Generic Patient Self-Report and Investigator Report Instruments of Therapeutic Safety and Tolerability

MARISSA N.D. LASSERE, KENT R. JOHNSON, SUSANNE VAN SANTEN, KERRI CARLTON, JOY RAPPO, RACHEL MICHAEL, JOHN R. KIRWAN, and JOHN P. EDMONDS

ABSTRACT. A patient self-report instrument was designed as a patient event index that maps to a parallel investigator instrument. Event importance (a composite of severity, frequency, and duration) was reported, but attribution was not required. The patient instrument used a checklist but also allowed for spontaneous reporting for new or unusual events. The investigator instrument (also a checklist) includes all events reported by the patient, as well as events such as signs, investigations, and diagnoses that would not generally be known to the patient. Presently, both patient and investigator instruments are to be used alongside current methods of adverse event reporting in clinical trials. The patient instrument would serve as a safety/tolerability index, whereas the investigator instrument would be a fully quantifiable (appropriately weighted), standardized adverse event index. As in many methodological projects in medicine, the overriding problem was the tradeoff between validity (comprehensiveness and accuracy) and feasibility (clarity and short administration time) in instrument development. A summary of pilot studies and results of instrument reliability and validity are presented. (J Rheumatol 2005;32:2033–6)

Key Indexing Terms:
ADVERSE EVENT PATIENT QUESTIONNAIRE RANDOMIZED CONTROLLED TRIALS

The importance of careful measurement of adverse events in clinical trials has long been recognized. It enables the reader to make a reasoned judgment of risk/benefit or safety/efficacy, from the individual patient to health policy levels. However, there is considerable asymmetry regarding the process on measuring efficacy and safety. Whereas the process of measuring efficacy is the aim of hypothesis-driven experimental design, measuring safety is less formalized from an inferential perspective. Whereas efficacy is measured across a few domains of interest, adverse event assessment is inevitably measured across multiple organ systems. Finally, whereas attribution of causality is implicit for efficacy, each adverse event requires a determination of attribution — traditionally using descriptors such as "unrelated," "possibly related," "probably related," and so on. These and other barriers to adverse event ascertainment and reporting were reviewed in the companion article of these proceedings[1].

With the maturing of programs directed at evidence-based decisions for therapeutics, there is greater need to quantitatively measure risk/benefit. Only in this manner can the full impact of therapeutic propositions be observed and compared. Cost-effectiveness methodology presumes an ability to quantify benefit and risk, and economic exercises often lack formal accounting of drug toxicity. This deficiency arises from the absence of an appropriate generic instrument to accurately and systematically record information on adverse events from interventions studied in clinical trials.

In the last 2 years our group has undertaken an iterative process of developing and testing instruments that can be used for standardized adverse event ascertainment and reporting in clinical trials. The goal was to develop a reliable, accurate, and comprehensive instrument (one that met the OMERACT filter[2]) that was fully quantitative. On a patient basis, events would be reported both as profile scores by organ system weighted by impact and as a single summary adverse event score. For populations it will allow standardized reporting of drug profiles.

A patient self-report instrument was designed as a patient event index that maps to a parallel investigator instrument. Event importance (a composite of severity-frequency-dura-

From the Department of Rheumatology, St. George Hospital, University of New South Wales, Sydney, Australia; Department of Clinical Pharmacology, Mater Hospital, University of Newcastle, Newcastle, Australia; Amsterdam Medical School, Amsterdam, The Netherlands; and the Academic Rheumatology Unit, Bristol Royal Infirmary, Bristol, England.

M.N.D. Lassere MB, BS, Grad Dip Epi, PhD, FRACP, FAFPHM, Associate Professor in Medicine, Conjoint Appointee University of NSW, Department of Rheumatology, St. George Hospital; K.R. Johnson, MD, Senior Research Officer, Department of Clinical Pharmacology, Mater Hospital, University of Newcastle; S. Van Santen, Amsterdam Medical School; K. Carlton, BSc, MSc, Research Officer; J. Rappo, BSc, M Epi, Research Nurse; R. Michael, BCom, Research Officer, St. George Hospital; J.R. Kirwan, Consultant and Reader in Rheumatology, Academic Rheumatology Unit, Bristol Royal Infirmary; J.P. Edmonds, Professor of Rheumatology, University of NSW, Director, Department of Rheumatology, St. George Hospital.

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

tion) was reported, but attribution was not required. The patient instrument used a checklist but also allowed for spontaneous reporting for new or unusual events. The investigator instrument (also a checklist) included all events reported by the patient, as well as events such as signs, investigations, and diagnoses that would not generally be known to the patient. Both patient and investigator instruments were to be used alongside current methods of adverse event reporting in clinical trials. The patient instrument would serve as a safety/tolerability index, whereas the investigator instrument would be a fully quantifiable (appropriately weighted) standardized adverse event index.

As in many methodological projects in medicine, the overriding problem was the tradeoff between validity (comprehensiveness and accuracy) and feasibility (clarity and ease and duration of administration) in instrument development. An outline of studies and results presented during the plenary session is published elsewhere in these proceedings[3].

The instruments were developed from the Rheumatology Common Toxicity Criteria (RCTC)[4], the Stanford Toxicity Index[5-7], and the symptom list included in the complete Health Assessment Questionnaire (not the shorter disability index that is familiar to rheumatologists)[8]. We used a combination of qualitative and quantitative research methods and conducted a series of studies to develop 3 versions of the instruments. Each version consisted of both the patient self-report instrument and the investigator report instrument. With each succeeding version, the instruments became more comprehensive but at a cost in the ease of administration. The aim was to find the optimal tradeoff between these two.

Qualitative research methods included a series of taped interviews of patients, nurse metrologists, clinicians, and clinician scientists, and feedback using semistructured questionnaires each time the instruments were tested. The themes emerging from these interviews were (1) the need to improve the quality of the information obtained at the time of the visit, (2) the need to record adverse events in a more systematic fashion, and (3) the need to code adverse events during the visit. At many points in this process the same provisional solution was suggested — move the burden of data acquisition back to the patient by building in a patient self-report instrument that mirrored and mapped to the investigator instrument. Both would be completed at the time of trial visit, first the patient instrument, then the investigator instrument, with the latter then informing the case report form. Feedback from the patients in particular was helpful and led to extensive modifications. Patients were also helpful regarding the comprehensibility of the language of the events, the logic of the severity grading order, and the overall organization of the instrument.

Summary information of the 3 versions is shown in Table 1, and an excerpt of the content of version 3 is provided in Table 2 for illustration. The patient component of the instrument has grown from 4 pages in version 1 to 7 in versions 2 and 3, and the investigator component from 6 pages to more than 17 in succeeding versions. Many further terms were added to increase comprehensiveness, and the grading system was changed from one where duration and severity are combined in versions 1 and 2 to one where they are separately assessed in the investigator version 3. An electronic version of the current version is being developed, which should enhance its acceptability. It will automatically also provide Medical Dictionary for Regulatory Activities (MedDRA), International Classification of Diseases, and other codings.

Table 1. The 3 versions of the patient self-report and investigator report adverse event instruments.

Instrument	Pages	Categories	Items	Severity Grades
Version 1				
Patient	4	11	72	3
Investigator	6	11	72	4
Version 2				
Patient	6	15	126	3
Investigator	15	15	178	4
Version 3				
Patient	7	17	152	3
Investigator	17*	17	104*	7

* Instrument unfinished.

Table 3 provides a summary of excerpted studies of the reliability and validity of different versions of patient and investigator instruments. The reliability of all patient and investigator instruments was very good. As expected, content validity was greater in successive versions. Only the results for version 1 are available for other studies of validity, and it performed well regarding both discriminant and criterion validity. Ease of administration varied from 5 to 15 minutes for patients and from 5 to 30 minutes for investigators depending on their experience and number of events marked. Clearly, 30 minutes is too long for investigators and we are looking at ways this can be reduced, including using an electronic version of the instrument. However, the nurse investigators felt more confident regarding ascertaining and reporting events and preferred the provision of clear definitions of event severity.

These data lend support regarding the OMERACT filter properties of these instruments, but further work is required before they are ready for recommendation; work includes finishing construction of the version 3 investigator instrument. How “long” and “wide” should it be? Would handheld computer usage help reduce the load? Should we add in attribution? A much broader spectrum of patients regarding age and medical conditions is needed. The question of a stand-alone patient instrument is very provocative and will be aggressively explored. To date, in the hands of the

Table 2. Example from version 3 of the patient and investigator instruments.

Patient Instrument	Investigator Instrument
Stem	Stem
In the last 4 weeks if you have experienced any of the complaints listed below, place a tick in the box to indicate whether it was mild, moderate or severe.	In the last 4 weeks has the patient experienced any of the complaints; do not determine attribution
Response Grade	Response Grade
Mild	1A. Mild: short duration
It was mild	1B. Mild: recurrent or persistent
Your activities did not change because of the complaint	2A. Moderate: short duration
You did not see a doctor or require prescription treatment for the complaint	2B. Moderate: recurrent or persistent
	3A. Severe: short duration
	3B. Severe: recurrent or persistent
	4. Includes life-threatening
Response event	The response grade is described for each event.
A. General complaints	For example "Af" from the patient form maps to "Af" on the investigator form
a. Tiredness or fatigue	"Af": Generally unwell; consider other definable conditions such as fever, fatigue before coding.
b. Fever, chills, or sweats	Grade 1A: Non-specifically unwell, less than two weeks, not interfering with function; Grade 1B: Non-specifically unwell, more than two weeks, not interfering with function.
c. Weight loss	
d. Weight gain	
e. Problems sleeping	
f. Did not feel well	
g. Other (please describe)	

Table 3. Summary of studies of the reliability and validity of different versions of patient and investigator instruments.

Version	Study Design	Study Results
Reliability		
Patient version 2	10 rheumatology clinic patients; test-retest at 7 days	Single measure ICC (95% CI); ICC = 0.91 (0.69, 0.98)
Patient version 2	67 rheumatology clinic patients, 62 other clinic patients (cardiology, renal, dermatology, HIV, endocrine, thoracic medicine, etc); test-retest at 24 h	Without including grade, single-measure ICC: ICC = 0.96 (0.94, 0.98), SDD = 18.6 SDD % of actual max = 12.6% With including grade: ICC = 0.94 (0.90, 0.97), SDD = 18.6 SDD % = 12.6%
Investigator version 1	8 paper patients, 7 medical investigators, interobserver	Average measure ICC = 0.92
Investigator version 2	8 paper patients, 7 medical investigators, interobserver	Average measure ICC = 0.89
Content and discriminant validity		
Investigator version 1	Retrospective coding of 12 mo, 3-arm blinded RA trial of treatment targets: low DAS target vs normal CRP target vs usual care. RA therapy escalated until target achieved or unacceptable toxicity. 485 adverse events in 40 patients	80% of total events capture as coded by a medical student and a rheumatologist on 2 separate occasions 95% agreement between 2 raters More adverse events in targeted arms, DAS, and CRP than usual care (DAS arm reduced damage progression on MRI at 12 mo)
Investigator version 2	As above	97% of total events capture as coded by rheumatologist
Investigator version 1	Retrospective coding of 12 week, 4-arm, double-blind RA trial of budesonide (2 doses) vs prednisone vs placebo: 632 adverse events in 143 patients	62% of total events capture as coded by rheumatologist
Investigator version 2	As above	91% of total events capture as coded by rheumatologist
Patient version 3	As above	94% of total events capture as coded by rheumatologist
Criterion validity		
Patient and Investigator version 1	Pilot testing in a prospective testing in second 3-arm RA trial of treatment targets: swollen joint count ≤ 2 target vs normal CRP target vs usual care. RA therapy escalated until target achieved or unacceptable toxicity. 18 patients, 1 nurse investigator	Traditional spontaneous reporting, 5 events in 3 patients Checklist reporting version 1:201 complaints in 18 patients 97% agreement between investigator and patient on event 85% agreement between investigator and patient on severity; disagreement on severity always rated higher by patient

ICC: intraclass correlation coefficient; SDD: smallest detectable difference, SDD % of actual maximum[8]; DAS: Disease Activity Score; CRP: C-reactive protein.

patients the instrument seems to have good metric properties. Further content development must be undertaken and formally tested. These include issues of attribution and weighting and aggregation. Further process aspects include developing reporting systems for drug profiles, and instrument compatibility with different data dictionaries. Finally, broader experimentation with the use of the instruments in randomized trials is essential.

REFERENCES

1. Lassere MND, Johnson KR, Woodworth T, et al. Challenges and progress in adverse event ascertainment and reporting in clinical trials. J Rheumatol 2005;32:2030-2.
2. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for outcome measures in rheumatology. J Rheumatol 1998;25:198-9.
3. Lassere MND, Johnson KR, Boers M, et al. Standardized assessment of adverse events in rheumatology clinical trials: Summary of the OMERACT 7 drug safety module update. J Rheumatol 2005;32:2037-41.
4. Woodworth TG, Furst DE, Strand V, 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.
5. Fries JF, Spitz PW, Williams CA, Bloch DA, Singh G, Hubert HB. A toxicity index for comparison of side effects among different drugs. Arthritis Rheum 1990;33:121-30.
6. Fries JF. ARAMIS and toxicity measurement (Arthritis Rheumatism and Aging Medical Information System). J Rheumatol 1995;22:995-7.
7. 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:106S-17S.
8. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23:137-45.