Standardized Assessment of Adverse Events in Rheumatology Clinical Trials: Summary of the OMERACT 7 Drug Safety Module Update

MARISSA N.D. LASSERE, KENT R. JOHNSON, MAARTEN BOERS, KERRI CARLTON, RICHARD O. DAY, MARTEN de WIT, I. RALPH EDWARDS, JAMES F. FRIES, DANIEL E. FURST, JOHN R. KIRWAN, PETER S. TUGWELL, THASIA G. WOODWORTH, and PETER M. BROOKS

ABSTRACT. A presentation, demonstration, and discussion of recently developed adverse event instruments were the topics for the OMERACT 7 Drug Safety Module. The module began with a plenary introducing the needs and challenges of adverse event ascertainment. It was followed by a review of module work from previous OMERACT meetings on a prototype coding instrument (Rheumatology Common Toxicity Criteria), then a brief description of the process behind the recently developed patient self-report and investigator report adverse event instruments. These current instruments are designed for use in controlled trials although they could be used in other settings. The instruments rely primarily on patient self-reporting using a checklist, which the investigator then folds into a parallel structured but more medically sophisticated instrument. In pilot testing, this innovative dual-use system has shown reliability and acceptability, while preserving validity. A "stakeholder panel" of representatives from 8 sectors followed — patient, nurse investigator, regulator, clinician scientist, industry, OMERACT, global public health/WHO, and Cochrane Collaboration — for their perspectives on the needs, challenges, and potential ways forward for adverse event ascertainment and reporting in clinical trials. At the breakout session small focus groups participated in hands-on interactive testing of one of 3 versions of the instruments, which differ in degree of comprehensiveness. Each focus group had a participatory patient with rheumatoid arthritis. At a second plenary there was group feedback by rapporteurs and presentation of results from pilot studies of iterative testing of validity, reliability, and feasibility of the instruments. During plenary discussion a frequent suggestion for improvement was to refine the process so that event ascertainment could be done entirely using the patient instrument with minimal input from the investigator at the visit, if patient-investigator agreement was high. Most found the patient checklist attractive, particularly if the patient instrument was shown to be reliable and valid. Finally, a future research agenda was discussed. (J Rheumatol 2005;32:2037–41)

Key Indexing Terms:
ADVERSE EVENTS MEASUREMENT
INSTRUMENT VALIDATION
RANDOMIZED CLINICAL TRIALS
DRUG TOLERABILITY INDEX

Introduction

The first article in the OMERACT 7 Drug Safety Module Update reviews inadequacies of the current approach to adverse event assessment and emphasizes the urgent need for a standardized measurement tool with appropriate properties and performance[1]. It also summarizes work that was

From the University of New South Wales, Department of Rheumatology, St. George Hospital, Sydney; Department of Clinical Pharmacology, Mater Hospital, University of Newcastle, Newcastle, Australia; Department of Clinical Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, The Netherlands; World Health Organisation Collaborative Centre for International Drug Monitoring, Uppsala, Sweden; Stanford University School of Medicine, Palo Alto; Geffen School of Medicine at the University of California in Los Angeles (UCLA), Los Angeles, California, USA; University of Bristol Academic Rheumatology Unit, Bristol, United Kingdom; Centre for Global Health, University of Ottawa, Institute of Population Health, Ottawa, Ontario, Canada; ABGH Novartis Pharma, Basel, Switzerland; and Faculty of Health Sciences, University of Queensland, Royal Brisbane Hospital, Herston, Australia.

M.N.D. Lassere, MB, BS, Grad Dip Epi, PhD, FRACP, FAFPHM, Associate Professor in Medicine, Conjoint Appointee, University of New South Wales, Department of Rheumatology, St. George Hospital; K.R. Johnson, MD, Senior Research Officer, Department of Clinical Pharmacology, Mater Hospital, University of Newcastle; M. Boers, MSc, MD, PhD, Professor of Clinical Epidemiology, Department of Clinical Epidemiology and Biostatistics, VU University Medical Center; K. Carlton, BSc, MSc, Research Officer, St. George Hospital; R.O. Day, MD, FRACP, Professor of Clinical Pharmacology, St. Vincent Hospital, University of New South Wales; M. de Wit, Dutch Arthritis Patients' League, The Netherlands; I.R. Edwards, MB, ChB, FRCP; Professor, Director, World Health Organisation Collaborative Centre for International Drug Monitoring; J.F. Fries, Professor of Medicine, Stanford University School of Medicine; D.E. Furst, Carl M. Pearson Professor of Rheumatology, Geffen School of Medicine at UCLA; J.R. Kirwan, BSc, MD, University of Bristol Academic Rheumatology Unit; P.S. Tugwell, MD, MSc, Director, Centre for Global Health, University of Ottawa, Institute of Population Health; T.G. Woodworth, MD, Executive Director, Global Therapeutic Area Profiling Head, ABGH Novartis Pharma; P.M. Brooks, Executive Dean, Faculty of Health Sciences, University of Queensland, Royal Brisbane Hospital.

Address reprint requests to Dr. M.N.D. Lassere, Department of Rheumatology, St. George Hospital, Belgrave Street, Kogarah, NSW 2217, Australia. E-mail: lasserem@sesahs.nsw.gov.au

undertaken towards developing such an instrument and that resulted in the Rheumatology Common Toxicity Criteria (RCTC)[2]. Since then, substantial further work has been undertaken in conceptualizing and then realizing ideas that emerged at OMERACT: (1) The goal was changed to developing an instrument that enabled investigator recording of an adverse event directly onto the case report form at time of visit, rather than coding the recorded adverse events at a later time. (2) The target was expanded from rheumatology to all fields of medicine and from rheumatology drug treatments to all therapeutic interventions. (3) The instrument was crafted for use in randomized trials where patients are seen about every month. (4) The instrument was developed to capture information derived from patient self-reporting. This led to the concept of 2 instruments, one for the patient and one for the investigator, constructed in parallel to allow for complementary use at the time of the visit. These results are reported elsewhere in this module[3].

In the Drug Safety Module Update of OMERACT 7 the development of these instruments was discussed and, as an important part of their continuing development, the instruments themselves were assessed for working feasibility in breakout sessions, which simulated their use in clinical encounters.

The Drug Safety Module Update — Overview and Introductory Session

The OMERACT 7 Drug Safety Module Update had 5 objectives: to review the background and conceptual basis for an adverse event index; to present the lineage of the instrument development and various iterations in its performance testing in relation to the OMERACT filter[4]; to explore the need for an adverse event instrument from the viewpoint of various stakeholders; to engage participants in breakout groups in a "hands-on" experience using the adverse event instrument with patients, to elicit critical feedback at the plenary session; and to conclude by discussing an agenda for future work.

The module leader briefly reviewed the background and conceptual arguments concerning assessment methods for acquisition of drug safety and tolerability[1,3]. Eight stakeholders each made one minute presentations addressing 3 questions from the perspective of their stakeholder constituency: Where are we now? Where do we want to be? What do we need to get there?

The Drug Safety Module Update — Stakeholder Panel Presentations

Patient. M. de Witt described the many uncertainties that patients constantly face and how these are as much of a challenge as the arthritis symptoms. One uncertainty occurs whenever a new treatment is started. The patient hopes that the treatment will help their condition, but also worries that it carries a risk of adverse effects. But how well are the adverse effects known? How will the patient know whether a new complaint or worsening of an existing complaint is due to the condition being treated, some other condition, or an adverse effect from treatment? To minimize this uncertainty and associated worry would be of immense benefit.

Research nurse. K. Carlton reviewed her experience of current *ad hoc* approaches to acquisition of adverse event data in clinical trials. The process is conducted with a case report form for recording events elicited in an open-ended fashion and may include an assessment of severity, attribution, resulting action, and outcome. She emphasized the highly variable nature and low reliability of the present approach and strongly encouraged a system designed to enhance the accuracy of its content and severity grading. The approach should be more prescriptive and provide better guidance to lessen the need for arbitrary and discretionary decisions both by the nurse investigator and by the patient.

Regulator. K. Johnson, formerly at the US Food and Drug Administration, argued that formal assessment of safety and tolerability lags far behind that for efficacy. There are historical and conceptual reasons for this, but the obstacles to a remedy are not insurmountable. In fact, regulatory agencies are increasingly engaged with pharmaceutical companies that develop products explicitly to be safer than existing marketed drugs — so-called safety claims. Enhanced ability to assess adverse events would mean better outcomes for patients and more informative marketing; more broadly, it would enable better estimates of risk/benefit and more valid cross-trial and cross-drug comparisons. Neither scientific limitations in trial design and statistics nor legal precedents in statutory law will impede rectifying the imbalance between safety/tolerability and efficacy. The biggest impediment is the lack of an adverse event assessment instrument with appropriate properties and performance characteristics.

Clinician scientist. J. Fries provided insight from his considerable experience of instrument development in the area of drug safety. He noted that, in the process of building an adverse event index, it is necessary to reconcile side effect profiles of different drugs. Thus compromise will be necessary to achieve consensus because indices are unavoidably arbitrary. He also argued that weighting of the importance of different adverse effects be driven by patient values rather than investigator values. His experience of instrument development suggested that to reliably assign drug attribution, an instrument needs a "causal filter." In his observational databases a simple symptom menu without such a filter was unable to separate the signal of true adverse events from the noise of background. Adverse events distinguishable with a "filter" instrument but not with a simple symptom menu included aspirin-related tinnitus and indomethacin-related headache. However, the setting of an observational database differs from that of a randomized controlled trial, and it is important to continue to explore the possibilities of the observational database. He asked and

answered some important questions: Can an ideal toxicity index be created?: No. Can a useful toxicity index be created?: Yes. Is it important?: Yes. Is it easy?: No.

Pharmaceutical scientist. T. Woodworth has participated for an extended period in safety assessment; she was an architect of the Rheumatology Common Toxicity Criteria. She noted that use of the RCTC in industry-sponsored trials was aimed to provide consistency of description and severity of adverse events and allow comparison of side effect profiles across different populations. Clearly, methods to validate the RCTC need to be developed. Does the RCTC instrument succeed in ensuring consistency of safety evaluation in trials and so enable comparison across different patient groups and treatments? Broader use of the RCTC requires better publicity and more forthright engagement of regulators regarding these conceptual approaches to adverse assessment generally.

The OMERACT perspective. M. Boers described the OMERACT filter as the triad of truth, discrimination, and feasibility[4]. This definition is a simple yet complete paradigm that can be applied to judge any new measure in rheumatology, including any adverse event instrument. The process of "meeting the filter" is iterative and informs the development strategy for a new measurement instrument. The proceedings of this module update are an important part of that process.

The Cochrane perspective. P. Tugwell presented the viewpoint of a section leader in the Cochrane Collaboration (www.cochrane.org). To date the Cochrane Library includes some 80 reviews in the musculoskeletal area, all easily web-accessed. The group is now expanding its conceptual base to "balance sheet" presentations, which attempt to capture and measure both the benefits and the risks of treatments. This reinforces the rationale for development of a standardized adverse event index that meets the OMERACT filter.

The global public health/World Health Organization perspective. R. Edwards discussed adverse event assessment from the perspective of global health through the offices of the WHO monitoring center in Sweden. His unit has had extensive experience in automated signal detection and pattern recognition to clarify issues of attribution. There are many important operational issues and technical challenges in data collection and analysis.

The Drug Safety Module Update — Breakout Group Hands-on Exercises

Several versions of paired patient and investigator adverse event instruments were brought forward to this session. (1) The RCTC instrument was paired with a patient self-report form devised specifically to complement the RCTC (Version 1). (2) An early (and relatively short) version of the new patient and investigator instruments was used (Version 2). (3) The most recently developed (and relatively long) version of the new patient and investigator instruments was included (Version 3). These 3 pairs of instruments had been developed sequentially in order to increase internal validity and reliability. However, with increasing accuracy and comprehensiveness of each version came increasing length. This is a common tradeoff between validity/reliability and feasibility. Table 1 shows the essential differences in the 3 versions.

Twelve breakout groups were held, with versions 1, 2, and 3 of the instruments each being tested by 4 groups. Patient participants played a particularly important role in the groups, as one patient in each group had volunteered to complete the patient self-report version of the questionnaire before the module update session was convened. The patient used recent personal experience to record adverse events. These patient participant volunteers were from Sweden, Denmark, Norway, The Netherlands, England, USA, Canada, and Australia; and for many English was a second language. Patients were asked to read and complete the questionnaire to the best of their ability. Group leaders received copies of the patient and investigator instruments for use in their group prior to the module update session; other participants received them at the session.

Group leaders and rapporteurs were given a copy of their respective instrument (both investigator report and patient report) prior to the breakout sessions. The breakout groups met in separate rooms for about 30 minutes, during which time the leader, now with the completed patient report-form available for reference, interviewed the patient in order to complete the investigator form. Copies of the completed patient questionnaire were also available to other group members, who completed their own copy of the investigator form as the interview proceeded. Group members were able to add questions or seek clarification from the patient interviewee or from the group leader. A rapporteur in each group recorded the process and noted points raised by other group participants during the interview, then reported back at the plenary session that followed.

The Drug Safety Module Update — Plenary Review of Instrument Performance

Patient participants reported having no difficulty filling out the patient report component of the instrument. However, in the process, patients became aware of many limitations. For example, certain deficiencies in content were noted, particularly in version 1, including the absence of some common complaints. The wording of some sections in versions 1 and 2 was also sometimes difficult to follow. These comments reinforced previous findings during the development of the 3 versions of the instruments[3]. The most recent version (Version 3) performed well in this regard.

Several themes emerged regarding use of the investigator instrument. First was the issue of the time: it took considerable time to collect information using the investigator

Table 1. The 3 versions of the patient self-report and investigator report adverse event instruments used by the "breakout" groups.

Instrument	No. of Pages	No. of Categories (body systems)	No. of Items (symptoms)	No. of Severity Grades
Version 1				
Patient-report	4	11	72	3
Investigator-report	6	11	72	4
Version 2				
Patient-report	6	15	126	3
Investigator-report	15	15	178	4
Version 3				
Patient-report	7	17	152	3
Investigator-report	*	17	*	7

* To be determined; instrument not yet complete.

instruments, particularly as they were unfamiliar to users. Breakout group members were concerned that there might not be a sufficiently rapid learning curve to increase speed of instrument administration given the time constraints of a clinical trial visit. However, it was recognized that the instrument would probably be administered by nurse investigators, who would develop their own expertise. First, because it is necessary to code all events at the time of the visit, more time was required than for simple recording. Second, there were concerns about the coding of severity. Coding was helped by provision of severity grades within the questionnaire format, but the difficulties of defining severity became apparent. Participants struggled with combining duration, frequency, intensity, and impact into an assessment of severity. Third, as a solution, it was suggested to convert the questionnaire to an electronic version in which large sections of questions would not be displayed unless a key question about adverse effects in a given body system was positive; the detailed questions about that system would then be answered.

One approach that had widespread support from participants was to complete the entire process using only or mainly the patient instrument. If there is good agreement between patient and investigator, then minimal input from the investigator will be required. Most (but not all) participants found this patient checklist approach attractive, particularly if the patient instrument is shown to be reliable and valid. Support for a fully patient-based instrument emerged unexpectedly, and was undoubtedly related to the hands-on experience in the breakout discussion groups. This approach may offer the optimal compromise between validity and feasibility. In this regard one group proposed a more patient-friendly system of categorizing symptoms than the medical body systems approach, but it is unclear what this approach might be. Another group preferred the goal of recording serious adverse events, rather than all adverse events. Another noted that severity grading would differ between investigator and patient. More fundamentally, a fully patient-based instrument would limit the validity of some elements of adverse event ascertainment (such as assessment of adverse laboratory values), as patients would be unaware of them or their implications. Perhaps a patient instrument could be used in conjunction with traditional methods.

Another unresolved concern raised in the plenary discussion was the question of drug attribution (that is, whether the event was caused by the drug). There are strong arguments for reliance on patient perceptions of events and their weightings of impact and importance; similarly, patient perceptions of attribution of adverse events could be used, since patients have actually experienced the events. Many participants, especially patients, found it troubling that in the questionnaires no effort was directed at obtaining drug attribution information. Some patients felt they were perfectly able to separate drug causation from events related to their disease or to other comorbidity. We were reminded that Fries and colleagues had found it necessary to use a "causal filter" for an adverse effect reporting instrument to have credible discriminatory power. However, that work had been conducted in the observational setting, not the randomized controlled clinical trial setting where the information dynamics are different because one has a controlled comparator.

Many participants noted that use of any of the 3 instruments, especially the most recent and longest one (Version 3), would result in the reporting of a large number of symptoms at every assessment. This would make it difficult to distinguish background "noise" of daily life from drug-related problems, and this was one of the reasons one might argue for inclusion of drug attribution. However, in the setting of a controlled clinical trial, patients (or at least the patient groups used in the study arms) could have the instrument administered at baseline and the noise "subtracted" from future reports.

Finally, the underlying philosophy of a patient-friendly and self-directed instrument was addressed. This too can be seen as a part of the process of balancing the need for more reliable and comprehensive methods without over-reporting due to "leading questions." Once again the potential for an

electronically administered "intelligent" questionnaire was raised, with areas of questioning omitted or hidden unless a key question revealed the need to enquire further. For example, if a patient answered "yes" to: "Have you had any problems with your breathing?", then a series of subsidiary questions would be presented seeking the details and severity.

Summary

The Drug Safety Module Update not only informed participants about progress in this area, but also allowed them to determine future directions of development. The feedback from the breakout discussion groups to the plenary session raised several issues, which will be valuable to those engaged in research in this area. There was general acknowledgment of the need for and importance of work on standardized recording of adverse drug reactions, as well as recognition of the substantial work to be done. Researchers were encouraged to continue the process, and to look at the issues of the signal-to-noise problem and the potential for electronic data collection. We anticipate hearing about further progress on this project at OMERACT 8.

REFERENCES

1. Lassere MND, Johnson KR, Woodworth TG, et al. Challenges and progress in adverse event ascertainment and reporting in clinical trials. J Rheumatol 2005;32:2030-2.
2. Woodworth TG, Furst DE, Strand CV, et al. Standardizing assessment of adverse effects in rheumatology clinical trials. Status of OMERACT Toxicity Working Group March 2000: Towards a common understanding of comparative toxicity/safety profiles for antirheumatic therapies. J Rheumatol 2001;28:1163-9.
3. Lassere MND, van Santen S, Carlton K, et al. Generic patient self-report and investigator report instruments of therapeutic safety and tolerability. J Rheumatol 2005;32:2033-6.
4. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for outcome measures in rheumatology. J Rheumatol 1998;25:198-9.