Towards a Reference Case for Use in Future Economic Evaluations of Interventions in Osteoarthritis

INTRODUCTION

In 2002, Gabriel, *et al*[1] outlined a process for the development of reference cases[2] for economic evaluations in rheumatology. The reference case for economic evaluations of rheumatoid arthritis (RA)[3] was accompanied by an in-depth discussion of the methodological rationale and the strengths and weaknesses of the supporting knowledge. The objective of this article is to provide a similar discussion of the key methodological approaches that might be suggested for a reference case in economic evaluations of osteoarthritis (OA). Our article is structured in the following manner: (1) 3 recent economic evaluations in the field of OA are described; (2) 13 methodological issues are discussed by using illustrations from the 3 studies, with additional examples, where relevant; and (3) some of the key unresolved issues are highlighted in order to determine the priority areas for further discussion.

RECENT EXAMPLES OF ECONOMIC EVALUATIONS IN OA

Economic Evaluation of Acetaminophen, NSAID, and Selective COX-2 Inhibitors in the Treatment of Symptomatic Knee OA

In our study a decision-analytic model was used to compare symptomatic therapy with analgesics over a 6-month period[4]. Two measures of effectiveness were used: (1) number of upper gastrointestinal (UGI) adverse events averted; and (2) number of patients who achieved perceptible pain relief. Separate analyses were conducted for all patients and for those who did not respond to acetaminophen.

The evaluation showed that acetaminophen dominates the other therapies in terms of cost per UGI event averted. Acetaminophen was also shown to be the preferred therapy if one values pain relief below US$ 275 per patient achieving minimal perceptible clinical improvement (MPCI), followed by ibuprofen US$ 275–$14,150) and rofecoxib (> $14,150). Rofecoxib was universally the preferred therapy for acetaminophen nonresponders at high risk of developing an adverse UGI event.

Economic Evaluation of Incorporating Hylan G-F 20 into the Treatment of Patients with Knee OA

The cost-effectiveness of injecting knee joints with hylan G-F 20 (a high-molecular weight viscosupplementation product) in addition to normal care as compared to normal care alone was assessed alongside a randomized clinical trial[5]. OA-related costs were collected from the societal viewpoint, clinical improvements were assessed with the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)[6] and the Health Utilities Index Mark 3 (HUI3)[7]. Hylan G-F 20 was associated with higher costs, but greater increases in quality-adjusted life-years (QALY). The incremental cost-effectiveness ratio of CDN$ 10,000 per QALY gained was thought to be below the threshold for adoption of health technologies typically applied in Canada.

The Cost-Effectiveness of Celecoxib and Rofecoxib in Patients with OA or RA

The objective of our study was to evaluate the cost-effectiveness of celecoxib in comparison to diclofenac and ibuprofen, and rofecoxib in comparison to naproxen, in patients with OA and RA not taking low-dose aspirin for the prevention of cardiovascular disease. Analyses were performed for patients at average risk of UGI events, and for higher-risk patients with a history of a UGI event. A Markov decision-analysis model extrapolated data from large clinical trials over a 5-year timeframe. Utilities for clinical events were obtained from the general public. For average-risk patients the incremental cost-effectiveness ratios for both cyclooxygenase-2 (COX-2) inhibitor nonsteroidal anti-inflammatory drugs (NSAID) exceeded CDN$ 200,000 per QALY, when compared with standard NSAID. For high-risk patients, the base-case results showed the COX-2 NSAID to be more effective and less costly than standard NSAID with co-prescription of proton pump inhibitors (PPI). This cost-effectiveness advantage was reduced as the rate of co-prescription of PPI with COX-2 was allowed to increase in the model.

KEY METHODOLOGICAL ISSUES

Model Horizon (Issue 1)

The generic reference case developed by Gold, *et al*[2] stated that the study (or model) horizon in an economic evaluation should be long enough to capture all the significant costs and benefits due to the therapy. The ideal study horizon would thus be the patient's lifetime. In practice, though, the longterm consequences of an event that occurs during the study period, such as a non-fatal myocardial infarction, are imputed, often over the remaining life-years of a patient. While the costs and consequences of symptom-modifying treatments such as NSAID might be adequately assessed within a short, 1-year time-horizon, disease-modifying agents would need to be evaluated over longer time horizons, such as 5 or 10 years.

How has this been addressed in practice? The study by Kamath, *et al*[4] used only a 6-month time horizon. The authors address this point and argue that their modelling study could have been extended beyond 6 months, possibly by using a Markov model, but that no longterm data were

available to inform such a model. In their study of selective COX-2 inhibitors, Maetzel, *et al*[8] used a Markov model with a 5-year timeframe but did not model potential reduction in mortality over the lifetime of the patient. The study by Torrance, *et al*[5] used a one-year time horizon, since this was the followup period of the relevant clinical trial. However, a longer time horizon might be necessary if hylan G-F 20 falls within the class of disease-modifying agents.

Duration of Therapy (Issue 2)

One of the key determinants of whether costs and consequences will continue into the future is the type of therapy. For instance, surgery for hip or knee replacement (including post-surgical followup and rehabilitation) might be seen as a one-time event. However, even here it is known that surgical revisions are sometimes required. Therefore the costs and consequences of these would appropriately be attributed to the initial therapy in an economic evaluation.

In the case of the use of NSAID for OA, therapy is often of a short duration, continuous or intermittent. Thus, where the benefit does not decline over time in patients who continue the drug, a short term analysis may suffice, since we can assume that the costs and outcomes observed during the study period (e.g., a one-year trial) would merely be replicated in future periods (i.e., the incremental cost per QALY estimate would be stable over time). However, new technologies, such as hylan G-F 20, might require longterm data.

How has this been addressed in practice? The studies by Kamath, *et al*[4] and Maetzel, *et al*[8] compared particular daily doses of the drugs of interest. The authors do not explicitly state their approach to dealing with compliance and discontinuation of therapy. However, this is likely to include discontinuation of NSAID therapy for those patients experiencing GI adverse events.

The study by Torrance, *et al*[5] considered the initial course of hylan G-F 20 injections, but also allowed for the possibility of additional treatment with hylan if required over the period of the clinical trial. The authors did not address the question of whether therapy could continue beyond the period of the clinical trial.

Extrapolation Beyond Trial Duration and Modelling Beyond Trial Duration (Issues 3 and 4)

If the desired study horizon for the economic evaluation is several years, or the patient's lifetime, several elements of extrapolation are required, one of which is the consideration of longterm therapy duration. Another important element is the extrapolation of treatment effect while the patient continues therapy, and the modelling of any change (i.e., decline) in treatment effect should the patient discontinue therapy.

Similarly, in addressing the question of what happens when therapy is discontinued, one might assume that the treatment effect (1) ceases (simplest), (2) declines over time, or (3) continues (optimistically). A very conservative position would be to assume a "catch-up" effect, where the decline in the patient's condition after discontinuation is worse than that for someone who had not been receiving therapy in the first place.

How has this been addressed in practice? The studies by Kamath, *et al*[4] and Torrance, *et al*[5] do not address the question of what happens beyond the period studied. Therefore their interpretation of the data is that they can tell us nothing about what might happen in the future. However, a possible assumption (not made by the authors) is that the continuation of therapy for another similar period of time would produce the same costs and the same effects. In essence, this is the tacit assumption being made where incremental cost-effectiveness ratios are being compared from studies with different time horizons. In the study by Maetzel, *et al*[8], it was assumed that the effect of NSAID (in terms of both efficacy and adverse events) was maintained as long as the patient continued therapy.

Synthesis of Comparisons Where Head-to-Head Trials Do Not Exist (Issue 5)

Indirect comparisons of drugs in the absence of head-to-head studies are fraught with difficulties, owing to the possibilities that the trials enroll different categories of patients, have different protocols (e.g., regarding the use of gastroprotective agents), and measure outcomes differently. Because of the high risk of bias in indirect comparisons, Gabriel, *et al*[1] advised against them. However, some analysts have proposed statistical approaches for dealing with the problem of indirect comparison. For example, model parameters can be estimated simultaneously using Bayesian Markov chain Monte Carlo methods[9]. The pros and cons of these approaches, as compared with arguing for head-to-head studies in all cases, need to be explored. Metaanalysis of randomized controlled trials (RCT) with common comparator adjusting for baseline covariates might be another alternative, and estimates of adverse events could be obtained by using all available single arms of RCT or cohort/observational studies of each intervention being compared, with adjustment for baseline covariates.

How has this been addressed in practice? In the study by Torrance, *et al*[5], this issue was not a problem, in that hylan G-F 20 was added to appropriate care and compared with appropriate care alone. Therefore, for the economic evaluation to be relevant we only need to be reassured that "appropriate care" as delivered in the trial is representative of regular practice.

In the studies by Kamath, *et al*[4] and Maetzel, *et al*, the authors undertook extensive reviews of the literature, including clinical trials and observational studies. They made judgments on the quality and appropriateness of the data from particular studies for inclusion in the model and

used extensive sensitivity analyses to deal with the uncertainties in the data. The study by Maetzel, *et al* also avoided indirect comparisons by evaluating celecoxib and rofecoxib separately with the data from the supporting clinical trials. No comparisons were made between the two COX-2 agents.

Outcome Measures (Issue 6)

In OA, outcomes are multifaceted and include pain, mobility, and activities of daily living. It was previously suggested to use criteria proposed by the Osteoarthritis Research Society International (OARSI) for symptom-modifying treatments[10], which define response as: (1) A reduction of 50% from baseline and an absolute reduction of 20 either in OA pain or in WOMAC LK 3.1 (DPDA subscale score, re-standardized on a scale of 0–100), or if there is: (2) a reduction of 20% from baseline and an absolute reduction of 10 in at least 2 of the 3 variables: overall OA pain, the WOMAC LK 3.1 (DPDA subscale score re-standardized on a scale of 0–100), or the patient's global assessment of disease activity.

Various disease-specific and generic quality of life measures have also been considered to be relevant in OA; these include the WOMAC and the Medical Outcome Study Short-Form 36 (SF-36). For disease-modifying drugs valid outcome measures such as imaging and biochemical markers that reflect future total joint replacement averted are required.

How has this been addressed in practice? The study by Torrance, *et al*[5] used both the WOMAC and the SF-36. The study by Kamath, *et al*[4] used a measure of effectiveness defined as the number of patients who achieved the MPCI on the WOMAC pain subscale. A previous study had determined that the MPCI was equivalent to a change of 9.7 units on a 0–100 visual analog scale administered as part of the WOMAC instrument. The majority of economic evaluations of NSAID (including those of COX-2 inhibitors) differentiate the drugs not in terms of efficacy outcomes, but in terms of their GI adverse events.

Mortality (Issue 7)

Mortality data are important in economic evaluations because the most common measures of benefit are life-years or QALY. Therapies for OA do not increase survival, so mortality data only enter into lifetime economic evaluations in order to model death from other causes, or to reflect mortality for adverse events. In a study of a surgical intervention for OA, the mortality risk from the operation would need to be considered. In a study of NSAID one would expect some mortality risk from serious GI complications.

How has this been addressed in practice? Neither of the studies by Kamath, *et al*[4] or Torrance, *et al*[5] consider mortality, owing to the short time horizons used. In the study by Maetzel, *et al*[8], which used a 5-year time horizon, complicated UGI events were assigned a probability of death in the calculation of QALY. In an evaluation of NSAID with different rates of GI complications there will therefore be different mortality rates for the 2 patient groups.

Valuation of Health (e.g., QALY) (Issue 8)

In line with the recommendations of other regulatory agencies[11], the reference case for RA recommends that values from the general public should be used for public policy decisions and values from patients for clinical decisions[10]. This decision facilitates the measurement of health state preference values in prospective studies with the HUI or the EuroQoL (EQ-5D)[2]. These measures involve the administration of a questionnaire during the study, in order to classify the patients' health states, with the health state "tariff" values being generated from previously-conducted surveys of the general public.

How has this issue been addressed in practice? Torrance, *et al*[5] used the HUI3 instrument to calculate the QALY gained from hylan G-F 20 therapy. Maetzel, *et al*[8] undertook a community survey in order to estimate utility values for their study. However, in general there seems to be no reason why the OMERACT recommendation for RA[1] should not be adopted for OA.

Resource Utilization (Issue 9)

Similar to other official guidelines[12], it is recommended that the reference case for OA should incorporate direct medical costs in the analysis, but report indirect and nonmedical costs separately. The main estimation problem is likely to relate to the downstream events, such as the costs of surgical revisions, or the costs of managing GI complications. Since some of these events are rare in clinical trials, it is usually better to estimate the resource utilization from observational studies or routine data sources, such as administrative databases.

How has this been handled in practice? Since the study by Torrance, *et al*[5] was conducted alongside a clinical trial, the authors were able to track actual resource utilization in detail and measured time off work and patients' expenditures to accommodate the perspective of the health care system and society. Kamath, *et al*[4] used institutional billing data (over a 5-year period) to estimate the cost of hospitalization and outpatient management of confirmed and suspected PUB. Maetzel, *et al* used largely routinely available sources to estimate the costs of managing clinical events.

Classification and Reporting of Adverse Events (Issue 10)

Toxicity is an important consideration in arthritis drug therapy and common toxicity criteria are under development by the OMERACT Toxicity Working Group[1]. It remains to be seen whether this classification will be useful for

economic evaluations. With new disease-modifying agents for the treatment of OA in development, the range of possible adverse events may be expanded.

How has this issue been addressed in practice? In the studies by Kamath, *et al*[4] and Maetzel, *et al*[8], a classification of GI adverse events was developed as a basis for the economic evaluation. In the study by Torrance, *et al*[5], the consequences of any adverse events were reflected in the overall utility scores and cost estimations for the 2 treatment groups. The quantity and nature of any adverse events were reported in the associated clinical evaluation (Raynauld, *et al*[13]).

Discontinuation of Therapy (Issue 11)

This issue has been largely addressed above. Discontinuation of therapy is likely to be an important consideration in the case of NSAID, where patients might suffer GI adverse events. Some data on discontinuation can be obtained from the clinical trials, but longterm data can only be obtained from observational studies.

How has this issue been addressed in practice? Maetzel, *et al*[8] used data from observational studies to estimate discontinuation rates beyond the duration of the clinical trials. However, they did not make direct comparisons between the two COX-2 agents. More longterm data are required on the compliance with different treatments. This is important because lack of compliance, or discontinuation, can affect both efficacy and costs. Clearly, in the case of NSAID, where patients take drugs intermittently over long periods, such studies will be difficult to mount.

Therapeutic Strategies (Issue 12)

While most clinical trials and economic evaluations compare defined treatment regimens (e.g., named drugs), physicians in regular practice often try one therapy and switch to another, if there is a lack of efficacy or in the presence of adverse events. Sometimes treatment strategies, as opposed to individual therapies, are specified in clinical practice guidelines (e.g., stepped care regimens).

For studies of existing NSAID, and possibly the COX-2 inhibitors, treatment strategies could be modelled if there were data on discontinuation rates with the various drugs. With the advent of disease-modifying agents, treatment strategies in OA may become more similar to those in RA, where there are questions about the appropriate sequence of different classes of drugs. In some clinical fields, such as depression, where discontinuation and treatment switches are common, longterm pragmatic clinical trials have been undertaken with incorporation of an economic component[14].

How has this issue been addressed in practice? The studies by Kamath, *et al*[4] and Torrance, *et al*[5] compared defined treatment alternatives rather than treatment strategies. However, Kamath, *et al* did consider that additional treatment would be required for patients who were acetaminophen nonresponders.

Population Risk Stratification (Issue 13)

In medical conditions where there is a range of treatment options, it is unlikely that the same therapy will be the most cost-effective for all patient groups. Indeed, in jurisdictions where economic evidence is required as part of the development of guidance for the use of health technologies, an explicit attempt is made to identify those subsets of the patient population for whom the technology would be most cost-effective[11].

In the evaluation of NSAID an important stratification is in relation to the patient's risk of GI adverse events. For example, in its technology appraisal of COX-2 inhibitors, the National Institute for Clinical Excellence in the UK considered patients were high-risk if they met certain criteria, such as age > 65 years, prior history of serious GI complications, use of concomitant medications known to increase the likelihood of upper GI adverse events, serious comorbidity, or requiring the prolonged use of maximum recommended doses of standard NSAID[15].

How has this issue been addressed in practice? The study by Torrance, *et al*[5] only applies, by definition, to the patients enrolled in the clinical trial of hylan G-F 20. The economic analysis applied to the patient population as a whole, and no subgroup analyses were reported. In the study by Kamath, *et al*[4], the patient population was restricted to those with symptomatic knee OA. In addition, a subgroup analysis was conducted for patients who failed to respond to acetaminophen. In the study by Maetzel, *et al*[8], a distinction was made between average-risk and high-risk patients. The latter were defined as those who had experienced prior complicated UGI events (e.g., perforation, obstruction, or major bleeding), or prior clinical symptomatic ulcers, as shown by endoscopy. Further, the analysis allowed for modelling of 2 additional risk factors, age, and a potential third risk factor, e.g., use of corticosteroids. Results were provided for various combinations of risk factor levels.

TOWARDS A REFERENCE CASE FOR ECONOMIC EVALUATIONS IN OA

It is clear from the discussion above that some of the 13 methodological issues raised by Gabriel, *et al*[1] are either easily resolved, or should be resolved by following the recommendations made in the reference case for RA[3]. However, several issues are more intractable and merit further discussion before a reference case for OA can be defined. These are:

1. Synthesis of comparisons where head-to-head trials do not exist
2. Duration of therapy/discontinuation of therapy/therapeutic strategies

3. Extrapolation beyond trial duration/modelling beyond trial duration
4. Outcome measures

Our article has identified a number of the key methodological issues in the evaluation of therapies for OA and indicated, where possible, options for their resolution. The next step would be to reach a consensus on the main issues. In the meantime, our aim is to stimulate discussion and debate.

MICHAEL DRUMMOND, DPhil,
Centre for Health Economics,
University of York, York, UK;
ANDREAS MAETZEL, MD, PhD,
Division of Clinical Decision Making,
University Health Network, and
Department of Health Policy Management and Evaluation,
University of Toronto, Toronto, Canada;
SHERINE GABRIEL, MD, MSc,
Department of Health Services Research and
Division of Rheumatology, Department of Internal Medicine,
Mayo Clinic, Rochester, Minnesota, USA;
LYN MARCH, MD, PhD,
University of Sydney, Institute of Bone and Joint Research, and
Department of Rheumatology, Royal North Short Hospital,
St. Leonards, NSW, Australia.

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. Gold MR, Siegel JC, Russell LB, Weinstein MC. Cost-effectiveness analysis in health and medicine. New York: Oxford University Press; 1986.
3. Maetzel A, Tugwell P, Boers M, et al. Economic evaluation of programs or interventions in the management of rheumatoid arthritis: defining a consensus-based reference case. J Rheumatol 2003;30:891-6.
4. Kamath CC, Kremers HM, Vanness DJ, O'Fallon WM, Cabanela RL, Gabriel SE. The cost-effectiveness of acetaminophen, NSAIDs, and selective COX-2 inhibitors in the treatment of symptomatic knee osteoarthritis. Value Health 2003;6:144-57.
5. Torrance GW, Raynauld JP, Walter V, et al. A prospective, randomized, pragmatic, health outcomes trial evaluating the incorporation of hylan G-F 20 into the treatment paradigm for patients with knee osteoarthritis: economic results. Osteoarthritis Cartilage 2002;10:518-27.
6. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833-40.
7. Boyle MH, Furlong W, Feeny D, Torrance GW, Hatcher J. Reliability of the Health Utilities Index — Mark III used in the 1991 cycle 6 Canadian General Social Survey Health Questionnaire. Qual Life Res 1995;4:249-57.
8. Maetzel A, Krahn M, Naglie G. The cost-effectiveness of rofecoxib and celecoxib in patients with osteoarthritis or rheumatoid arthritis. Arthritis Rheum 2003;49:283-92.
9. Ades AE, Cliffe S. Markov chain Monte Carlo estimation of a multiparameter decision model: consistency of evidence and the accurate assessment of uncertainty. Med Decis Making 2002;22:359-71.
10. 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.
11. National Institute for Clinical Excellence. Technical guidance for manufacturers and sponsors. Technical Appraisals Process Series No. 5. London: NICE; 2000.
12. Hjelmgren J, Berggren F, Andersson F. Health economic guidelines — similarities, differences and some implications. Value Health 2001;4:225-50.
13. Raynauld J-P, Torrance GW, Band PA, et al. A prospective, randomized, pragmatic, health outcomes trial evaluating incorporation of hylan G-F 20 into the treatment paradigm for patients with knee osteoarthritis: clinical results. Osteoarthritis Cartilage 2002;10:506-17.
14. Simon GE, Vonkorff M, Heiligenstein JH, et al. Initial antidepressant choice in primary care: effectiveness and cost of fluoxetine vs tricyclic antidepressants. JAMA 1996;275:1897-902.
15. National Institute for Clinical Excellence. Cyclo-oxygenase (COX) II selective inhibitors, celecoxib, rofecoxib, meloxicam and etodolac for osteoarthritis and rheumatoid arthritis. Technology appraisal No. 27, August 2001. London: NICE; 2001. Also available at: http://www.nice.org.uk/cat.asp?c=18033 [Cited August 14, 2003]