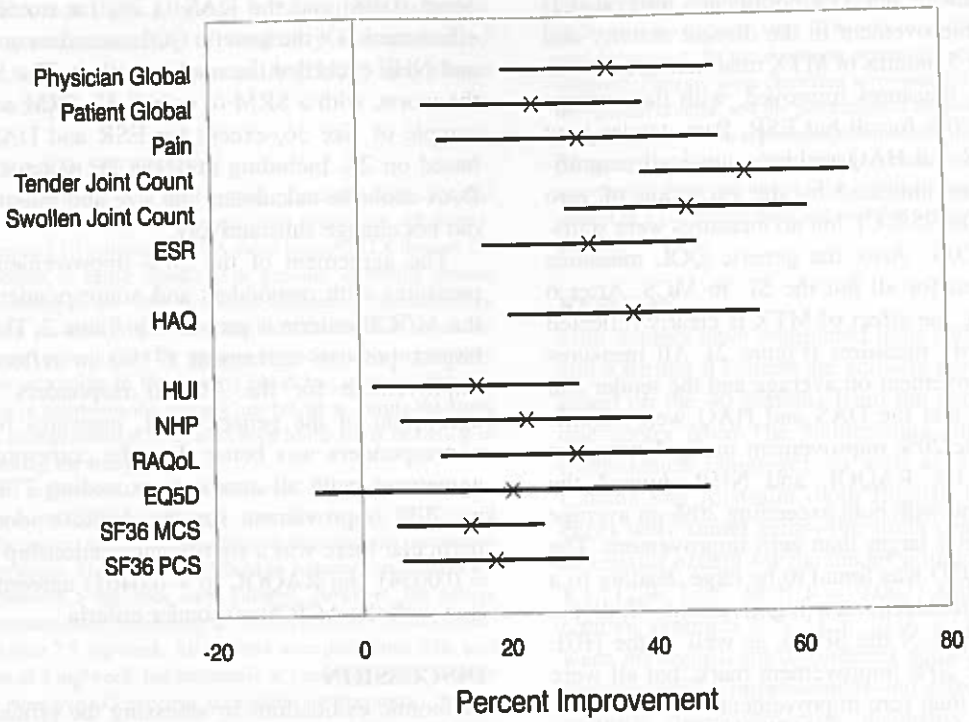

Figure 2. Percentage improvement of study measures after 6 months of treatment with methotrexate: mean and 95% CI.

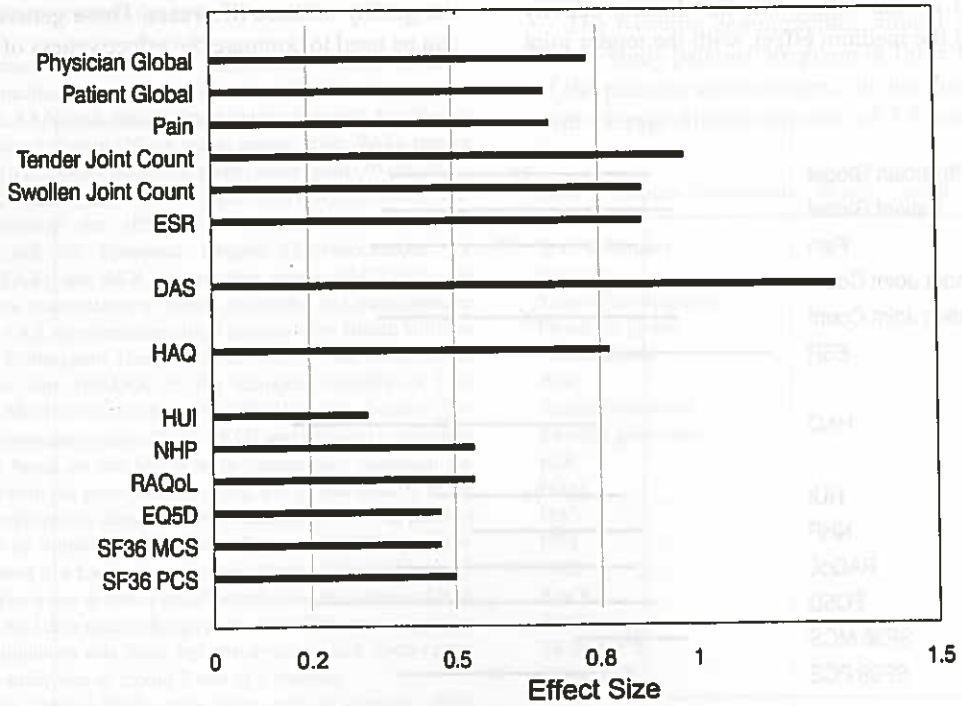

Figure 3. Effect sizes of study measures after 6 months of treatment with methotrexate.

different diseases and so provide data that can be used directly to help determine the best allocation of scarce health care resources. From the perspective of musculoskeletal diseases, it is essential to know and understand how these generic QOL measures reflect improved performance.

Several methodologic concerns needed to be considered and evaluated in the study. Although there were some center

Table 2. Positive* and negative** agreement of the generic QOL measure with the ACR20 responder criteria.

Measure	Responder (n = 14)	Nonresponder (n = 13)	p***
HUI	38.5	76.9	0.6728
NHP	71.4	84.6	0.0034
RAQOL	57.1	84.6	0.0461
EQ5D	50.0	84.6	0.1032
SF-36 MCS	42.9	66.7	0.7015
SF-36 PCS	57.1	83.3	0.0511

*Percentage of responder patients with 20% improvement in the QOL measure (positive agreement).
**Percentage of nonresponder patients with less than 20% improvement in the generic QOL measure (negative agreement).
***Fisher's exact test.
For abbreviations see Table 1.

differences in disease activity measures, namely tender and swollen joint counts, subgroup analysis indicated that the results of the study were not confounded by these differences. A concern during the study was the possibility of a problem with missing observations. However, there were few missing observations, with the exception of the measurement of ESR at 3 and 6 months. As a result of these missing ESR values, the DAS and the EULAR responder criteria could be calculated only for 70% of the patients at 6 months. There was a substantial improvement in all the disease activity and disease-specific disability measures after 6 months of treatment with MTX compared to 3 months of treatment. The choice of the 6 month treatment period appears to be the appropriate option, with relevant change occurring against which the performance of the generic QOL instruments could be gauged. The Latin square design for the 5 generic QOL questionnaires was implemented with no apparent difficulties and was successful in balancing any order effects.

The early results on the sensitivity of the generic QOL measures applied in RA are quite promising. All the generic measures revealed average improvements after 6 months of treatment of about 20% from baseline and average SRM between small and medium. In particular, the NHP and the RAQOL instruments performed well in terms of the percentage of improvement, SRM, and positive and negative agreement with the ACR responder groups. Although the SF-36 Health Survey did not reach this same level of sensitivity to change, both the mental component score and physical component score had the least variable results yielding the narrowest CI. Of particular note for the SF-36 is that there was essentially no improvement in the mental component score after the 3 month treatment period. The results for the EQ5D instrument were quite variable, and further assessment of this variability will be undertaken as more patients are recruited to the study. The relative performance of the HUI was poor.

There was concern regarding the sensitivity of the HUI prior to the study. In particular, the selection options for 3 questions appeared to have a ceiling effect when applied to musculoskeletal disorders. A modified version of the HUI was used in the study and any increase in sensitivity due to this modification will be assessed once the scoring algorithm is determined and more patients are available for evaluation. In addition, the standard gamble and utility feeling thermometer will be assessed at that time.

Data collection in the other study centers is progressing and other non-English speaking centers are currently enrolling patients. A subsequent report will further explore and provide more precise results for the 5 generic QOL measures, as well as the modified HUI, standard gamble, and utility feeling thermometer.

REFERENCES

1. Bellamy N, Boers M, Felson DT, et al. Workshop report: health status instruments/utilities. J Rheumatol 1995;22:1203–7.
2. Tugwell P, Boers M, Brooks PM. OMERACT II conference-outcome measures in RA. Clinical trials: conclusion. J Rheumatol 1995;22:1431–2.
3. Arnett FC, Edworthy SM, Bloch DA, et al. The American Rheumatism Association revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315–24.
4. The OMERACT Committee. Conference on outcome measures in rheumatoid arthritis clinical trials, Maastricht, the Netherlands, April 29-May 3, 1992. J Rheumatol 1993;20:527–91.
5. van der Heijde DM, van't Hof M, van Riel PL, van de Putte LB. Development of a disease activity score based on judgment in clinical practice by rheumatologists. J Rheumatol 1993;20:579–81.
6. Van Gestel AM, Prevoo ML, van't Hof MA, van Rijswijk MH, van de Putte LB, van Riel PL. 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 Rheum 1996 Jan;39:34–40.
7. Felson DT, Anderson JJ, Boers M, et al. American College of Rheumatology preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 1995;38:727–35.
8. Torrance GW, Feeny DH, Furlong WJ, Barr RD, Zhang Y, Wang Q. Multiattribute utility function for a comprehensive health status classification system. Health Utilities Index Mark 2. Medical Care 1996;34:702–22.
9. Hunt SM, McEwen J, McKenna SP. Measuring health status: a new tool for clinicians and epidemiologists. J Roy Coll Gen Pract 1985;35:185–8.
10. de Jong Z, van der Heijde D, McKenna SP, Whalley D. The reliability and construct validity of the RAQOL: a rheumatoid arthritis-specific quality of life instrument. Br J Rheumatol 1997;36:878–83.
11. Euro Qol Group. EuroQol: a new facility for the measurement of health-related quality of life. Health Policy 1990;16:199–208.
12. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473–83.
13. Ware JE, Kosinski M, Bayliss MS, et al. Comparison of methods for the scoring and statistical analysis of the SF-36 health profile and summary measures: results for the medical outcomes study. Medical Care 1995;33:AS264–279.
14. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale: Laurence Erlbaum; 1988.