

The OMERACT Initiative. Towards a Reference Approach to Derive QALY for Economic Evaluations in Rheumatology

ANNELIES BOONEN, ANDREAS MAETZEL, MICHAEL DRUMMOND, MARIA SUAREZ-ALMAZOR, MARK HARRISON, VIVIAN WELCH, and PETER S. TUGWELL

ABSTRACT. Within the OMERACT Economics Workgroup, an initiative was started to work towards consensus on the approach to calculate quality-adjusted life-years (QALY) in rheumatology. We report on a first meeting May 7, 2008, in Toronto attended by rheumatologists and experts in QALY. Following a summary of an international QALY workshop of pharmacoeconomists conducted under the umbrella of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR), participating experts identified a series of high-level generic principles to be considered for QALY estimations. The OMERACT workgroup then addressed specific issues rheumatologists should concentrate on in research to build consensus on QALY in rheumatology; discussion was based on results of a Web-based survey, conducted prior to the meeting, to identify attributes of a QALY considered important and approaches considered suitable for the QALY estimations in rheumatology. One priority was to further explore indirect approaches to QALY estimation as representation of patients' preference for health in clinical decisions and to explore the additional value of patients' preferences versus societal preferences in allocation decisions. The role of the different descriptive systems and their influence on the QALY, the role of the visual analog scale to value preferences, and comparison of methods to integrate utility over time were also identified as research priorities. Approaches should be easy to apply, easy to understand by different parties, reproducible, and sensitive to change. (J Rheumatol 2009;36:2045–9; doi:10.3899/jrheum.090355)

Key Indexing Terms:

MUSCULOSKELETAL DISEASE
OMERACT

QUALITY-ADJUSTED LIFE-YEARS
RECOMMENDATION

From the Department of Internal Medicine, Division of Rheumatology, Maastricht University Medical Center, Maastricht, The Netherlands; Centre for Health Economics, University of York, York, UK; Department of General Internal Medicine, University of Texas M.D. Anderson Cancer Center, Houston, TX, USA; Arthritis Research Campaign (arc) Epidemiology Unit, The University of Manchester, Manchester, UK; and Centre for Global Health, University of Ottawa, Ottawa, Ontario, Canada.

Supported by an OMERACT grant provided for organization of the Special Interest Group.

A. Boonen, MD, PhD, Department of Internal Medicine, Division of Rheumatology, Maastricht University Medical Center, Maastricht; A. Maetzel, MD, MSc, PhD, Director, Health Economics, Amgen (Europe), Assistant Professor, Clinical Epidemiology, Department of Health Policy, Management and Evaluation, University of Toronto; M. Drummond, MSc, PhD, Professor of Health Economics, Centre for Health Economics, University of York; M.E. Suarez-Almazor, MD, PhD, Professor and Deputy Chair, Department of General Internal Medicine, University of Texas M.D. Anderson Cancer Center; M. Harrison, MSc, PhD, arc Epidemiology Unit, The University of Manchester; V. Welch, MSc, Centre for Global Health, University of Ottawa; P.S. Tugwell, MD, MSc, FRCPC, Canada Research Chair, Professor of Medicine and Epidemiology, University of Ottawa.

Address correspondence to Dr. A. Boonen, Maastricht University Medical Center, Department of Internal Medicine, Division of Rheumatology, P Debyelaan 25, 6229 HX Maastricht, The Netherlands.
E-mail: a.boonen@mumc.nl

The OMERACT Economics Group first convened in 1997, and various members have been working together actively. The group's aim is to promote the development and standard use of rigorous scientific methods for economic evaluations in rheumatology¹. One of the issues identified as especially relevant for economic evaluations is "how to value health" as an adjustment factor for the health-related quality of the life-years spent with a rheumatic disorder². While the quality-adjusted life-year (QALY) was accepted as conceptually appropriate, different approaches to health valuation can provide different QALY, to the extent that this might change clinical conclusions³ and economic decisions⁴. This inconsistency led members of the Economics Group to concentrate on the question of whether there could be a standard recommended approach to QALY estimation for patients with rheumatic disorders.

Why "value health" when we have in rheumatology validated instruments to assess effectiveness of treatments, including instruments to assess overall health status? The quest for an approach to value health came from health economists and decision analysts who needed a method to

synthesize the most relevant dimensions of health in a single, universal indicator (a) to allow comparisons of cost-effectiveness ratios between interventions for various diseases, and (b) to weight these ratios according to values and preferences⁵. It was in 1968 that the concept (not yet the term) of utility was proposed by decision analysts as a method to value health by assessing preference for health states⁶. Over the years, utility became accepted as the theoretically most defendable approach for weighting life-years to calculate QALY. A utility values a specific health state by asking about the preference subjects have for that health state when confronted with a decision problem, being the choice between health states, on a scale from zero (the value for death) to one (the value for perfect health)⁵. The QALY combines the utility with the time lived in that health state. Although some health economists argue that health gains or losses should be valued not in terms of QALY but in terms of costs (such as willingness to pay), the QALY became the approach favored to inform decision-makers on health changes when allocating resources in healthcare⁷⁻¹⁴.

While it is clear that the QALY was developed and used most frequently for economic decisions in healthcare, its theoretical origin within decision-making theory allows this “universal value for health” to be used as an instrument for comparing clinical decisions. However, this aspect of the QALY remained largely unexplored. Assessment of preference for health under conditions of uncertainty was perceived as being closest to decisions made in real life, where choices about treatments are made with varying degrees of uncertainty about the outcomes. This resulted in the promotion of the standard gamble (SG) method as the theoretically most defendable approach to utility derivation (choice between 2 uncertain outcomes). However, because the standard gamble exercise is often not well understood by respondents, other direct, experience-based approaches to “utility” derivation emerged, such as the time tradeoff (TTO) method (one health outcome is uncertain in the choice experiment) or the feeling thermometer (no choice experiment, no uncertainty about the chosen outcome). Further simplification and the need to adopt the societal perspective led to the use of indirect approaches to the derivation of utilities for health states. Short questionnaires were developed, such as the EuroQOL (EQ-5D)¹⁵, the Health Utilities Index (HUI)¹⁶, or the Medical Outcome Study Short Form Survey 6D (SF-6D)¹⁷, for which the authors then obtained utility tariffs from the general population via the SG or TTO method⁵. Other investigators applied mathematical transformation algorithms to disease-specific health status measures, e.g., the Health Assessment Questionnaire (HAQ), to translate these into utilities¹⁸. As a result, different QALY may be calculated, depending on the approach used and favored by the author or study promoter. Discrepancies among QALY, on the other hand, defeat the initial purpose of using QALY to

promote comparability and transparency between cost-effectiveness ratios of competing interventions.

The OMERACT Economics Group sought to reach potential consensus on a standard approach to QALY estimation in rheumatic disorders. We report on the OMERACT Economics Group meeting held May 7, 2008, in Toronto, which was attended by 9 health economists or health scientists, 5 rheumatologists with an interest in health economics, and 5 other experts with experience in the use of QALY. During that meeting, results of preliminary work were presented and discussed, and some issues towards a research agenda were formulated. For each part of the meeting, a brief summary is provided.

THE ISPOR INITIATIVE “MOVING THE QALY FORWARD”

Michael Drummond described the workshop “Moving the QALY forward, building a pragmatic road,” organized under the umbrella of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR), to address health economists’ similar concerns about QALY. The workshop discussed concerns and explored possible agreement on a general method for estimating QALY. In view of the diverse visions and opinions, the workshop concluded that it is not yet possible to specify a preferred method for estimating QALY. However, consensus was reached on a set of higher level principles relating to QALY. To be reported in full in a workshop report by Smith, *et al*¹⁹, the principles are briefly: (a) that QALY can be used at various levels in the healthcare system (e.g., for broad resource allocation decisions and patient-level choices); (b) that both community preferences and individual patient experiences were relevant; and (c) that distributive issues (i.e., equity) need to be addressed when using QALY. In addition, research priorities relating to each of the principles were identified. Finally, the workshop concluded that there was a need for a reference case for the estimation of QALY if their use was to be encouraged¹⁹.

The discussion that followed by the OMERACT QALY subgroup identified potential roles for OMERACT when further exploring the QALY as an instrument in decision-making in rheumatology to:

1. Further explore the role of QALY in clinical decision-making at the level of groups of patients and individual patients and not only for decisions on resource allocation. This is appealing to OMERACT and would fit in the needs for the grades of recommendation, assessment, development, and evaluation (GRADE) system (a structured system for rating quality of evidence and grading strength of recommendation in clinical practice), in which the patient preference of health is important²⁰. At this point, there is insufficient evidence of the extent QALY truly reflect (revealed) patient preferences;
2. Gain insight into the discrepancy between patient prefer-

ence and societal preference and the additional role of patient preference in allocation decisions;

3. Continue appraisal and improvement of instruments and methods to calculate utility and derive QALY. Specific issues were to explore the role of the differences of the descriptive systems; the effect of approaches to integrate preference over time; and the role of the visual analog scale (VAS) to value health;

4. Investigate the relation between well-being and QALY: This is of interest to OMERACT and could be considered as an area of further study.

EXISTING EVIDENCE ON VALIDITY OF INDIRECT UTILITY INSTRUMENTS IN RHEUMATOID ARTHRITIS

A number of generic utility scales for weighting QALY have been applied in rheumatoid arthritis (RA). However, each must be validated for use within specific diseases. The evidence published between 1980 and mid-2006 regarding the validity and comparative performance of generic utility scales in RA was reviewed and presented at the meeting by Mark Harrison. The review focused on reports describing primary evidence of validity or performance of a generic utility scale, and evidence was selected, summarized, and reviewed using the OMERACT filter³.

Of the 923 instruments identified in the systematic search, 26 related to validation of generic utility instruments in RA comprising EQ-5D¹⁵, Health Utility Index-2 (HUI2) and HUI3¹⁶, SF-6D¹⁷, and Quality of Well-Being Scale²¹. The EQ-5D, HUI2 and HUI3, and SF-6D all have consistent evidence of construct validity and responsiveness in RA, but each has limitations. Further, differences between the measures exist in the aspects of health that are measured, the method of estimation of the utility values, and the range of possible values. The EQ-5D and HUI3 have been the most widely studied instruments in RA, while the SF-6D is relatively new but appears to have potential for use particularly in milder RA. The EQ-5D based on currently used tariffs is a nonlinear scale with a long tail in the low values. Although standardized measures of responsiveness such as effect size (change in units of standard deviations, SD; an effect size of 1 represents change equal to 1 SD) suggest each of the measures is responsive to change, they may underestimate the impact of using alternative measures in practice. For example, in patients experiencing an improvement in symptoms, the SF-6D is one of the most consistently responsive measures. However, absolute mean change for the SF-6D is smaller, leading to lower effect estimates. In an intervention study in RA the gains in QALY over 10 years were modelled using different utility measures (all estimated from HAQ). The SF-6D provided much more pessimistic QALY (0.89 ± 0.34) than the HUI3 (1.95 ± 0.62) and EQ-5D (1.35 ± 0.46). When accepting a threshold of US\$50,000 per QALY, 91% of simulations would be accepted based on HUI3 versus

63% and 12% for EQ-5D and SF-6D, respectively⁴. In view of the differences between the properties and performance of the measures, there is need for further research into the impact of applying utility values collected using different generic utility measures on the resulting cost-effectiveness conclusions.

RESULTS OF THE OMERACT QALY SURVEY

In preparation for the OMERACT QALY subgroup meeting, an E-mail survey was conducted among all persons on the OMERACT mailing list and experts in the field of QALY and research experience in rheumatology. The first part of the survey assessed opinions on attributes and suitability of the QALY approach. The second part evaluated individual familiarity with various instruments and priorities for research on these instruments.

Respondents. The survey was sent out to 600 persons and was (at least partially) completed by 92 respondents. Sixty-eight percent were male, 81% had an academic affiliation, 10% a pharmaceutical industry background, 72% performed clinical as well as research activities, 72% worked in North America or Europe, and 48% were not experienced in health economics.

Attributes of the QALY. Respondents were asked to score a list of 15 attributes about QALY according to relevance on a 5-point scale, where 1 = not relevant at all, 3 = neutral, 5 = very relevant. The list of attributes was based on aspects of the OMERACT filter of validity and on a published checklist specifically for judging preference-based measures of health-related quality of life (HRQOL)²². Respondents considered 4 attributes as most relevant (average on a 1–5 scale): ease of understanding (4.65); responsiveness (4.48); reproducibility (4.34); use of the patient's perspective (4.03).

The least important attributes (despite an average score > 3.4) were consistency with underlying (decision-making or economic) theory (3.77); ease of obtaining societal values (3.73); applicability towards other diseases (3.64); allowance for values worse than death (3.48).

Additional attributes raised by the respondents were free of charge for academic use; ease of understanding for decision-makers; terminology accessible to patients; applicability across legislations; and availability in different languages.

Suitability of approaches for a QALY estimate. Respondents were asked to score the suitability of 15 different approaches to estimate QALY on a 5-point scale (1 = not suitable at all, 3 = neutral, 5 = very suitable). The 3 most suitable approaches were (average on a 1–5 scale) transforming disease-specific measures (e.g., HAQ; 3.90); indirect measures (in general; 3.80); EQ-5D (3.71); SF-6D (3.65); and TTO (patients; 3.73).

Familiarity with instruments to assess utility and prioritization in QALY research. Of a list of the 9 utility instruments

used most frequently to calculate QALY, respondents were most familiar with VAS (93% of respondents), feeling thermometer (79%), and EQ-5D (71%), followed with similar scores for SF-6D (69%), TTO (69%), and SG (67%). There was no agreement on which of these should receive priority in further research. No instrument received the support ("yes") of at least 50% of the respondents.

Results of the survey. First, the working group discussed the bias in the results of the questionnaire in view of the low response and the overrepresentation of clinicians in contrast to persons experienced in QALY and economic evaluations in healthcare. The rheumatologists and rheumatology clinical researchers clearly favored the incorporation of patient values into QALY estimation approaches and were attracted by the role of the QALY in clinical decisions. Since it is known that patients value their health quite differently than society, it was thought reasonable to derive both patient and societal valuations into QALY estimates. It was noted that the EuroQol instrument contains both a population valuation of health (utility weights based on the 5 dimensions) and a patient rating (based on the thermometer). Although theoretical and clinimetric disadvantages of the thermometer or VAS are recognized, it might be possible to integrate the EQ-5D profiles and the VAS preferences into a patient's valuation of health. This could be compared with societal preferences alone. Also, similar approaches could be applied to instruments that contain health state descriptions.

Workgroup participants were surprised that "transforming disease-specific measures (HAQ) with preference weights" was a preferred approach for respondents of the survey. The HAQ reflects only a subset of dimensions relevant to HRQOL and, for example, does not include the influence of fatigues or sleep disturbances. Moreover, some felt that the relation between HAQ and QALY is substantially influenced by other population-specific confounders. The workgroup felt it was preferable to have a simple generic instrument that can be used to estimate the QALY across conditions.

The issue whether profiles within the health states should be different for the various rheumatological conditions remained unresolved. It was noted that the UK National Institute for Health and Clinical Excellence (NICE) recently decided to prioritize the EQ-5D for QALY estimates. The EQ-5D has recognized advantages, such as ease of applicability, availability of translations in many languages, and availability of country-specific weights. Moreover, participants noted an increasing familiarity among researchers and decision-makers with EQ-5D values for health. OMERACT might want to test a set of (recently proposed) tariffs that reflect the perspective of patients with rheumatological conditions and validate these with data from rheumatology clinical trials of biologic products submitted to NICE. It was also noted that the International Classification of Functioning, Disability, and Health might offer possibilities

to derive patients' values for health based on a universal classification.

WRAPUP AND CONSIDERATIONS FOR A FUTURE AGENDA (Michael Drummond)

Wrapup. QALY is an outcome in healthcare policy decisions and an accepted standard of valuing time spent with illness. Rheumatologists, health economists, and decision-makers are all concerned about the proliferation of approaches towards measuring QALY and the resulting difficulty comparing cost-effectiveness ratios across interventions. Nevertheless, the QALY has great potential. OMERACT can and should make an independent contribution to QALY research, alongside other international initiatives. For example, this could include further development of a reference case for deriving QALY in rheumatology.

Recommended next steps for the OMERACT QALY subgroup.

The focus of next steps includes:

1. Further development of a reference case for deriving QALY in rheumatology;
2. Collaboration with developers of the EQ-5D or other generic instruments to test methods to derive patient values and address the sensitivity of cost-effectiveness conclusions to the perspective of estimating utility values;
3. Greater insight into the role of descriptive systems when deriving values across different rheumatological conditions;
4. Review of global VAS and exploration of similarities and differences between the Global Patient VAS used in the American College of Rheumatology remission criteria ACR 20/50/70 and the EQ-5D VAS, to assess whether using one might suffice;
5. Exploration of the role that patient values can have in clinical choices; and
6. Further exploration of indirect methods for estimating QALY (i.e., using generic instruments such as EQ-5D and HUI3) in view of their ease of use, comparative performance across the range of health states in rheumatology, and the sensitivity of cost-effectiveness conclusions to the method of estimating utility values.

ACKNOWLEDGMENT

The authors acknowledge Leanne Idzerda for coordinating the e-mail survey and organizing the meeting, and all persons participating in the e-mail survey. The many personal comments were welcomed and useful.

Other Attendees at the QALY Subgroup meeting: Dorcas Beaton, Claire Bombardier, Peter Conway, Duncan Gordon, Jeffrey Hoch, Leanne Idzerda, Paul Kind, Lara Maxwell, Shirlina Prahardhi, J.L. (Hans) Severens, Erin Ueffing, and John Wong.

REFERENCES

1. Gabriel SE, Tugwell P, Drummond M. Progress towards an OMERACT-ILAR guideline for economic evaluations in rheumatology. Ann Rheum Dis 2002;61:370-3.
2. Gabriel S, Drummond M, Maetzel A, et al. OMERACT 6

- Economics Working Group report: a proposal for a reference case for economic evaluation in rheumatoid arthritis. *J Rheumatol* 2003;30:886-90.
3. Harrison MJ, Davies LM, Bansback NJ, Ingram M, Anis AH, Symmons DP. The validity and responsiveness of generic utility measures in rheumatoid arthritis: a review. *J Rheumatol* 2008;35:592-602.
 4. Marra CA, Marion SA, Guh DP, et al. Not all "quality-adjusted life years" are equal. *J Clin Epidemiol* 2007;60:616-24.
 5. Drummond MF, Sculpher MJ, Torrance GW. Methods for the economic evaluation of health care programmes. Oxford: Oxford Medical Publications; 2005.
 6. Raiffa H. Decision analyses: lectures on choices under uncertainty. Reading, MA: Addison-Wesley; 1968.
 7. Tsuchiya A, Williams A. Welfare economics and economic evaluation. In: Drummond MF, McGuire A, editors. Economic evaluation in health care: Merging theory with practice. Oxford: Oxford University; 2001:22-45.
 8. Johannesson M. Theory and methods of economic evaluation of health care. *Dev Health Econ Public Policy* 1996;4:1-245.
 9. Johannesson M, Jonsson B, Karlsson G. Outcome measurement in economic evaluation. *Health Econ* 1996;5:279-96.
 10. Gafni A, Birch S. Preferences for outcomes in economic evaluation: an economic approach to addressing economic problems. *Soc Sci Med* 1995;40:767-76.
 11. Torrance GW. Utility approach to measuring health-related quality of life. *J Chronic Dis* 1987;40:593-603.
 12. Torrance GW. Preferences for health outcomes and cost-utility analysis. *Am J Manag Care* 1997;3 Suppl:S8-20.
 13. Torrance GW. Utility measurement in healthcare: the things I never got to. *Pharmacoeconomics* 2006;24:1069-78.
 14. Torrance GW. Toward a utility theory foundation for health status index models. *Health Serv Res* 1976;11:349-69.
 15. The EuroQol Group. EuroQol — a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199-208.
 16. Horsman J, Furlong W, Feeny D, Torrance G. The Health Utilities Index (HUI): concepts, measurement properties and applications. *Health Qual Life Outcomes* 2003;1:54.
 17. Brazier J, Usherwood T, Harper R, Thomas K. Deriving a preference-based single index from the UK SF-36 Health Survey. *J Clin Epidemiol* 1998;51:1115-28.
 18. Bansback N, Marra C, Tsuchiya A, et al. Using the Health Assessment Questionnaire to estimate preference-based single indices in patients with rheumatoid arthritis. *Arthritis Rheum* 2007;57:963-71.
 19. Smith M, Drummond M, Brixner D. Moving the QALY forward: rationale for change. *Value Health* 2009;12:S1-S4.
 20. Atkins D, Best D, Briss PA, et al; GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328:1490-8.
 21. Brazier J, Deverill M. A checklist for judging preference-based measures of health related quality of life: learning from psychometrics. *Health Econ* 1999;8:41-51.
 22. Rosser RM. A health index and input measure. In: Walker SR, Rosser RM, editors. Quality of Life: Assessment and application. Lancaster: MTP Press; 1988.