Challenges and Progress in Adverse Event Ascertainment and Reporting in Clinical Trials


ABSTRACT. Toxicity, safety, and tolerability are integral facets of patient risk/benefit decisions, yet the capacity to define, measure, and compare these aspects is underdeveloped compared to aspects of efficacy. There are many reasons for this, scientific and administrative, but all are surmountable. Probably the greatest primary obstacle is the absence of a measurement instrument designed specifically for this purpose. There are increasing calls from various stakeholders for better evidence, and therefore better ascertainment, in this area, especially in randomized trials, and for these reasons OMERACT began deliberations about these concepts in 1994. A prototype coding instrument (the Rheumatology Common Toxicity Criteria) was developed and discussed at OMERACT 5. In the 2 years before OMERACT 7, a process of concept development and iterative design and testing were conducted to develop a patient self-report and investigator-reported adverse event instruments designed for use in trials at the time of visit. The predominant workload is performed by the patient in a self-report checklist, which is then mapped by the trialist onto a medically sophisticated version. This article presents background on the process of developing a dual adverse event instrument, which was presented and critically discussed in detail at OMERACT 7. (J Rheumatol 2005;32:2030–2)

Key Indexing Terms:
ADVERSE EVENT REPORTING PATIENT QUESTIONNAIRES RANDOMIZED CONTROLLED TRIALS

Challenges and Progress in Adverse Event Ascertainment and Reporting

The decision to treat a patient should be predicated on the premise that the treatment will do more good than harm. If it is likely that the treatment will work as well as or better than other treatments and its associated risks are minimal, then it would be recommended unless prohibited by cost or nonavailability. Given 2 equally effective treatments, the treatment with lesser side effects would be preferred. Knowledge on whether treatment has been shown to work (efficacy), how it works compared to others (relative efficacy), how it works in practice (effectiveness), what risks are associated with it (side effects, adverse events), and thus its overall utility compared to others (relative risk-benefit) comes from diverse sources, including clinical trials. Clinical trials, however, evaluate efficacy and toxicity asymmetrically. Although decision-making by practitioners and reimbursement authorities requires both efficacy and safety input, drug registration has been dominated by formal efficacy trials with toxicity assessed less rigorously in comparison. Drugs are registered if they are effective with little toxicity, not if they are safe with little efficacy. This is appropriate. However, the absence of symmetry has probably contributed to less formalism regarding the assessment of safety. There is no universal instrument or summary index of drug safety akin to, say, the Medical Outcome Study Short-Form 36 (SF-36)2,3 instrument to assess physical and mental health status. But there is clearly a need for such an instrument. Individual clinical decisions need better information to assess risk-benefit. Systematic metaanalyses, economic assessments, and policy decisions need standardized quantitative data for risk-benefit evaluations. Clearly, we need standardized data collection methods and instruments to facilitate the symmetry of measuring risks, as well

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; Novartis Pharma, Basel, Switzerland; Geffen School of Medicine at the University of California in Los Angeles, Los Angeles, CA, USA; Academic Rheumatology Unit, University of Bristol, Bristol, UK; Centre for Global Health, University of Ottawa, Institute of Population Health, Ottawa, Canada; St. Vincent Hospital, University of New South Wales, Sydney; 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, Joint Appointee University of NSW, Department of Rheumatology, St. George Hospital, Sydney; K.R. Johnson, MD, Senior Research Officer, Department of Clinical Pharmacology, Mater Hospital, University of Newcastle; T.G. Woodworth, MD, Executive Director, Global Therapeutic Area Profiling Head, ABGH, Novartis Pharma; D.E. Furst, Carl M. Pearson Professor of Rheumatology, Geffen School of Medicine, UCLA; J.F. Fries, Professor of Medicine, Stanford University School of Medicine; J.R. Kirwan, BSc, MD, Academic Rheumatology Unit, University of Bristol; P.S. Tugwell, Director, Centre for Global Health, University of Ottawa, Institute of Population Health; R.O. Day, Professor of Clinical Pharmacology, St. Vincent Hospital, University of New South Wales; 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, Kogarah, 2217 NSW, Australia. E-mail: lasserem@sesahs.nsw.gov.au
as benefits, and the place to begin is in clinical trials. Comparative clinical trials are required to enable rigorous benefit versus risk evaluations and reimbursement decisions, but the challenge of standardized toxicity assessment has not yet been met by the development of measurement tools for this purpose.

In reflecting on his work on the Stanford Toxicity Index, Jim Fries noted in 1991:

The central attributes of therapeutics are efficacy and toxicity. Unlike efficacy, the assessment of toxicity has received essentially no attention. Rather, toxicity has been assessed by the [US] Food and Drug Administration by looking at individual toxic events such as abdominal pain, rash, or bone marrow depression to see if they do or do not fall within “acceptable” limits...

There is an urgent need for quantitative summary measures for drug toxicity. The drug that is associated with rare fatal bone marrow toxicity may prove far safer than a drug that causes frequent gastrointestinal hemorrhage. A rational basis for clinical choice or for reaching a conclusion in clinical trials cannot be established unless comparisons between drugs are made across all potential toxicities, taking into account their frequency and severity. It is hard to imagine how a clinical trial can be adequate without such measurements.

What Are the Challenges of Adverse Event Ascertainment and Reporting?
Although there is no fundamental conceptual barrier to this work, the challenges are large and must not be underestimated.

1. Safety is inherently multi-organ and thus more complex than most efficacy assessments, which concentrate on a relatively narrow therapeutic goal (such as reducing pain).
2. Important patient-centered aspects of adverse events, such as intensity, frequency, duration, and impact on activities, need to be considered.
3. Careful attention to proper instrument design and testing (components, weighting, aggregation to profiles, collapse to single summary index) is essential.
4. The process of adverse event data collection should be easy to use for patients and investigators, and calculation and interpretation should be straightforward for the analyst. These issues might be seen as equivalent to those that have been addressed for other universal instruments such as the SF-36 for measuring health status or the utility measure QALY (quality-adjusted life years).

There is a large background literature on the challenges of adequately characterizing the side effect profile of treatments in clinical trials. For example, adverse event ascertainment in phase II trials is central because toxicity is often the main endpoint. This setting would greatly benefit from a universal adverse event instrument.

Early approaches to characterizing toxicity were conducted using recommendations from the World Health Organization (WHO) and US National Cancer Institute (NCI). These systems (WHO and NCI grading systems) were developed to code toxicity by degree of severity. They were sometimes used during clinical trials as criteria for termination, but more commonly were used after trial completion for coding and analysis of volunteered adverse events. Both grading systems have been used in oncology trials, but the process of toxicity collection and reporting remains suboptimal in cancer treatment, and in other medical areas. Only 39% of the trial reports had either a detailed description of the severity or used a published toxicity index. These authors also noted the great variation in toxicity reporting within and across different fields of medicine and called for better standardization of adverse event reporting.

What Has OMERACT Done in This Area to Date?
OMERACT is an international collaborative group functioning since 1992. Its explicit aim is to develop outcome measures, primarily in the setting of clinical trials. A working group on outcome measures of drug safety in rheumatology began in preparation for OMERACT 2 in 1994 (led by Peter Brooks), and concluded that the issue needed more attention and that adverse event reporting must be all-inclusive and harmonized. Deficiencies in classification and collecting these data were seen as causing an unnecessary loss of information, and solving this problem was considered a priority. A critical review of the methodological quality of toxicity assessment in clinical trials of different medical fields concluded that reports in rheumatology failed to meet over half the criteria they had selected. Among the deficiencies commonly found were (1) lack of consistency in who obtained adverse events and how they were obtained, (2) lack of adequate information regarding the timeframes involved, (3) absence of severity rating, and (4) absence of judgment regarding causation.

In 1998 at OMERACT 4 the Drug Toxicity Working Party defined 7 attributes of interest in toxicity assessment: frequency, severity, importance to patient, importance to clinician, impact on activities, impact on economic resources, and integration of benefit with adverse events. These attributes were later used as criteria to evaluate 4 existing toxicity instruments in rheumatology, the Stanford Toxicity Index (STI), the Morgan Index, the JAQQ, and the Juvenile Arthritis Quality of Life Questionnaire (JAQQ), none of which achieved more than 3 criteria.

At the OMERACT 5 meeting in 2000 there were 2 further developments. The toxicity index (STI) was revised to include an item rating patients’ overall satisfaction, taking into consideration both benefits and adverse effects. The revised STI was considered to have met 6 of the 7 attributes of interest in toxicity assessment, with the remaining one, impact on activities, being recorded elsewhere in the Health...
Assessment Questionnaire, of which the the STI is part. The next step was to test whether the STI met the OMERACT filter (truth, discrimination, and feasibility)26, applied to codify performance characteristics desired of new measures. The meeting also recommended that rheumatology trials develop a coding instrument similar to the Common Toxicity Criteria (CTC) developed by the NCI27 for use in oncology trials. The result was called the Rheumatology Common Toxicity Criteria28.

Preparations for OMERACT 7

These developments illustrate the progress that was made in this area over the previous meetings of OMERACT, and laid the groundwork for further developments in preparation for OMERACT 7. Efforts were directed toward an instrument that enabled investigator recording of an adverse event directly onto the case report form at the time of the patient visit (rather than coding the recorded adverse events at a later time). The instrument was crafted for use specifically in randomized trials where patients are seen about every month. Further, in a move to increase ease of use (i.e., improve feasibility), the idea of 2 instruments emerged. One would be completed by the patient in advance of interview, and would be used to aid the investigator when completing the definitive record. These 2 new tools have been developed (the Patient Self-Report Adverse Event Instrument and the Investigator Report Adverse Event Instrument) and the major effort to test and refine these new tools, which was presented at OMERACT 7, is reported elsewhere in these proceedings29,30.

REFERENCES

4. Fries JF. The hierarchy of quality-of-life assessment, the Health Assessment Questionnaire (HAQ), and issues mandating development of a toxicity index. Control Clin Trials 1991;12 Suppl:1065-17S.