Psychometric Evaluation of the SF-12 Health Survey in Rheumatoid and Osteo- arthritis Clinical Trials

Sanjay K. Gandhi,[1] Ph.D.

J. Warren Salmon,[2] Ph.D.

Sean Z. Zhao,[1] Ph.D.

Bruce L. Lambert,[2] Ph.D.

Prasanna R. Gore,[2] Ph.D.

Kendon Conrad,[2] Ph.D.

[1] Pharmacia Corp., [2] University of Illinois at Chicago.

Corresponding author: Sanjay K. Gandhi, Manager, Global Health Outcomes, Pharmacia (Searle), 5200 Old Orchard Rd, Skokie, IL 60077.

Abstract

Objective: To evaluate psychometric properties of the SF-12 health survey as a generic health-related quality of life (HQL) measure in osteoarthritis (OA) and rheumatoid arthritis (RA) patient populations in clinical trials.

Method: Data were aggregated from three clinical trials, evaluating efficacy of different NSAIDs in OA (N=651) and RA (N=693) patients. Patient assessments were made using SF-36 and commonly used clinical measures in OA and RA at baseline, and up to week 6. The SF-12 items were extracted from the SF-36 items. For the SF-12, item missing rate, computability of component scores, factor structure, item-component correlations, and floor and ceiling effects were evaluated. Correlations of SF-12 physical (PCS12) and mental component summary scores (MCS12) with SF-36 component summary scores (PCS36 and MCS36), and clinical variables were also examined. Analyses for OA and RA patients were conducted separately.

Results: A low individual SF-12 item missing rate (0.14% to 2.3%) and a high percentage score computability (91-94%) were observed at baseline. No floor or ceiling effects at baseline were observed. The scree plot confirmed two factor structure of the SF-12 items. Items belonging to the physical component correlated more strongly with the PCS12 than the MCS12, and vice versa. The correlations between PCS12 and PCS36, and MCS12 and MCS36 ranged from 0.92-0.96 ($p<0.0005$), at baseline, and week 4 or 6. Significant correlations of -0.09 to -0.58 ($p<0.05$) between SF-12 scores and clinical variables, at baseline, week 4 or 6 were observed. A similar trend was observed between SF-12 and clinical variable change scores at week 2, and week 4 or 6.

Conclusion: The SF-12 is a psychometrically sound tool for the assessment of HQL in osteo- and rheumatoid arthritis patients.

Introduction

Patient quality of life is considered a key health outcome indicator.[1] This is especially true in chronic or terminal diseases. By definition, Chronic diseases are almost always impossible to cure and the emphasis is on ameliorating symptoms and thus improving quality of life of patients rather than ridding them of the disease.[2,3] Hence health-related quality of life (HQL) is an important outcome variable of different health care interventions. An increase in the use of HQL variables in clinical trials of drug therapy has been observed over the last 10 years.[4] This can partly be attributed to an increased awareness among clinical practitioners and health policy decision makers regarding the utility of the HQL as a measure of drug therapy effectiveness.[2,5-7] Though the majority of clinical trials include HQL as a secondary variable,[4,8] an increase in the number of clinical trials incorporating HQL variables indicates greater popularity of HQL as outcome variable of drug therapy trials.[9] The most important single trait that justifies the use of HQL measures in the assessment of health care interventions is that it is the most relevant patient-centered outcome assessment tool.

In 1997, the number of people suffering from arthritis, in the US, was estimated to be 40 million, and this number is expected to increase to 60 million by 2020.[10] Because of its high prevalence, arthritis represents a significant burden to the population, the US health care system, and society. Arthritis (osteo and rheumatoid) is shown to adversely affect the functional status/quality of life of patients.[11,12] It is noteworthy that the quality of life has been included as one of the recommended measures in the outcome assessment of osteoarthritis drug clinical trials.[13] The inclusion of quality of life in the core set of measures for osteoarthritis clinical trials highlights the increasing popularity of quality of life as a major end-point in such studies. It has also been suggested that validated quality of life scales would complement the anthropomorphic, clinical and laboratory data in RA clinical trials.[14] Arthritis interventions can be assessed over time by examining their impact on patients' HQL, as measured by either a disease-specific or a generic HQL instrument.[15]

Instrument psychometric properties are important in selecting HQL instruments for use in clinical trials. The instrument should possess satisfactory reliability and validity properties for credible study results.[16] In general, without satisfactory instrument psychometric properties, the study results would be

of questionable value.[17] Before incurring the expenses and patient burden of an instrument in clinical trials, it is desirable to understand its psychometric properties in a population similar to the study population. Another consideration in selection of HQL instruments is respondent burden or length of the questionnaire.[16] Periodic measure of HQL by means of lengthy questionnaires at relatively frequent time-points is often not feasible. Hence, HQL instruments with less respondent burden are desirable.[16]

The Medical Outcomes Study Short Form-36 is a widely used and well validated instrument to assess generic HQL.[18-22] The Medical Outcomes Study Short Form-12[23] (SF-12), a generic HQL instrument, was constructed using a subset of items from SF-36. The respondent burden is obviously reduced. Thus, especially when the respondent burden is a concern, the SF-12 could provide an efficient alternative as compared to SF-36.[23] Psychometric properties of the SF-12 have been tested in the general population and certain disease conditions[23], but no literature is available on its exclusive use in osteo- and rheumatoid arthritis patient populations. The present study is aimed at evaluating the psychometric properties of the SF-12 as a measure of generic HQL instrument in clinical trials involving patients with osteo- and rheumatoid arthritis.

The comparative item structure covering domains assessed by the SF-36 is shown in Table I. Individual items of the SF-12 cover various HQL domains. These domains are: i) physical functioning (2 items), ii) role-physical (2 items), iii) bodily pain (1 item), iv) general health (1 item), v) vitality (1 item), vi) social functioning (1 item), vii) role-emotional (2 items), and viii) mental health (2 items). The physical (domains i to iv) and mental (domains v to viii) component summary scores are computed from these domains. The 12 items were selected such that these items explain at least 90% of variability in the physical component and mental component summary scores of the SF-36. The SF-36 was used as a criterion for the validation of SF-12. The validity of the criterion measure (SF-36) becomes important to ascertain the validity of the test measure (SF-12).[24]

The present study is aimed at testing the psychometric properties of the SF-12 health survey in osteoarthritis and rheumatoid arthritis patient populations under clinical trial setting. Recommendation of different measures of clinical variables for RA[13] and OA [25] underlines the differences in these diseases. Hence results in these populations are reported separately.

METHODS

Data from five clinical trials were used for the study. These data were provided by G. D. Searle & Co. (Skokie, IL). These clinical trials involved assessment of efficacy of drug treatments in arthritis and used the SF-36 as an HQL measure. Data with sole administration of SF-12 were not available. The SF-12 is comprised of 12 items which are a subset of items from the larger SF-36. Patient responses on these 12 items were used for this study ('embedded' form).

Considering inherent differences in the prognoses of osteoarthritis (OA) and rheumatoid arthritis (RA), psychometric evaluation of SF-12 was conducted separately in OA and RA patient populations. However, the methodology used was the same in OA and RA patient populations. Hence a common methods section is provided for OA and RA trials. Whenever necessary, distinctions between OA and RA patient populations are made. The present study is limited to the acute form (1 week version) of the SF-12. For the purpose of the study, data from different clinical trials were aggregated to obtain a larger pool of patients for data analysis.

Psychometric evaluation of the SF-12 in arthritis clinical trials

Data source: For the psychometric assessment of the SF-12 in OA, data from three clinical trials (OA Trial I, OA Trial II, and OA Trial III) were aggregated; whereas data from two trials (RA Trial I and RA Trial II) were aggregated for the psychometric evaluation of SF-12 in RA. The clinical variables used in OA and the RA trials are described in Table 1. All the trials were randomized, double blind, placebo controlled clinical trials. The drugs used in this trials were NSAIDs.

Analytic plan, data processing and data analysis

The criteria evaluated across three OA and two RA trials were: completeness of data in terms of item-level and component summary level missing data, score computability, features of scale score distribution (such as ceiling and floor effects), factor structure, item discriminant validity and scaling success rate, item-component correlation, amount of variability in SF-36 scores as explained by SF-12 scores, and correlation of SF-12 with clinical variables.

Described below is the framework adopted to attain each of the above mentioned objectives. For all comparisons, the significance level was set at $p<0.05$, unless specified otherwise.

Completeness of data and score computability

The percentage of patients missing individual items in the SF-12 were recorded and qualitatively compared. Based on the criteria set forth by authors[26,27] computability of the scores for the SF-12 and PCS and MCS were calculated. In case of the SF-12, the PCS-12 and MCS-12 scores cannot be computed unless responses to all 12 items are available.[27]

Features of score distribution

Features of score distribution and the percent of trial participants scoring the lowest possible scores (floor effect) and the highest possible scores (ceiling effect) were calculated to determine the ability of the items to capture the full range of health states.[24] The SF-12 summary scores were qualitatively compared with the SF-36 summary scores.

Factor structure

Factor structure of the SF-12 items were examined using the principal components analysis with varimax rotation.[28] It was hypothesized that two factors will be obtained (physical and mental component). This was confirmed using the scree-test criteria.[29] In addition, items originally belonging to physical functioning, role-physical, bodily pain, and general health scales were hypothesized to load higher on the "physical health", whereas vitality, social functioning, role emotional, and mental health were hypothesized to load higher on the "mental health" factor.[27] The developers have suggested that general health, vitality and social functioning items cross-load on physical and mental components. Cross loading of 0.40 is considered to be meaningful in the social sciences research[30] and hence was accepted as a convention for this study.

Item-component correlation

As suggested by developers, correlation between 'physical component' items and physical component summary score should be higher than the correlation between 'mental component' items and mental component summary scores. These relationships were examined for SF-12 items.

Amount of variability in PCS-36 and MCS-36 as explained by PCS-12 and MCS-12 scores

Pearson's correlation coefficients (r) were computed between PCS-12 and MCS-12 with PCS-36 and MCS-36, respectively. The individual r^2s indicated the amount of variability in SF-36 summary

scores as explained by SF-12 component summary scores. It was hypothesized that about 90% of variance in the SF-36 component summary scores would be explained by the SF-12 component summary scores.[23] The correlations were computed for cross-sectional and difference scores.

Correlation between SF-12 component summary scores and clinical indicators

The relationships between SF-12 PCS and MCS, and arthritis severity indicators were examined by computing zero-order correlation coefficients between arthritis severity indicators and SF-12 PCS and MCS scores. The correlation coefficients were also computed between changes in arthritis severity indicators and corresponding changes in the SF-12 summary scores. The clinical variables used for the OA were the physician global assessment,[31] patient global assessment,[32,31] pain intensity,[33] knee pain on weight bearing,[34] knee pain on motion,[34] and time to walk 50 feet.[34] For the RA condition the clinical variables used were physician global assessment,[31] patient global assessment,[31,32] pain intensity,[319] duration of morning stiffness,[318] assessment of joint swelling,[302,318] assessment of joint tenderness/pain,[35] and functional capacity classification.[36]

RESULTS

Patient demographic characteristics of osteo- and rheumatoid arthritis trials

The OA patients were predominantly caucasian (84%) and female (71%). The mean age of OA patients was 61.89 years (SD - 11.06). The average duration of OA condition was 9.35 years (SD - 8.51). Similarly, the RA patients were predominantly caucasian (86.9%) and female (75.8%). The mean age of RA patients was 56.29 years (SD 11.87). The average duration of RA was 11.16 years (SD 8.92).

Completeness of data

The percentage missing rate of individual SF-12 items in OA ranged from 0.46 to 2.30. Similarly, the percentage missing rate of individual SF-12 items in RA ranged from 0.29 to 1.73. Based on the baseline data, SF-12 summary scores could be computed for 90.94% in OA and 94.32% in RA.

Features of score distribution

Features of score distribution and the percent of trial participants scoring the lowest possible scores (floor effect) and the highest possible scores (ceiling effect) at baseline for OA and RA are

presented in Table 2. As can be seen in Table 2, the component summary scores of the SF-12 did not show any floor or ceiling effects.

Factor structure

The scree plot and principal components analysis were used to establish the factor structure of SF-12 items in the arthritis patient population.[28] Figures 1 and 2 show the scree plots for OA and RA, respectively. Using the scree plot criteria, the two factor structure of the SF-12 in OA and RA was confirmed. This provided confirmation of the two component structure of health as measured by the SF-12 health survey. Table 3 shows item-loadings of the SF-12 items using principal components analysis after varimax rotation for OA and RA patients, respectively. Items assessing Physical Functioning, Bodily Pain, and Role Physical load higher on factor 1 (Physical Component), whereas items assessing Mental Health and Role Emotional load highly on the second component (Mental Component). Items assessing general health, vitality and social functioning are loaded on both the components. Similar results have been observed in other patient populations.[23]

Item-component correlations

Table 4 shows the correlation of individual items and SF-12 summary scores. In both patient populations (OA and RA), physical functioning, role physical, bodily pain and general-health items correlated higher with physical component summary score, whereas, vitality, role-emtional, social functioning and mental health items correlated higher with mental component summary score. The results confirmed the hypothesized item-scale correlations.

Corelations between SF-12 and SF-36 summary scores

The correlations between physical component summary scores of SF-12 and SF-36 ranged from 0.92 - 0.95 ($p < 0.0005$) at baseline, 2 week, 4 week or 6 week scores in OA and RA patients. Similarly, correlations between mental component summary scores of SF-12 and SF-36 ranged from 0.95 - 0.96 ($p < 0.0005$) at baseline, 2 week, 4 week or 6 week scores in OA and RA patients.

Correlation between SF-12 component summary scores and clinical indicators

Significant correlations (-0.18 and -0.55) between SF-12 components and clinical variables were obtained at baseline and week 6 in OA and RA (Tables 5 and 6). Significant correlations (-0.14 to -0.46) in week 4 and 6 change scores of SF-12 component summaries and clinical variables were also observed in OA and RA patient populations. Moderate correlation coefficients between SF-12 scores with clinical variables indicated construct validity of SF-12 and SF-36 scores.

DISCUSSION AND CONCLUSION

Health-related quality of life research has become increasingly popular in the past decade.[37] More and more clinical trials have incorporated HQL end-points as treatment efficacy measures.[9] The U.S. National Institutes of Health is supporting research in the area of HQL.[38] The American College of Physicians in their position statement has asserted that patient well-being is at the core of medical practice.[39] The importance of HQL in health care resource allocation and clinical practice decisions is predicted to increase.[40] Thus, HQL research is propelled by multiple forces within the health sector, most notably for greater precision in methodology, instrumentation, and measurement by the pharmaceutical industry. While the US FDA has not issued regulations on the HQL assessment, pharmaceutical manufacturers and government agencies have supported the efforts of academics and contract research organizations in developing and testing HQL instruments.

In recent reviews of HQL evaluations in clinical trials, a serious deficit was noted in reporting of psychometric properties of instruments used to assess HQL.[4,41] In addition, few trials reported psychometric properties of generic instruments in the disease condition being studied. One possible explanation could be the lack of availability of such information. The present study aims to fill part of this gap in the literature by assessing the psychometric properties of SF-12, a generic HQL instrument, in arthritis patients. The importance of SF-12 as a generic HQL instrument stems from its brevity and ability to explain about 90% of variance in the extensively validated SF-36 component summary scores.[23] The present study reports findings of a validation study of SF-12 in OA and RA patients in clinical trial setting.

The present study used the embedded form as opposed to the unembedded form. The comparability of SF-12 psychometric properties when administered alone as compared to using subset of SF-36 items is reported.[27] In brief, it has been shown that: a) embedded and unembedded forms showed close similarity in terms of ordering of items; and b) the factor structure of the SF-12 and the factor contents were virtually the same in the embedded and the unembedded forms. Furthermore, the performance of the unembedded SF-12 in discriminating groups known to differ in physical health and mental health suggest that the psychometric properties of the unembedded form is similar to that of the embedded form. It has also been reported that no trends in the results of several studies of the unembedded form are different than the embedded form. Hence, though 'unembedded' form would have been ideal for the study purpose, the use of 'embedded' form does not pose threats to the validity of study results.

Features of score distribution of SF-12

It was observed that the percent missing rate of individual SF-12 items was very low for OA and RA patients. The score computability of SF-12 was 91.96% in OA and 94.38% in RA. This has implications for SF-12's practical utility as an outcomes assessment tool. A low score computability would adversely affect the cost of the study by increasing the number of subjects needed to detect the differences between groups. No floor or ceiling effects in SF-12 scores were observed in OA and RA patients. This is a desirable property of any psychometric instrument because it indicates the ability of the SF-12 component scores to capture a full range of health states and provides an indication of the discriminative ability of the instrument.

The two factor structure of SF-12 was confirmed in OA and RA patients using the scree plot criteria. The hypotheses regarding item-factor loadings were confirmed using principal component analyses. The hypotheses regarding the correlation of individual item-component correlations were tested and confirmed in OA and RA patient populations. These results demonstrated the validity of SF-12 in OA and RA patient populations. However, validity is an incremental process. Further evidence of SF-12's validity will strengthen its position as a health outcomes measure in arthritis patient populations.

Correlation of SF-12 with clinical variables

Significant correlations of SF-12 with clinical variables were observed. The patients reflecting poorer health on clinical variables indicated the same on SF-12 scores. This indicated construct validity of SF-12 component scores. The magnitude of most of these correlations were moderate (about -0.20 to -0.30) with extreme correlations being -0.10 to -0.55. This implies that SF-12 scores capture a different portion of health status of patients than clinical variables. In addition, significant moderate correlation coefficients of HQL scores with clinical indicators reflect the 'face validity' of patient HQL as an outcomes assessment tool.

Correlation of SF-12 with SF-36

A very high correlation of PCS12 with PCS36 and MCS12 with MCS36 was observed in OA and RA patient populations at baseline, and at week 2, 4 and 6. The amount of variability in PCS36 as explained by PCS12 ranged from 85% to 90%. Similarly, the amount of variability in MCS36 as explained by MCS12 ranged from 90% to 92%. Such high correlation of SF-12 with SF-36 scores also indicates convergent validity of SF-12 scores where SF-36 scores are considered as criterion variables. A large amount of variability in SF-36 scores explained by SF-12 scores justifies the attempts to substitute SF-36 measures with SF-12 measures. The result of this substitution can be reduced respondent burden without the loss of significant amount of information.

Reliability of SF-12

Reliability of the SF-12 is reported using the test-retest approach.[26,28] This study involved psychometric evaluation using placebo-controlled clinical trial data. It could be hypothesized that repeated measure of HQL in placebo could be considered as a test-retest measure. However, placebo-effect (psychological relief) in medical drug therapy literature is widely reported.[42-45] Hence, this study could not assess reliability coefficients for the SF-12.

Arthritis is one of the most prevalent chronic disease condition in the US. The prevalence of arthritis is expected to increase to about 60 million people in next two decades.[10] Arthritis has been reported to impair patient HQL and work productivity.[46] The deleterious

impact of arthritis on the US health care system and the lives of the patients is well-documented.[11,47] Hence, the focus of medical interventions in arthritis centers around improving the quality of life of patients. It is important to make available valid health status assessment tools for arthritis patients. The present study established psychometric properties of SF-12 generic instrument in osteo- and rheumatoid arthritis patient population.

The health care settings that may facilitate patients' HQL in the real world could be pharmacies and physician offices. There are several barriers to the acceptance and use of this outcome measure for patient health monitoring by the health care professionals.[48] Some of the obstacles include respondent burden, cost of data processing, complex data analysis and interpretation, and difficulty in responding to items. The SF-12 has proven to be a HQL instrument with less respondent burden and satisfactory content validity. Due to its brevity, SF-12 may prove to be a useful patient assessment tool for clinicians in their routine practice. The burden and cost of data processing would be expected to be less than other lengthier HQL instruments. Results indicated a very low item missing rate, probably indicating comprehensibility of questions and brevity of the instrument. These factors indicate a more likely acceptability of this instrument than others by medical practitioners.

CITED LITERATURE

1. Stewart AL, Ware JE. Measuring Functioning and Well-Being. The Medical Outcomes Study approach. London: Duke University Press: 1992.
2. Epstein RS, Sherwood LM. From outcomes research to disease management: A guide to perplexed. Ann Intern Med 1996;124:832-837.
3. Kong SX, Stubbings J. Assessing the economic value of pharmaceutical products and services. J Pharm Prac 1995;8:143-155.
4. Kong SX, Gandhi SK. Methodologic assessments of quality of life measures in clinical trials. Ann Pharmacother 1997;31:830-836.
5. Joe RH, Carter CB. Formulary decision making for evolving therapies. Am J Managed Care 1996;2(7):868-875.
6. Johannesson M. Economic evaluation of drugs and its potential uses in policy making. PharmacoEconomics 1995;8(3):190-198.
7. Greenfield S, Nelson EC. Recent developments and future issues in the use of health status assessment measures in clinical studies. Med Care 1992;30 (5 Suppl):MS23-MS41.
8. Bergner M. Quality of life, health status, and clinical research. Med Care 1989;27:S148-S156.
9. Spilker B. Quality of life and pharmacoeconomics in clinical trials. Second edition. Philadelphia, PA:Lippincott-Raven publishers 1996, pg 1-10.
10. Anon. The baby boomers go gray. Arthritis Today 1997;Jan-Feb:29.
11. Bendtsen P, Hornquist JO. Change and status in quality of life in patients with rheumatoid arthritis. Qual Life Res 1992;1:297-305.
12. Hampson SE, Glasgow RE, Zeiss AM. Personal models of osteoarthritis and their relation to self-management activities and quality of life. J Behav Med 1994;17(2):143-158.
13. Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. J Rheumatol 1997; 24(4):799-802.
14. Bell MJ, Bombardier C, Tugwell P. Measurement of functional status, quality of life, and utility in rheumatoid arthritis. Arthritis Rheum 1990;33:591-601.
15. Buchbinder R, Bombardier C, Yeung M, Tugwell P. Which outcomes measures should be used in rheumatoid arthritis clinical trials. Arthritis Rheum 1995;38:1568-1580.
16. Bowling A. Measuring disease: A review of disease specific quality of life instruments. Philadelphia: Open University Press; 1995:1-19.
17. Patrick DL, Erickson P. Health status and health policy: Quality of life in health care evaluation and resource allocation. New York; Oxford University Press: 1993.
18. Ware JE, Sherbourne CD. The MOS 36-Item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30(6):473-483.
19. Ware JE, Kosinski M, Bayliss MS, McHorney CA, Rogers WH, Raczek A. Comparison of methods for scoring and statistical analysis of SF-36 health profile and summary measure: Summary of results from the Medical Outcomes Study. Med Care 1995;33(4):AS264-AS279.

20. McHorney CA, Ware JE, Raczek AE. The MOS 36-item short-form health survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care 1993;31(3):247-263.
21. McHorney CA, Ware JE, Lu JFR, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): III. Tests of data quality, scaling assumption, and reliability across diverse patient groups. Med Care 1994;32(1):40-66.
22. McHorney CA, Ware JE. Construction of an alternative form general mental health scale for the MOS SF-36 Health Survey. Med Care 1995;33:15-28.
23. Ware JE, Kosinski M, Keller SD. A 12-item Short-Form Health Survey. Construction of scales and preliminary tests on reliability and validity. Med Care 1996;34(3):220-233.
24. Nunnally JC, Bernstein IH. Psychometric theory. 3rd edition. New York; McGraw Hill:1994.
25. Tugwell P, Boers M. Developing consensus on preliminary core efficacy endpoints for rheumatoid arthritis clinical trials. OMERACT Committee. J Rheumatol 1993 Mar;20(3):555-6.
26. Ware JE, Kosinski M, Keller SD. SF-36 Physical and Mental Component Summary Scales: A User's Manual. Boston, MA: The Health Institute, 1994.
27. Ware JE, Kosinski M, Keller SD. SF-12: How to score the SF-12 physical and mental health summary scales (second edition). The Health Institute, New England Medical Center, Boston, Massachusetts, 1995.
28. Kerlinger FN. Foundations of Behavioral Research. New York: Holt, Rinehart, and Winston, 1973.
29. Cattell RB. The scree-test for the number of factors. Multivariate Behav Res 1966;1:245-276.
30. Hair JF, Anderson RE, Tatham RL, Black WC. Multivariate data analysis, Fourth edn. New Jersey: Prentice Hall, 1995.
31. Cooperating Clinics Committee of American Rheumatism Association. A seven-day variability study of 499 patients with peripheral rheumatoid arthritis. Arthritis Rheum 1965;8:302-334.
32. Ward J, Williams HJ, Egger MJ, et al. Comparison of auranofin, gold sodium thiomolate and placebo in the treatment of rheumatoid arthritis. Arthritis Rheum 1993;26(11):1303-1315.
33. Fries JF. The assessment of disability: from first to future principles. Brit J Rheumatol 1983;22(suppl):48-58.
34. Searle. A multicenter, double-blind, placebo controlled comparison study of the safety and efficacy of Daypro® (oxaprozin) 1200 mg QD and Relafen® (nabumetone) 1500 mg QD in the treatment of osteoarthritis of the knee. Protocol number: N65-94-02-022. G.D. Searle & Co. Skokie, IL, U.S.A., 1995.
35. Ritchie DM, Boyle JA, McInnes JM. Clinical studies with an articular index of the assessment of joint tenderness in patients with rheumatoid arthritis. Q J Med 1968;147:393-496.
36. Steinbroker O, Traeger CH, Batterman RC. Therapeutic criteria in rheumatic arthritis. JAMA 1949;140:659.
37. Revicki DA. Quality of life research and the health care industry. J Res Pharm Econ 1990;2:41-53.

38. Meenan RF, Mason JH, Anderson JJ, Guccione AA, Kazis LE. AIMS2: the content and properties of revised and expanded Arthritis Impact Measurement Scales health status questionniare. Arthritis Rheum 1992;35(1):1-10.
39. American College of Physicians. Comprehensive functional assessment for elderly patients. Position paper. Philadelphia: American College of Physicians, January 15, 1988.
40. Drummond MF, Stoddard GL, Torrance GW. Methods for the economic evaluation of health care programs. Oxford: Oxford University Press; 1986.
41. Gandhi SK, Kong SX. Quality of life measures in the evaluation of antihypertensive drug therapy: Reliability, validity and quality of life domains. Clin Ther 1996;18:1276-1295.
42. Bouchet C, Guillemin F, Briancon S. Nonspecific effects in longitudinal studies: impact of quality of life measures. J Clin Epidemiol 1996;49(1):15-20.
43. Bienenfeld L, Frishman W, Glasser SP. The placebo effect in cardiovascular disease. Am Heart J 1996;132(6):1207-1221.
44. Piercy MA, Sramek JJ, Kurtz NM, Cutler NR. Placebo response in anxiety disorders. Ann Pharmacother 1996;30(9):1013-1019.
45. Hrobjartsson A. The uncontrollable placebo effect. Eur J Clin Pharmacol 1996;50(5):345-348.
46. Blake DJ, Maisiak R, Alarcon GS, Holley HL, Brown S. Sexual quality-of-life of patients with arthritis compared to arthritis-free controls. J Rheumatol 1987; 14(3):570-576.
47. Frank RG, Beck NC, Parker JC, Kashani JH, Elliott TR, Haut AE, Smith E, Atwood C, Brownlee-Duffeck M, Kay DR. Depression in rheumatoid arthritis. J Rheumatol 1988;15:920-925.
48. Deyo RA, Patrick DL. Barriers to use of health status measurement in clinical investigation, patient care, and policy research. Med Care 1989;27:S254-S268.

Table 1. Trial demographics included in the study

Trial	Duration	N	Clinical variables
OA Trial I	6 wks	328	Physician global assessment of OA condition, Patient global assessment of OA condition, Pain intensity, Functional capacity classification, Knee pain on weight bearing, Knee pain on motion, Time to walk 50 feet
OA Trial II	6 wks	347	Physician global assessment of OA condition, Patient global assessment of OA condition, Pain intensity, Functional capacity classification, Knee pain on weight bearing, Knee pain on motion, Time to walk 50 feet
OA Trial III	6 wks	572	Physician global assessment of OA condition, Patient global assessment of OA condition, Pain intensity, Functional capacity classification
RA Trial I	12 wks	380	Physician global assessment, Patient global assessment, Duration of morning stiffness, Pain intensity, Assessment of Joint Swelling, Assessment of Joint Tenderness/ Pain, Functional capacity classification, Erythrocyte sedimentation rate
RA Trial II	4 wks	300	Physician global assessment, Patient global assessment, Duration of morning stiffness, Pain intensity, Assessment of Joint Swelling, Assessment of Joint Tenderness/ Pain, Functional capacity classification, Erythrocyte sedimentation rate

TABLE 2. FEATURES OF SCORE DISTRIBUTION FOR OSTEOARTHRITIS TRIALS

Scale/Summary Score	N†	% missing	Mean	SD	Min Score	Floor Effect		Max Score	Ceiling Effect	
						n	%		n	%
Osteoarthrisis										
PCS12	592	9.06	33.45	9.31	13.21	0	0	60.98	0	0
MCS12	592	9.06	49.90	10.48	22.48	0	0	69.32	0	0
Rheumatoid arthritis										
PCS12	654	5.62	33.09	9.04	12.34	0	0	58.49	0	0
MCS12	654	5.62	47.70	11.02	15.86	0	0	69.41	0	0

†number of patients for whom the score can be computed out of total of 651 patients for OA and … for RA; -- Not applicable; Baseline data reported; PCS12 - Physical Component Summary Score of SF-12; MCS12 - Mental Component Summary Score of SF-12. The minimum score is 0 and the maximum score is 100.

FIGURE 1. SCREE PLOT OF EIGENVALUES OF INDIVIDUAL FACTORS OF SF-12 ITEMS IN OSTEOARTHRITIS TRIALS

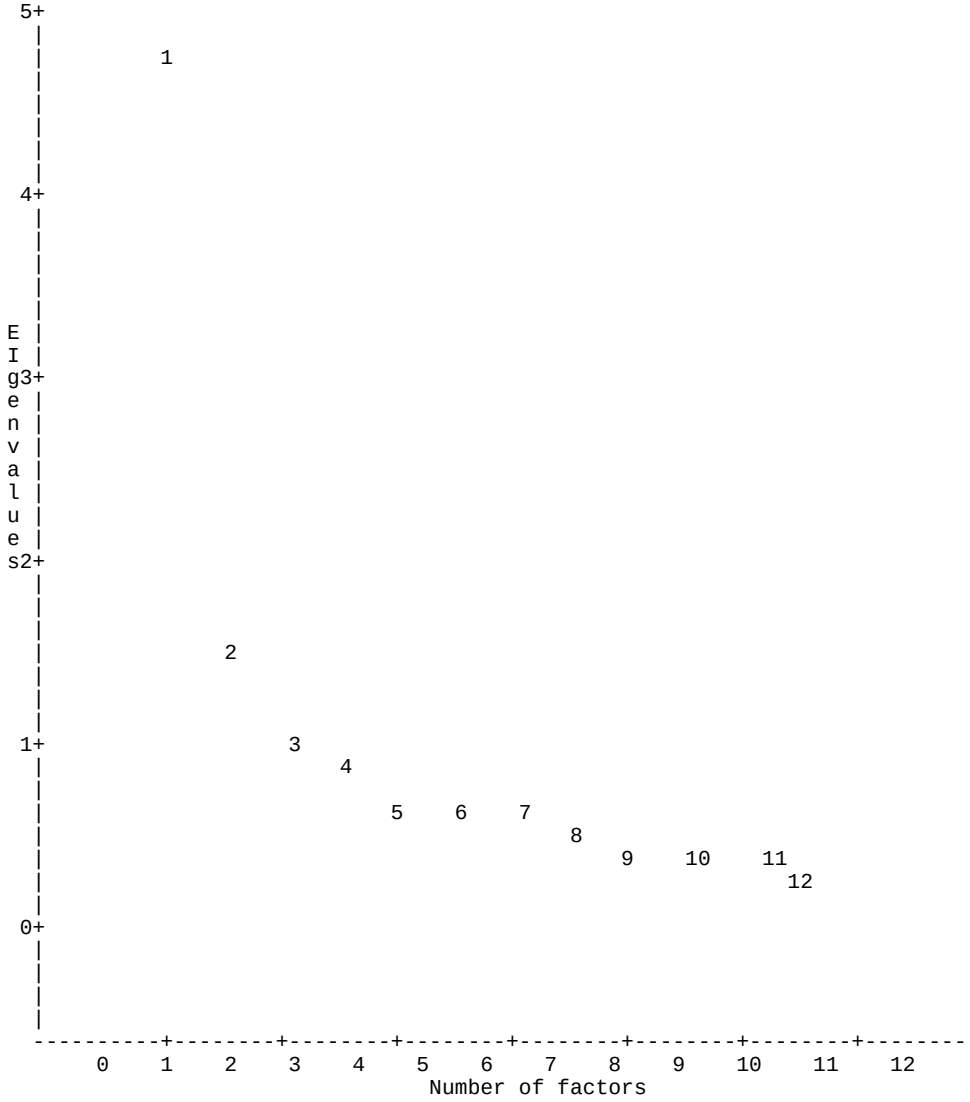

FIGURE 2. SCREE PLOT OF EIGENVALUES OF INDIVIDUAL FACTORS OF SF-12 ITEMS IN RHEUMATOID ARTHRITIS TRIALS

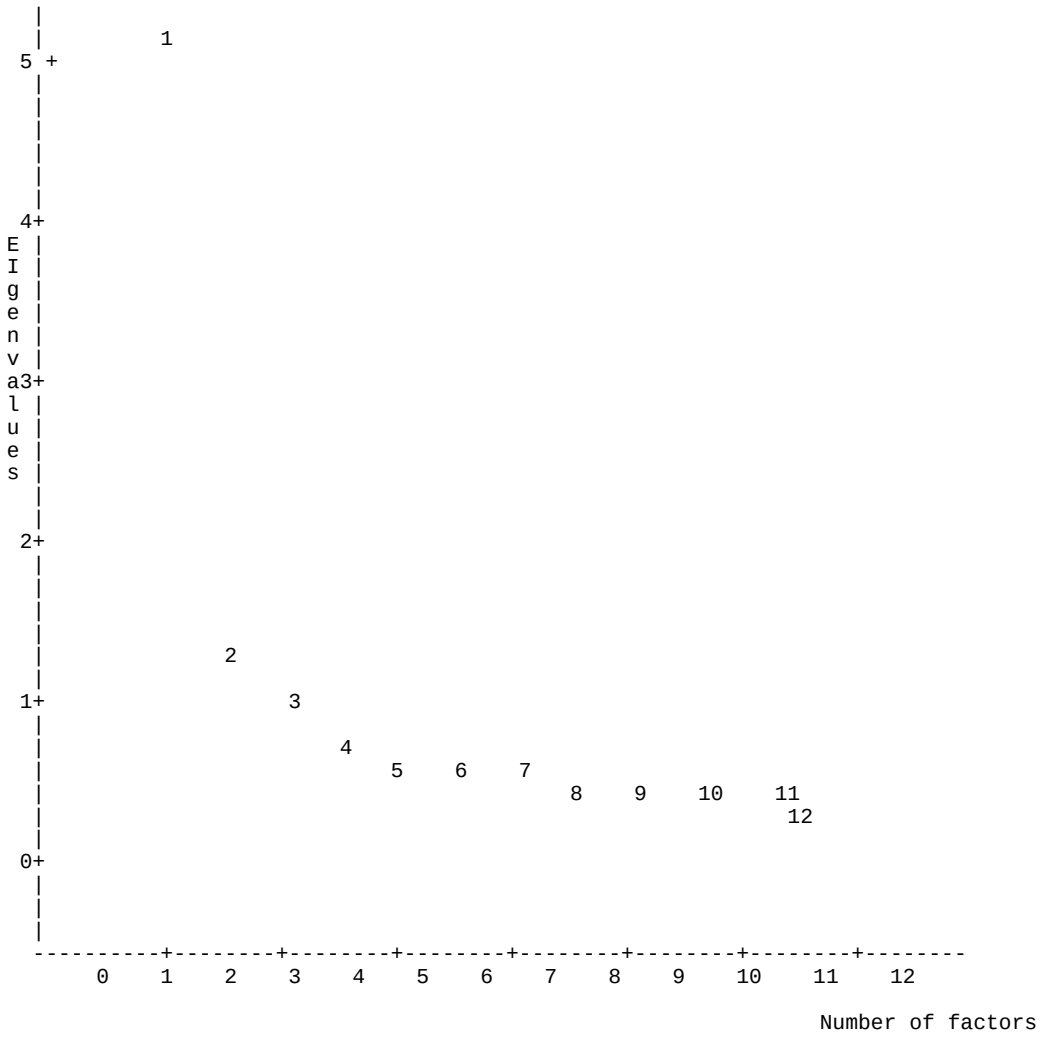

TABLE 3. FACTOR STRUCTURE OF SF-12 ITEMS USING PRINCIPAL COMPONENT ANALYSIS (VARIMAX ROTATION) IN OSTEOARTHRITIS

	Osteoarthritis		Rheumatoid arthritis	
ITEM	FACTOR 1	FACTOR 2	FACTOR 1	FACTOR 2
PF2	0.71*	0.12	0.74*	0.04
PF4	0.69*	-0.03	0.71*	0.07
RP2	0.70*	0.22	0.68*	0.33
RP3	0.70*	0.13	0.70*	0.24
BP2	0.74*	0.27	0.72*	0.33
GH1	0.48*	0.42*	0.41*	0.46*
VT2	0.48*	0.48*	0.54*	0.42*
SF2	0.61*	0.44*	0.53*	0.56*
RE2	0.24	0.72*	0.21	0.74*
RE3	0.15	0.73*	0.20	0.69*
MH3	0.11	0.72*	0.20	0.67*
MH4	0.06	0.72*	0.04	0.78*

The loadings above 0.4 are flagged by '*'
A brief description of individual items is provided in Appendix B.

TABLE 4. INDIVIDUAL SF-12 ITEM-COMPONENT CORRELATIONS IN OSTEOARTHRITIS

Score	PF2	PF4	RP2	RP3	BP2	GH1	VT2	SF2	RE2	RE3	MH3	MH4
Osteoarthritis												
PCS12	0.70*	0.64*	0.68*	0.70*	0.76*	0.54*	0.47	0.53	0.16	0.20	0.11	0.08
MCS12	0.19	0.11	0.30	0.21	0.36	0.43	0.55*	0.61*	0.76*	0.73*	0.70*	0.68*
Rheumatoid arthritis												
PCS12	0.70*	0.64*	0.68*	0.70*	0.76*	0.54*	0.47	0.53	0.16	0.20	0.11	0.08
MCS12	0.19	0.11	0.30	0.21	0.36	0.43	0.55*	0.61*	0.76*	0.73*	0.70*	0.68*

$p < 0.05$, and n=592.
* Indicates higher correlation of a particular item with a SF-12 component.
A brief description of individual items is provided in Appendix B.

PCS12 - Physical Component Summary Score of SF-12; MCS12 - Mental Component Summary Score of SF-12.

TABLE 5. CORRELATION BETWEEN SF-12 COMPONENT SUMMARY MEASURES AND CLINICAL VARIABLES IN OSTEOARTHRITIS

Observation Period	Patient Global	Physician Global	Knee Pain on Weight Bearing	Knee Pain on Motion	Functional Capacity	Pain	Time to walk 50 feet
Baseline							
PCS12	-0.41 (592)	-0.31 (592)	-0.44 (422)	-0.35 (422)	-0.29 (483)	-0.43 (591)	-0.36 (420)
MCS12	-0.24 (592)	-0.22 (592)	-0.22 (422)	-0.22 (422)	-0.24 (483)	-0.18 (591)	-0.22 (420)
6 weeks							
PCS12	-0.53 (431)	-0.55 (454)	-0.55 (316)	-0.46 (316)	-0.37 (138)	-0.52 (450)	-0.35 (316)
MCS12	-0.23 (431)	-0.28 (454)	-0.27 (316)	-0.25 (316)	-0.30 (138)	-0.24 (450)	-0.27 (316)
Change score at week 6							
PCS12	-0.40 (399)	-0.37 (419)	-0.43 (291)	-0.30 (311)	-0.26 (128)	-0.43 (416)	-0.22 (290)
MCS12	N.S.	-0.17[a] (419)	-0.17[a] (291)	N.S.	-0.18[a] (128)	-0.21 (416)	-0.15 (290)

Values in the parantheses indicate number of subjects for each comparison. Zero-order Pearson's correlation coefficients are reported.[a] $p < 0.05$. For all other comparisons $p < 0.0005$.

PCS36 - Physical Component Summary of SF-36; MCS-36 - Mental Component Summary Score of SF-36; PCS12 - Physical Component Summary Score of SF-12; MCS12 - Mental Component Summary Score of SF-12.

TABLE 6. CORRELATION BETWEEN SF-12 COMPONENT SUMMARY MEASURES AND CLINICAL VARIABLES IN RHEUMATOID ARTHRITIS

Observation Period	Patient Global	Physician Global	Joint Tenderness / pain score	Joint swelling score	Pain	Duration of morning stiffness	Erythrocyte sedimentation rate	Functional capacity
Baseline								
PCS12	-0.49 (653)	-0.41 (653)	-0.30 (653)	-0.18 (654)	-0.45 (653)	-0.21 (654)	-0.21 (644)	-0.21 (654)
MCS12	-0.33 (653)	-0.22 (653)	-0.16 (654)	-0.10[a] (654)	-0.28 (653)	-0.10[a] (654)	-0.11[a] (654)	-0.10[a] (654)
4 weeks¥ and 6 weeks¤								
PCS12	-0.57 (494)	-0.54 (494)	-0.36 (494)	-0.24 (494)	-0.58 (494)	-0.29 (494)	-0.26 (477)	-0.37 (532)
MCS12	-0.30 (494)	-0.26 (494)	-0.14[a] (494)	-0.11[a] (494)	-0.34 (494)	-0.15[a] (494)	-0.09[a] (477)	- 0.11[a] (532)
4 week¥ and 6 week¤ change score								
PCS12	-0.39 (468)	-0.36 (468)	-0.23 (468)	-0.10[a] (468)	-0.46[a] (468)	-0.14[a] (468)	N.S.	-0.15 (505)
MCS12	-0.23 (468)	-0.23 (468)	N.S.	N.S.	-0.23 (468)	N.S.	N.S.	N.S.

Values in the parantheses indicate number of subjects for each comparison. Zero-order Pearson's correlation coefficients are reported. [a] - $p < 0.05$, For all other comparisons $p = 0.0001$.N.S. - not significant ($p > 0.05$). ¥ RA Trial II, ¤ RA Trial I. PCS12 - Physical Component Summary Score of SF-12; MCS12 - Mental Component Summary Score of SF-12.

Appendix A

Definition of clinical variables used in clinical trials

Trial	Clinical Variable	Severity level	Definition of severity levels
OA Trial I	KPWB[a]		
OA Trial II	0	None	No pain evident
	1	Mild	Pain evident but does not interfere with activities
	2	Moderate	Pain evident which interferes, but does not prevent performance of activities
	3	Severe	Pain evident which prevents performance of most activities
	4	Very Severe	Intolerable pain which prevents performance of all activities
OA Trial I	KPM[b]		
OA Trial II	0	None	No pain evident
	1	Mild	Pain evident but does not interfere with activities
	2	Moderate	Pain evident which interferes, but does not prevent performance of activities
	3	Severe	Pain evident which prevents performance of most activities
	4	Very Severe	Intolerable pain which prevents performance of all activities
OA Trial I OA Trial II OA Trial III RA Trial I RA Trial II	VAS[c] (pain assessment)*	N/A	Assessed on a 100 mm scale ranging from no pain to severe pain
OA Trial I	PhGA[d]		
OA Trial II	0	Very Good	Asymptomatic and no limitation of normal activities
OA Trial III¶ RA Trial I¶ RA Trial II¶	1	Good	Mild symptoms and no limitation of normal activities
	2	Fair	Moderate symptoms and limitation of some normal activities
	3	Poor	Severe symptoms and inability to carry out most normal activities
	4	Very Poor	Very severe symptoms and inability to carry out all normal activities
OA Trial I	PaGA[e]		
OA Trial II	0	Very Good	Asymptomatic and no limitation of normal activities
OA Trial III¶ RA Trial I¶ RA Trial II¶	1	Good	Mild symptoms and no limitation of normal activities
	2	Fair	Moderate symptoms and limitation of some normal activities
	3	Poor	Severe symptoms and inability to carry out most normal activities
	4	Very Poor	Very severe symptoms and inability to carry out all normal activities
OA Trial I OA Trial II	TTWF[f] (seconds)*	N/A	Time needed to walk a straight continuous distance of 50 feet as fast as possible without running
OA Trial I§ RA Trial I† RA Trial II†	DMS[g,*]	N/A	§Length of time patient's morning stiffness lasted within the past 24 hours † Interval of time of stiffness between time of awakening and time when patient is limber (average for 3 days)
RA Trial I RA Trial II	FCC[h]		
	1	Class I	Complete functional capacity with ability to carry on all usual duties without handicap
	2	Class II	Functional capacity adequate to conduct normal activities despite handicap or discomfort or limited mobility
	3	Class III	Functional capacity adequate to conduct only a few or none of the duties of usual occupation or self-care
	4	Class IV	Largely or wholly incapacitated with patient bedridden or confined to wheelchair, permitting little or no self-care

Appendix A (continued)

RA Trial I RA Trial II	JPT[i] 0 1 2 3	Higher response indicates worse pain/ tenderness	Response to pressure / motion: None (joint not tender) Positive response (joint tender) Spontaneous response (joint tender and winced) Withdrawal by patient (joint tender, winced and withdrew)
RA Trial I RA Trial II	JSS[j] 0 1 2 3	Higher response indicates worse joint swelling	Assessment of joint swelling: None Detectable synovial thickening without loss of bony contours Loss of bony contours Bulging synovial proliferation with cystic characteristics

* Patients were divided into groups (continuous variable converted to categorical variable) based on the scores on these measures for the purpose of analysis.
¶ The response categories were 1-5 instead of 0-4
[a]KPWB - Knee pain on Weight Bearing (Searle 1995)
[b]KPM - Knee Pain on Motion (Searle 1995)
[c]VAS pain - Visual Analogue Scale pain assessment (Fries 1983)
[d]PhGA - Physician Global Assessment of Arthritis Condition (Cooperating Clinics Committee of American Rheumatism Association 1965)
[e]PaGA - Patient global assessment of arthritis condition (Ward, Williams, et al, 1983; Cooperating Clinics Committee of American Rheumatism Association 1965)
[f]TTWF - Time to Walk 50 Feet (Searle 1995)
[g]DMS - Duration of Morning Stiffness (Cooperating Clinics Committee of American Rheumatism Association 1965)
[h]Functional Capacity Classification (Steinbrocker et al, 1949)
[i]AJT - Assessment of Joint Tenderness / Pain (Ritchie, Boyle, McInnes 1968)
[j]AJS - Assessment of Joint Swelling (Ward, Williams, et al, 1983; Cooperating Clinics Committee of American Rheumatism Association 1965)

Appendix B

SF-12 items and Respective Domains

Items	Scale	Item
Moderate Activities	Physical functioning (P)	PF2
Climb Several Flights	Physical functioning (P)	PF4
Accomplished Less	Role physical (P)	RP2
Limited in kind	Role physical (P)	RP3
Pain interfere	Bodily pain (P)	BP2
Health in general	General health (P)	GH1
Energy	Vitality (M)	VT2
Social-time	Social functioning (M)	SF2
Accomplished less	Role emotional (M)	RE2
Not careful	Role emotional (M)	RE3
Peaceful	Mental health (M)	MH3
Blue/sad	Mental health (M)	MH4

P - Physical component, M - Mental component.