Reading and Interpreting Economic Evaluations in Rheumatoid Arthritis: An Assessment of Selected Instruments for Critical Appraisal

MICHAEL H. WEISMAN, ANACLETO D. GANO Jr, SHERINE E. GABRIEL, MARC C. HOCHBERG, ARTHUR KAVANAUGH, MATTHEW H. LIANG, JOSHUA J. OFMAN, MARK PRASHKER, MARIA E. SUAREZ-ALMAZOR, EDWARD YELIN, SHOSHANNA D. NAKELSKY, and JOSEPH D. CROFT Jr, on behalf of the Evidence-Based Medicine Working Groups in Rheumatology

ABSTRACT. ***Objective.*** To describe and compare the relative attributes (reliability, ease of use, applicability, and relevance) of different assessment tools for economic analyses as they pertain to rheumatoid arthritis (RA) literature.

Methods. An expert panel, comprising rheumatology researchers and clinicians, operationalized 2 economic appraisal instruments and applied them to 11 articles used for analysis. Each expert reviewed 3 articles, with each article independently reviewed by a pair of experts. A summary score for each article per appraisal instrument was calculated by dividing the number of items that received a "positive" response by the total number of items in the appraisal instrument.

Results. Scores for each article were similar across reviewers and appraisal instruments.

Conclusion. There is a need for a more comprehensive approach for evaluating this rapidly growing body of economic literature that is not only valid and reliable, but also easy to apply and understand. Although consistency between reviewers was good on both guidelines, inter-guideline discrepancies were noted and reviewers reported some difficulty in using the operationalized format. (J Rheumatol 2003;30:1739–47)

Key Indexing Terms:
ECONOMICS RHEUMATOID ARTHRITIS

Rheumatoid arthritis (RA) is a chronic, inflammatory disease of unknown etiology that affects roughly 2 million American adults[1,2]. The economic burden of RA is substantial. Annual medical care costs in the United States are estimated to exceed $4,647 per patient, with total direct medical costs estimated at $4.2 billion annually[3]. Indirect costs, estimated between $13,180 and $65,880 per patient, accrue because the majority of patients with RA experience work disability within 10 years of disease onset[4]. These indirect costs are likely underestimated because of a lack of data on sustained losses in activities other than employment and the difficulty of pricing the costs of intangibles associated with pain and suffering from RA.

Understanding the economic burden of illness is critical

From the Evidence-Based Medicine Working Groups in Rheumatology and Zynx Health Incorporated, Beverly Hills, CA, USA.

Supported by an unrestricted educational grant by Amgen, Inc., Thousand Oaks, CA, USA.

J.D. Croft Jr, MD, Clinical Professor of Medicine (Rheumatology) Georgetown University Medical School, Washington, DC; S.E. Gabriel, MD, MSc, Chair, Department of Health Sciences Research, Professor of Medicine and Epidemiology, Health Sciences Research and Division of Rheumatology, Department of Internal Medicine, Mayo Clinic, Rochester, MN; A.D. Gano Jr, MPH, Research Associate, Zynx Health Inc., Beverly Hills, CA; S.D. Nakelsky, MPH, Research Associate; J.J. Ofman, MD, MSHS, Assistant Professor, Departments of Medicine and Health Services Research, UCLA School of Medicine, Los Angeles, CA, Senior Vice President of Research, Zynx Health Inc., Beverly Hills, CA; M.C. Hochberg, MD, MPH, Professor of Medicine, Department of Medicine, University of Maryland School of Medicine, Head, Division of Rheumatology and Clinical Immunology, Department of Medicine, Maryland Veterans Affairs Health Care System, Baltimore, MD; A. Kavanaugh, MD, The Center for Innovative Therapy, Division of Rheumatology, Allergy and Immunology, University of California, San Diego, CA; M.H. Liang, MD, MPH, Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Professor, Department of Health Policy and Management, Harvard School of Public Health, Boston, MA; M. Prashker, MD, MPH, Director, Center for Health Quality, Outcomes and Economic Research, Bedford VA Medical Center, Chair, Health Services Department, Boston University School of Public Health, Boston, MA; M.E. Suarez-Almazor, MD, MSc, PhD, Associate Professor of Medicine, Health Services Research, Veterans Affairs Medical Center, Baylor College of Medicine, Houston, TX; M.H. Weisman, MD, Director, Division of Rheumatology, Department of Medicine/Rheumatology, Cedars-Sinai Health Systems, Los Angeles, CA; E. Yelin, PhD, Professor of Medicine and Health Policy, Department of Medicine and Rosalind Russell Medical Research Center for Arthritis, University of California, San Francisco, CA.

The Evidence-Based Medicine (EBM) Working Groups in Rheumatology comprises subgroups in Disease Management, Health Economics, and Outcomes Assessment Instruments: Chair, J.D. Croft Jr, MD. Health Economics Working Group: M.H. Weisman, MD; S.D. Nakelsky, MPH, A.D. Gano Jr, MPH; S.E. Gabriel, MD, MSc; M.C. Hochberg, MD, MPH; A. Kavanaugh, MD; M.H. Liang, MD, MPH; M. Prashker, MD, MPH; J.J. Ofman, MD, MSHS; M.E. Suarez-Almazor, MD, MSc, PhD; E. Yelin, PhD.

Address reprint requests to Dr. J.J. Ofman, Zynx Health Inc., 9100 Wilshire Blvd., East Tower, Suite 655, Beverly Hills, CA 90212. E-mail: jofman@cerner.com

Submitted December 27, 2001; revision accepted January 16, 2003.

Personal, non-commercial use only. The Journal of Rheumatology Copyright © 2003. All rights reserved.

Downloaded on March 10, 2026 from www.jrheum.org

to determining optimal policy for disease management. Most randomized clinical trials, ranging from a few months to many years, cannot fully address the longterm effects of chronic illnesses, such as RA, with disease durations lasting up to several decades. It is imperative that employers, insurers, and government agencies recognize the economic impact of RA so that a fair, efficient, and appropriate allocation of scarce healthcare resources can be developed and implemented. The volume of medical literature addressing economic issues has increased greatly in the last decade: 65% of the health economics literature in RA has been published within the last 6 years.

The methodological quality of this body of work remains suboptimal[5]. Criteria have been adopted to guide critical appraisal of economic evaluations and to assist consumers in evaluating the validity and quality of these analyses[6-9]. Nevertheless, there is little information regarding their comparative reliability and clinical relevance.

Our primary objective was to describe and compare the reliability, ease of use, applicability, and relevance of various assessment tools for economic analyses as they pertain to the economic literature in RA. Our secondary objective was to apply these tools in appraising current RA economics literature and to identify areas for future research. We believe our exploratory analysis should enhance the ability of investigators and clinicians to use and interpret economic analyses.

MATERIALS AND METHODS

Article Selection and Assignment

We conducted a systematic review of the published literature from January 1995 through June 2001. The Medline, Cochrane, and Healthstar computerized bibliographic databases were searched using the following Medical Subject Headings (MeSH) and headings developed and approved by the working group: work disability, direct cost(s), indirect cost(s), employment, unemployment, downsizing, work, intangible costs, medical care expenditures, health expenditures, quality of life, quality adjusted life years, economics, cost, cost allocation, cost control, cost of illness, cost savings, cost sharing, cost benefit analysis, costs and cost analysis, or technology, high costs, and RA. We limited our search to English language articles focused on human subjects. The following MeSH headings were excluded: foreign language, non-human subjects, addresses, monograph, editorial, legal cases, review articles, abstracts, and interview.

We utilized a systematic review strategy in which explicit acceptance and rejection criteria were set for the title, abstract, and paper review. Any article related to economic analyses of RA treatments met the basic acceptance criteria for each review stage. The paper review incorporated the most stringent inclusion criteria, allowing for acceptance only articles that contained primary economic data related to the treatment of RA. Rejection criteria at each stage of the review excluded articles that (a) did not explicitly mention cost or a related concept, and (b) was unrelated to RA. This process was applied to each article, with the exception of one published paper[10] submitted by the working group, which was not identified by our search strategy. This paper did meet our inclusion criteria and was included in the critical appraisal.

Eight experts in the field participated in the working group. The members of the group had various backgrounds, including health economics, clinical trials, and clinical care. Eleven papers were ultimately selected through the review process and randomly allocated to 7 of the 8 working group experts for appraisal. Two experts independently reviewed each paper, but reviewer pairs did not necessarily overlap across articles. Three articles were randomly assigned to each expert for appraisal. Seven of the experts participated in the review.

Selection of Appraisal Criteria

We conducted a literature search of appraisal criteria and guidelines for economic analyses. We identified 7 articles[6-9,11-13] describing appraisal criteria, methodological criteria, and guidelines for conducting economic analyses. We identified 4 guidelines with clearly defined checklists[9,11-13]. For each of these instruments, all open-ended items were operationalized into "yes" or "no" questions for ease of use during article appraisal. Questions with 2 parts were divided into 2 separate appraisal items and included in the total item count. The 4 identified guidelines, along with their respective operationalized forms, were submitted to the working group for review. Each member of the working group ranked the operationalized appraisal instruments in terms of face validity, applicability, breadth, depth, relevance, simplicity, length of time, comprehensiveness, ease of use, and the desire to use appraisal criteria that differed from each other. The 2 appraisal instruments that ranked the highest by the majority of experts were used in this exploratory analysis[11,12].

Scoring

Seven experts independently reviewed 3 articles and applied both appraisal instruments to each article, so that 2 experts employing both criteria reviewed each article. One expert abstained from article appraisal. Methods for scoring were not provided by the authors of either appraisal guideline. Thus, we adopted the scoring technique used in the study by Gerard, *et al*[14] in which a percentage score was calculated by dividing the number of items with a "positive" response by the total number of items in the appraisal instrument. A "positive" response was dependent on the nature of the question and was assessed in one of 2 ways: (1) by a "yes" or "no" answer to the question, or (2) through comments that indicated a positive or negative response to the question. Questions with missing responses received no score/a score of zero. It was decided *a priori* that these items would still be included in the total number of items when calculating the percentage scores (Scoring Method 1). In addition, we also calculated the percentage scores under the assumption that items without responses were assigned as "positive" responses. As above, we followed the *a priori* decision that these items were to be included in the total number of items used when calculating the percentage scores (Scoring Method 2). A higher percentage score indicated that the article met more of the "high quality" criteria for an economic evaluation. This scoring technique assumed that (1) the weight for all questions was equal, and (2) all questions could be categorized into "yes" and "no" responses.

Working Group Survey

After the working group evaluated the selected articles, we surveyed each reviewer to assess their reaction to the 2 appraisal instruments regarding ease of use, breadth, depth, and relevance. A 5-point Likert-type scale[15] was used, where 1 equaled "poor" and 5 equaled "excellent." The experts were asked which instrument they preferred using and what improvements they would make to the instruments.

RESULTS

Consistency Between Appraisal Criteria

Two appraisal instruments, developed by Drummond, *et al*[11] and Sanchez, *et al*[12], were selected based on their perceived ease of use, level of comprehensiveness, and ability to operationalize. The Sanchez appraisal instrument included 37 questions (or items) organized into 11 domains (objective, perspective, type of analysis, study design, choice of interventions, costs and consequences, discounting, results,

Personal, non-commercial use only. The Journal of Rheumatology Copyright © 2003. All rights reserved.

Downloaded on March 10, 2026 from www.jrheum.org

sensitivity analysis, conclusions, and sponsorship) (Appendix 1). The Drummond appraisal instrument was composed of 43 questions in 10 domains (Appendix 2). The 2 instruments differed from each other in that the Drummond appraisal instrument utilized more objectively defined criteria and appeared to emphasize specific costs that were not considered by Sanchez, including market values and capital costs.

In general, scores were similar for both appraisal instruments (for each paper and each reviewer). Figure 1 shows that only one expert's evaluation (of the article by Al, *et al*[16] using the Drummond appraisal instrument) was not consistent with that expert's evaluation of the Al, *et al* article using the Sanchez appraisal instrument (circled in Figure 1).

Consistency Between Reviewers

For 7 of the 11 studies[10,17-22], scores were fairly consistent between the 2 independent reviewers when using each appraisal instrument (Figure 2). Scores for 3 studies[16,23,24] were clearly inconsistent when the Drummond appraisal instrument was used. In another study[25], reviewers' scores were inconsistent for both appraisal instruments.

Discrepant Responses for the Sanchez Appraisal Instrument

Out of the 37 questions on the Sanchez instrument, the independent reviewers achieved agreement of at least 50% or greater on a majority of the study items (89%). The question most often answered inconsistently was, "Was any justification given for the discount rate used?" (72% of the time). Two questions associated with discounting, "Was the study performed over time?" and "Did the study use dollar values from previous years?", were disagreed upon 45% of the time. Sixty-four percent of the reviewer pairs also responded differently to the question of whether any appropriate alternatives were omitted.

Discrepant Responses for the Drummond Appraisal Instrument

Out of the 43 questions on the Drummond instrument, the independent reviewers achieved agreement of 50% or more on 28 of the study items. In contrast, 15 questions were disagreed upon by at least 50% of the reviewer pairs. Frequently all questions about the accuracy and appropriateness of measurements of costs and consequences received inconsistent scores. Two additional questions about the discounting of costs and consequences were also sources of disagreement between the reviewers.

Appraisal of Economic Analyses in RA (Table 1)

The range of percentage scores for Scoring Method 1 was 43–95% for the Sanchez appraisal instrument and 26–81% for the Drummond instrument. Only one study[10] received consistently low scores, with only one appraisal scoring above 50%. As detailed in Table 1, the scores for the remaining 10 studies varied widely across reviewers and appraisal instruments. Using the Sanchez instrument, the mean score across RA studies was 75%, with 5 studies scoring above the mean for both reviewers[18,19,21-23]. The mean score using the Drummond instrument was 55%, with 3 studies scoring above the mean for both reviewers[21,22,24].

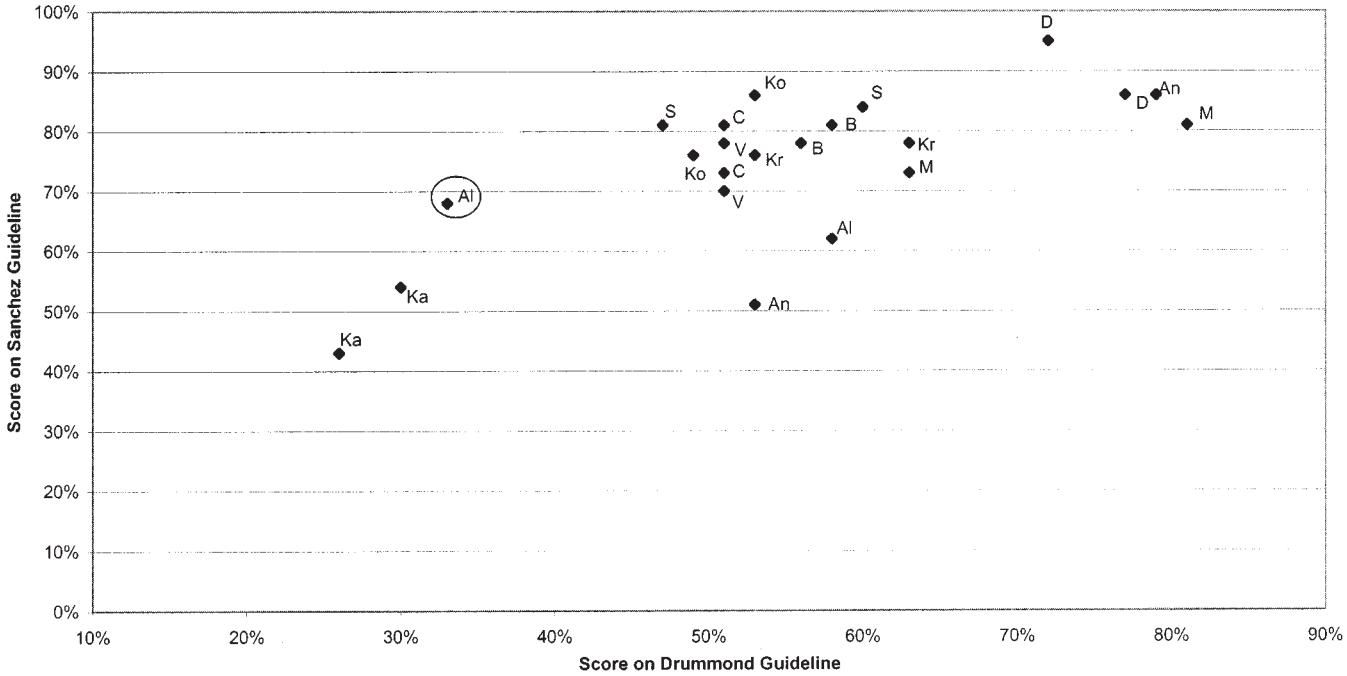

Figure 1. Correlation between scores on Drummond and Sanchez guidelines (points identified by initial "M," first author). Inconsistent evaluation circled.

Personal, non-commercial use only. The Journal of Rheumatology Copyright © 2003. All rights reserved.

Downloaded on March 10, 2026 from www.jrheum.org

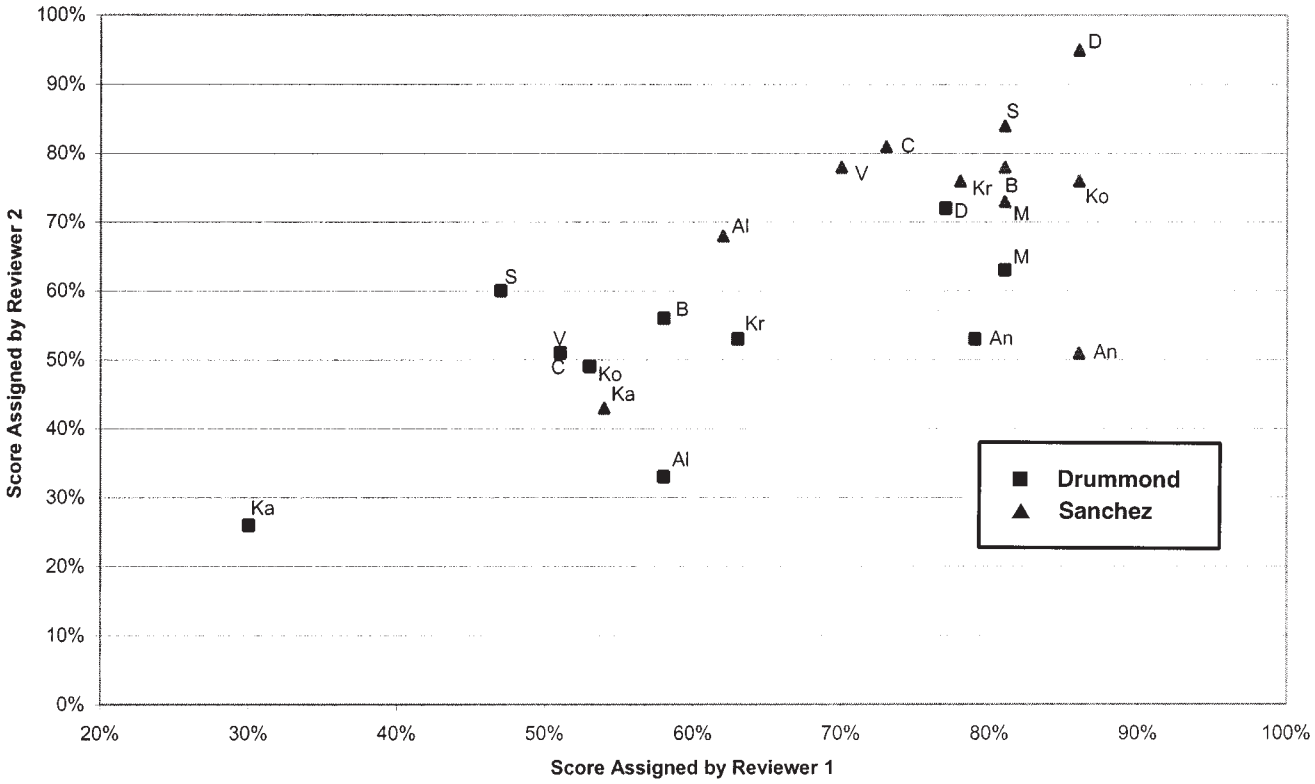

Figure 2. Correlation between scores assigned by Reviewer 1 and Reviewer 2 (points identified by first author's initial).

Table 1. Percentage (%) of "positive' responses for each study from each reviewer using the Sanchez[12] and Drummond[11] appraisal instruments.

	Appraisal Instrument					
	Sanchez (n = 37)			Drummond (n = 43)		
Article	Reviewer 1 Scoring Method 1 (Scoring Method 2)	Reviewer 2 Scoring Method 1 (Scoring Method 2)	Mean Scoring Method 1 (Scoring Method 2)	Reviewer 1 Scoring Method 1 (Scoring Method 2)	Reviewer 2 Scoring Method 1 (Scoring Method 2)	Mean Scoring Method 1 (Scoring Method 2)
Kavanaugh[10]	54 (65)	43 (65)	49 (65)	30 (44)	26 (37)	28 (41)
Al[16]	62 (78)	68 (68)	65 (73)	58 (74)	33 (60)	45 (67)
Choi[17]	73 (76)	81 (81)	77 (78)	51 (63)	51 (70)	51 (66)
Solomon[23]	81 (84)	84 (84)	82 (84)	47 (58)	60 (65)	53 (62)
Kristiansen[18]	78 (78)	76 (78)	77 (78)	63 (74)	53 (74)	58 (74)
Kobelt[19]	86 (86)	76 (76)	81 (81)	53 (53)	49 (51)	51 (52)
Verhoeven[20]	70 (70)	78 (78)	74 (74)	51 (53)	51 (51)	51 (52)
Maetzel[24]	81 (81)	73 (92)	77 (86)	81 (81)	63 (77)	72 (79)
Davey[21]	86 (86)	95 (95)	91 (91)	77 (79)	72 (93)	74 (86)
Bergquist[22]	81 (81)	78 (78)	80 (80)	58 (67)	56 (70)	57 (69)
Anis[25]	86 (86)	51 (59)	69 (73)	79 (79)	53 (70)	66 (74)
Overall mean			75 (79)			55 (66)

Scoring Method 1: Items with missing responses were assumed "zero" and included in the calculation of percentage scores. Scoring Method 2: Items with missing responses were assumed "positive" and included in the calculation of percentage scores.

Only 2 studies[21,22] received scores above the overall means of both appraisal instruments from both reviewers. Thus, the use of the different instruments may result in different assessments of the RA economic literature.

Table 1 also shows the percentage scores for each article if missing responses are assigned "positive values" rather than zero values (Scoring Method 2). The calculated mean scores for the Sanchez and Drummond instruments are higher when these items are included under the assumption of Scoring Method 2 (79% and 66%, respectively). For the Sanchez instrument, 3 studies[21,23,24] received scores above the mean from both reviewers. When the Drummond instru-

Personal, non-commercial use only. The Journal of Rheumatology Copyright © 2003. All rights reserved.

Downloaded on March 10, 2026 from www.jrheum.org

ment was used, 5 studies scored above the mean for both reviewers[18,21,22,24,25]. Two studies[21,24] attained scores above the overall means of both instruments from both reviewers. Regardless of the scoring method used, it was observed that the distribution of the missing values did not affect the overall results of this exercise.

Working Group Survey Results (Table 2)

To summarize the experts' experience using the 2 appraisal instruments, we conducted a brief informal survey using a 5-point Likert-type scale. The results of the survey are shown in Table 2. A score of 1 was equal to "poor," whereas a score of 5 equaled "excellent." Although not significant from a statistical standpoint, trends were apparent among the experts' responses. Compared to the Sanchez appraisal instrument, the Drummond instrument scored slightly higher in 3 categories: "applicability," "breadth," and "depth." However, in addition to scoring higher on "ease of use," the Sanchez instrument was also preferred more often than the Drummond instrument. Overall, 2 experts preferred the Drummond instrument, 3 the Sanchez instrument, and one expert had no preference.

DISCUSSION

The results of this exploratory analysis reveal that, in general, there was reasonably good consistency between reviewers on both appraisal instruments. However, the Sanchez appraisal instrument resulted in more consistent responses between reviewer pairs than the Drummond instrument. While consistent responses inform us about reliability, they do not provide any indication of the validity of the instruments. Alhough it appears that the Sanchez instrument was simpler and easier to use (thereby resulting in fewer discrepancies), the Drummond instrument may have enhanced comprehensiveness at the expense of ease of use. As a result, more discrepancies might be observed between reviewers using this instrument.

This difference in consistency was reflected in the experts' comments about the appraisal instruments. While the working group indicated that the Sanchez instrument was easier to use and "preferred," the Drummond instrument had greater perceived "applicability," "breadth," and "depth." Both appraisal instruments had similar perceived relevance to the RA literature. In addition to the appraisal instruments' characteristics, the differences in scores may also be attributable to the backgrounds of the reviewers, including the level of comfort and experience each reviewer has in working with economics literature. Further, the expert group felt there was a learning curve intrinsic to the process of grading economic evaluations with the appraisal criteria. For example, through our discussions of the results, the experts who indicated having more comfort and familiarity with the Drummond instrument also stated that they preferred the Drummond instrument versus the Sanchez instrument. Thus, it appears that some introduction to the appraisal criteria would aid in the review process and provide improved consistency between reviewers.

Table 2. Mean score on the working group survey (n = 6) (range).

	Drummond	Sanchez
Ease of use	3.2 (1–5)	3.7 (3–5)
Applicability	4.0 (3–5)	3.7 (3–4)
Breadth	4.2 (3–5)	3.8 (3–4)
Depth	4.0 (3–5)	3.3 (3–4)
Relevance	4.0 (3–5)	4.2 (3–5)
Preferred instrument*	2	3

* One expert found no overall preference for either instrument.

The use of RA-specific health economic information is increasingly relevant to stakeholders involved in decisions about RA disease management. These include formulary and policy decision-makers, journal editors and reviewers, investigators performing systematic reviews, and clinicians. As policy and clinical decision-makers integrate health economic information into their practices, familiarity with qualitative analysis and the ability to discriminate poor from high quality analyses become critical.

While many tools have been developed to address the methodological deficiencies of this body of literature and appraise their quality[6], several limitations exist[7-9]. The results of our study reveal several potential problems. None of the tools have undergone formal validation. None had scoring systems with weights assigned to the items acknowledging that not all items contribute equally to the methodological quality of such analyses. Finally, as demonstrated in this analysis, reliability is poor, suggesting that the items are often not well understood. Moreover, the tools need to be relevant to specific clinical areas, as there may be certain characteristics of economic analyses that are more relevant to the RA literature, yet are not captured by existing tools. For example, it is of particular importance that the patient populations of RA efficacy trials be comparable in regard to such variables as duration of illness and background therapy.

Our study identified a number of problems associated with the appraisal of economic analyses in RA. The results point to the need for additional research to confirm and extend our findings regarding reliability, ease of use, and clinical relevance, and to assess the validity of the existing instruments. In addition, the appraisals of RA studies using these tools appeared to be quite variable, some studies complying with most recommendations while others did not. These data emphasize the need for suitable, valid, and comprehensive guidelines for both researchers and research consumers in this area.

Because a number of experts indicated some difficulty or discomfort with using the appraisal instruments to evaluate the papers in this exercise, we collected suggestions on how

Personal, non-commercial use only. The Journal of Rheumatology Copyright © 2003. All rights reserved.

Downloaded on March 10, 2026 from www.jrheum.org

to improve the appraisal criteria and process. The experts suggested that a more explicit set of criteria might help to make the process easier. Clinicians would benefit more if the language was less technical and the concepts more direct and to the point. In addition, the applicability and relevance to RA would be improved by adding disease-specific criteria. However, the addition of disease-specific criteria could reduce the ability to compare the quality of studies in RA to studies involving other diseases. This trade-off should be considered in the development of any disease-specific instrument. A delicate balance needs to be established between the specificity to certain disease conditions and the comparability among broader categories of illness.

This exercise was subject to several limitations. The methodology was not designed for quantitative analysis, with the lack of a scoring algorithm being evident. The study design did not allow for measurement of statistical performance, such as agreement or correlation, and was not powered to address specific hypotheses. In addition, the development of a reliable instrument is hampered not only by differences between generic and disease-specific evaluations, but also by the lack of consensus on several issues concerning health economics. We acknowledge that the majority of existing quality appraisal instruments do not enable the user to weigh and assess the relative importance of the different study items. However, our goal was to highlight important issues in an effort to stimulate further discussion and additional empirical research. We wanted to assess the relevance, applicability, and reliability of the currently used appraisal instruments. Our objective was not to validate the existing instruments, but rather to explore the strengths and weaknesses of these instruments in their ability to provide insights into the quality and relevance of economic analyses as they pertain to RA. We found that a more explicit set of appraisal criteria or guidelines would be helpful to better understand the quality and relevance of the economic literature pertaining to RA. Another limitation is the relatively small number of articles available for review. This simply reflects the available literature in the context of our review strategy. Lastly, the group of experts selected for this exercise may have more experience with economic evaluations than the average clinician, leading to an overestimation of the ease of use of the appraisal criteria.

Summary

A comprehensive literature search for economic articles relating to RA was undertaken, and 2 guidelines for evaluating the economic medical literature were operationalized. A panel of rheumatology researchers and clinicians applied the operationalized guidelines to selected RA economic evaluations. Although consistency between reviewers was good on both guidelines, inter-guideline discrepancies were noted and reviewers reported some difficulty in using the operationalized format. There is need for a more comprehensive approach for evaluating this rapidly growing body of literature that is not only valid and reliable, but also easy to apply and understand. Such a tool should strengthen the rigor of future economic evaluations, and at the same time make their results more accessible and useful to both clinicians and policy-makers. This is particularly important for patients with RA, in view of the high cost of new and emerging therapies. With the introduction of new treatments and the likelihood of additional pharmacoeconomic analyses in the near future, valid and easy to operationalize evaluation tools are becoming increasingly important to interpret the growing body of RA-specific health economic literature.

REFERENCES

1. Lawrence RC, Helmick CG, Arnett FC, et al. Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the United States. Arthritis Rheum 1998;41:778-99.
2. Goronzy JJ, Weyand CM. Rheumatoid arthritis: A. Epidemiology, pathology, and pathogenesis. In: Klippel JH, editor. Primer on the Rheumatic Diseases, 12th ed. Atlanta: The Arthritis Foundation; 2001:209-17.
3. Ward MM, Javitz HS, Yelin EH. The direct cost of rheumatoid arthritis. Value in Health (ISPOR) 2000;3:243-252.
4. Callahan LF. The burden of rheumatoid arthritis: facts and figures. J Rheumatol 1998;25 Suppl 53:8-12.
5. Neumann PJ, Stone PW, Chapman RH, Sandberg EA, Bell CM. The quality of reporting in published cost-utility analyses, 1976-1997. Ann Intern Med 2000;132:964-72.
6. Neumann PJ, Hermann RC, Berenbaum PA, Weinstein MC. Methods of cost-effectiveness analysis in the assessment of new drugs for Alzheimer's disease. Psychiatr Serv 1997;48:1440-4.
7. Detsky AS. Guidelines for economic analysis of pharmaceutical products: a draft document for Ontario and Canada. Pharmacoeconomics 1993;3:354-61.
8. Gold MR, Siegel JE, Russell LB, Weinstein MC. Cost-effectiveness analysis in health and medicine. New York: Oxford University Press; 1996.
9. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the British Medical Journal. BMJ 1996;313:275-83.
10. Kavanaugh A, Heudebert G, Cush J, Jain R. Cost evaluation of novel therapeutics in rheumatoid arthritis (CENTRA): a decision analysis model. Semin Arthritis Rheum 1996;25:297-307.
11. Drummond M, O'Brien B, Stoddart G, Torrance GW. Critical assessment of economic evaluation. Methods for the economic evaluation of health care programmes. Oxford: Oxford Medical Publications; 1997:27-51.
12. Sanchez LA. Evaluating the quality of published pharmacoeconomic evaluations. Hosp Pharm 1995;30:146-2.
13. Sacristan JA, Soto J, Galende I. Evaluation of pharmacoeconomic studies: utilization of a checklist. Ann Pharmacother 1993; 27:1126-33.
14. Gerard K, Seymour J, Smoker I. A tool to improve quality of reporting published economic analyses. Int J Technol Assess Health Care 2000;16:100-10.
15. Psychology of Communication: Attitudes (The Likert Scale). The CultSock web site. Available at: http://www.cultsock.ndirect.co.uk/MUHome/cshtml/psy/likert.html - 2001.
16. Al MJ, Michel BC, Rutten FF. The cost effectiveness of diclofenac plus misoprostol compared with diclofenac monotherapy in patients with rheumatoid arthritis. Pharmacoeconomics 1996;10:141-51.

Personal, non-commercial use only. The Journal of Rheumatology Copyright © 2003. All rights reserved.

Downloaded on March 10, 2026 from www.jrheum.org

Appendix 1: Sanchez Appraisal Instrument with overall response rates (Eleven articles assessed by 2 reviewers = 22 possible responses per item)

Appraisal Instrument Questions		Positive	Zero	Missing
I. Objective				
1. Was the study objective clearly defined?	Y=1	22	0	0
2. Is the study question(s) clear, defined and measurable?	Y=1	21	1	0
II. Perspective				
3. What is the perspective(s) of the analysis?	Y=1	21	1	0
4. Is it appropriate given the scope of the problem?	Y=1	16	4	2
III. Type of Analysis				
5. Was a pharmacoeconomic tool used?	Y=1	16	4	2
6. Is it appropriate given the problem?	Y=1	15	3	4
7. Is it what actually was conducted?	Y=1	15	2	5
IV. Study Design				
8. What was the study design?	Y=1	22	0	0
9. What were the data sources?	Y=1	21	1	0
10. Is the evaluation suitable if carried out in a clinical trial?	Y=1	12	9	1
V. Choice of Interventions				
11. Were all appropriate alternatives considered?	Y=1	11	11	0
12. Were any appropriate alternatives omitted?	N=1	9	12	1
13. Are the alternatives relevant to the perspective and clinical nature of the study?	Y=1	20	1	1
14. Were the alternatives appropriately described?	Y=1	17	4	1
VI. Costs and Consequences				
15. What are the costs and consequences (outcomes)?	Y=1	21	0	1
16. Were all the important and relevant costs and consequences for each alternative identified?	Y=1	9	13	0
17. Are the costs and outcomes relevant to the perspective chosen?	Y=1	18	3	1
18. Do they include negative outcomes treatment failures, adverse events)?	Y=1	17	5	0
19. Were all sources of values clearly identified?	Y=1	20	2	0
20. Were costs and consequences measured in the appropriate physical units?	Y=1	18	4	0
VII. Discounting				
21. Was the study performed over time?	Y=1	17	5	0
22. Did the study use dollar values from previous years?	Y=1	10	11	1
23. Are any costs or consequences that occur in the future discounted to reflect their present values?	Y=1	10	11	1
24. Was any justification given for the discount rate used?	Y=1	6	11	5
VIII. Results				
25. Are the results accurate?	Y=1	21	1	0
26. Are the results practical for medical decision makers?	Y=1	18	4	0
27. Were the appropriate statistical analyses performed?	Y=1	19	1	2
28. Was an incremental analysis performed?	Y=1	19	2	1
29. Are all the assumptions and limitations of the study discussed?	Y=1	16	6	0
30. Was the generalizability of the results to other settings or patient populations discussed?	Y=1	14	8	0
IX. Sensitivity Analysis				
31. Are cost ranges for significant variables tested for sensitivity?	Y=1	21	1	0
32. Are the appropriate and relevant variables varied?	Y=1	20	2	0
33. Do the findings follow the anticipated trend?	Y=1	18	3	1
X.Conclusions				
34. Are the conclusions of the study justified?	Y=1	20	2	0
35. Is it possible to extrapolate the conclusions to daily clinical practice?	Y=1	12	10	0
XI. Sponsorship				
36. Was this an industry-sponsored or an industry-conducted study?	N=1	10	12	0
37. Was there any bias to sponsorship of the study?	N=1	17	4	1

Personal, non-commercial use only. The Journal of Rheumatology Copyright © 2003. All rights reserved.

Downloaded on March 10, 2026 from www.jrheum.org

Appendix 2: Drummond Appraisal Instrument with overall response rates (Eleven articles assessed by 2 reviewers = 22 possible responses per item)

Appraisal Instrument Questions		Positive	Zero	Missing
1. Was a well-defined question posed in answerable form?	Y=1	20	0	2
1.1 Did the study examine both costs and effects of the service(s) or program(s)?	Y=1	20	2	0
1.2 Did the study involve a comparison of alternatives?	Y=1	21	1	0
1.3 Was a viewpoint for the analysis stated and was the study placed in any particular decision-making context?	Y=1	19	2	1
2. Was a comprehensive description of the competing alternatives given (i.e.: can you tell who did what to whom, where and how often)?	Y=1	11	7	4
2.1 Were any important alternatives omitted?	N=1	7	12	3
2.2 Was (Should) a *do-nothing* alternative (be) considered?	Y=1	7	12	3
3. Was the effectiveness of the programs or services established?	Y=1	11	4	7
3.1 Was this done through a randomized, controlled clinical trial?	Y=1	12	10	0
3.11 If so, did the trial protocol reflect what would happen in regular practice?	Y=1	7	14	1
3.2 Was the effectiveness established through an overview of the clinical studies?	Y=1	9	13	0
3.3 Were observational data or assumptions used to establish effectiveness? If so, what are the potential biases in results?	N=1	10	12	0
4. Were all the important and relevant costs and consequences for each alternative identified?	Y=1	10	8	4
4.1 Was the range wide enough for the research question at hand?	Y=1	17	4	1
4.2 Did it cover all relevant viewpoints? (Possible viewpoints include the community or social viewpoint, and those of patients and third-party payers. Other viewpoints may also be relevant depending upon the particular analysis.)	Y=1	8	14	0
4.3 Were capital costs, as well as operating costs, included?	Y=1	2	19	1
5. Were costs and consequences measured accurately in appropriate physical units (e.g.: hours of nursing time, number of physician visits, lost work days, gained life-years)?	Y=1	17	5	0
5.1 Were any of the identified items omitted from measurement?	N=1	16	5	1
5.11 If so, does this mean that they carried no weight in the subsequent analysis?	N=1	8	5	9
5.2 Were there any special circumstances (e.g.: joint use of resources) that made measurement difficult?	N=1	15	5	2
5.21 Were these circumstances handled appropriately?	Y=1	8	10	4
6. Were costs and consequences valued credibly?	Y=1	11	1	10
6.1 Were the sources of all values clearly identified? (Possible sources include market values, patient or client preferences and views, policy-makers' views and health professionals' judgments.)	Y=1	19	2	1
6.2 Were market values employed for changes involving resources gained or depleted?	Y=1	17	5	0
6.3 Where market values were absent (e.g.: volunteer labor), or market values did not reflect actual values (such as clinic space donated at a reduced rate), were adjustments made to approximate such market values?	Y=1	13	8	1
6.4 Was the valuation of consequences appropriate for the question posed (i.e.: has the appropriate type or types of analysis - cost-effectiveness, cost-benefit, cost-utility - been selected)?	Y=1	18	3	1

Appendix 2 ***continued***

Personal, non-commercial use only. The Journal of Rheumatology Copyright © 2003. All rights reserved.

Downloaded on March 10, 2026 from www.jrheum.org

Appendix 2 *continued*

7. Were the costs and consequences adjusted for differential timing?	Y=1	3	11	8
7.1 Were costs and consequences, which occur, in the future 'discounted' to their present values?	Y=1	9	11	2
7.2 Was any justification given for the discount rate used?	Y=1	5	14	3
8. Was an incremental analysis of costs and consequences of alternatives performed?	Y=1	12	3	7
8.1 Were the additional (incremental) costs generated by one alternative over another compared to the additional effects, benefits, or utilities generated?	Y=1	18	4	0
9. Was allowance made for uncertainty in the estimates of costs and consequences?	Y=1	14	2	6
9.1 If data on costs or consequences were stochastic, were appropriate statistical analyses performed?	Y=1	12	6	4
9.2 If a sensitivity analysis was employed, was justification provided for the range of values (for key study parameters)?	Y=1	16	6	0
9.3 Were study results sensitive to changes in the values (within the assumed range for sensitivity analysis, or within the confidence interval around the ratio of costs to consequences?	Y=1	8	14	0
10. Did the presentation and discussion of study results include all issues of concern to users?	Y=1	9	6	7
10.1 Were the conclusions of the analysis based on some overall index or ratio of costs to consequences (e.g.: cost-effectiveness ratio)?	Y=1	18	4	0
10.11 If so, was the index interpreted intelligently or in a mechanistic fashion?	Y=1	18	3	1
10.2 Were the results compared with those of others who have investigated the same question?	Y=1	13	8	1
10.21 If so, were allowances made for the potential differences in study methodology?	Y=1	11	9	2
10.3 Did the study discuss the generalizability of the results to other settings and patient/client groups?	Y=1	12	9	1
10.4 Did the study allude to, or take account of, other important factors in the choice or decision under consideration (e.g.: distribution of costs and consequences, or relevant ethical issues)?	Y=1	6	16	0
10.5 Did the study discuss issues of implementation, such as the feasibility of adopting the 'preferred' program given existing financial or other constraints, and whether any freed resources could be redeployed to other worthwhile programs?	Y=1	6	16	0

17. Choi HK, Seeger JD, Kuntz KM. A cost-effectiveness analysis of treatment options for patients with methotrexate-resistant rheumatoid arthritis. Arthritis Rheum 2000;43:2316-27.
18. Kristiansen IS, Kvien TK, Nord E. Cost effectiveness of replacing diclofenac with a fixed combination of misoprostol and diclofenac in patients with rheumatoid arthritis. Arthritis Rheum 1999;42:2293-302.
19. Kobelt G, Eberhardt K, Jonsson L, Jonsson B. Economic consequences of the progression of rheumatoid arthritis in Sweden. Arthritis Rheum 1999;42:347-56.
20. Verhoeven AC, Bibo JC, Boers M, Engel GL, van der Linden S. Cost-effectiveness and cost-utility of combination therapy in early rheumatoid arthritis: randomized comparison of combined step-down prednisolone, methotrexate and sulphasalazine with sulphasalazine alone. COBRA Trial Group. Combinatietherapie Bij Reumatoide Artritis. Br J Rheumatol 1998;37:1102-9.
21. Davey PJ, Meyer E. The cost effectiveness of misoprostol prophylaxis alongside long-term nonsteroidal anti-inflammatory drugs. Implications of the MUCOSA trial. Pharmacoeconomics 2000;17:295-304.
22. Bergquist SR, Felson DT, Prashker MJ, Freedberg KA. The cost-effectiveness of liver biopsy in rheumatoid arthritis patients treated with methotrexate. Arthritis Rheum 1995;38:326-33.
23. Solomon DH, Kuntz KM. Should postmenopausal women with rheumatoid arthritis who are starting corticosteroid treatment be screened for osteoporosis? A cost-effectiveness analysis. Arthritis Rheum 2000;43:1967-75.
24. Maetzel A, Ferraz MB, Bombardier C. The cost-effectiveness of misoprostol in preventing serious gastrointestinal events associated with the use of nonsteroidal antiinflammatory drugs. Arthritis Rheum 1998;41:16-25.
25. Anis AH, Tugwell PX, Wells GA, Stewart DG. A cost effectiveness analysis of cyclosporine in rheumatoid arthritis. J Rheumatol 1996;23:609-16.

Personal, non-commercial use only. The Journal of Rheumatology Copyright © 2003. All rights reserved.

Downloaded on March 10, 2026 from www.jrheum.org