Patient Based Method of Assessing Adverse Events in Clinical Trials in Rheumatology: the Revised Stanford Toxicity Index

VIVIAN WELCH, GURKIRPAL SINGH, VIBEKE STRAND, JIM FRIES, MAARTEN BOERS, DENA RAMEY, RICHARD O. DAY, PETER BROOKS, PETER TUGWELL, JENNIFER CLINCH, and BETSY KRISTJANSSON

ABSTRACT. We describe the progress towards developing a patient rated toxicity index that meets all of the patient-important attributes defined by the OMERACT Drug Safety Working Party. These attributes are frequency, severity, importance to patient, importance to the clinician, impact on economics, impact on activities, and integration of adverse effects with benefits. The Stanford Toxicity Index (STI) has been revised to collect all attributes with the exception of impact on activities. However, since the STI is a part of the Health Assessment Questionnaire (HAQ), impact on activities is collected by the HAQ. In particular, a new question asks patients to rate overall satisfaction, taking into consideration both benefits and adverse effects. The next step in the development of this tool is to ensure that the STI meets the OMERACT filter of truth, discrimination, and feasibility. Although truth and feasibility have been confirmed by comparisons within the ARAMIS database, discrimination needs to be assessed in clinical trials. (J Rheumatol 2001;28:1188–91)

Key Indexing Terms:
CLINICAL IMPORTANCE QUESTIONNAIRE TOXICITY PATIENT RATINGS ADVERSE EVENTS

The measurement of side effects and safety in clinical trials in rheumatology has received increasing attention[1]. However, even in recent trials, adverse events captured by clinicians and patient reported symptoms are not collected by validated instruments[2-4]. For this reason, discontinuations due to side effects are used as an indicator of patient rated intolerance of side effects in Cochrane metaanalyses[5].

OMERACT is in the process of defining a grading system for rheumatology common toxicity criteria (CTC)[6]. However, this grading system is designed for use by clinicians. The CTC are based on diagnosis (if possible) so that a patient reported side effect of "nausea" would likely be coded as "gastroenteritis severity 1" by the clinician.

Empiric evidence reveals differences in patient rated and clinician rated severity of symptoms, with clinicians rating either lower[7] or higher severity than patient ratings[8]. Further, patient rated importance of side effects may not be captured by ratings of severity and frequency[9].

In 1997, the OMERACT Drug Safety Working Party defined 7 important attributes for patient based collection of safety, including: frequency, severity, importance to patient, importance to clinician, impact on activities, impact on economic resources, and integration of benefit with adverse events. A literature search identified only 4 patient based methods of collecting safety data in rheumatology trials. These instruments are the Stanford Index[10], the patient oriented symptom index[11,12], the Morgan index[13], and the juvenile arthritis quality of life questionnaire[14,15].

None of these instruments measured all attributes of interest[1]. However, the Working Party considered the Stanford Toxicity Index (STI) to have potential for being able to capture most attributes. The Working Party proposed that the STI be revised to incorporate missing attributes and that it be validated for use as an outcome measure in clinical trials.

Our paper describes discussions to: (1) propose revisions to the STI that will satisfy the attributes of the OMERACT Drug Safety Working Party for patient based safety assessment, and (2) propose methods for validation of the revised STI as an outcome for clinical trials according to the OMERACT filter[16].

From the Clinical Epidemiology Unit, Loeb Research Institute, Ottawa, Ontario, Canada; Department of Medicine, Stanford University, Palo Alto, California, USA; Department of Clinical Epidemiology, VU University Hospital, Amsterdam, The Netherlands; Department of Clinical Pharmacology, St. Vincent's Hospital, University of New South Wales, Sydney; Faculty of Health Sciences, University of Queensland, Sydney, Australia; Department of Medicine, Ottawa Hospital, Ottawa, Canada.

V. Welch, MSc, Loeb Research Institute; G. Singh, MD; C.V. Strand, MD; J. Fries, MD, Department of Medicine, Stanford University; M. Boers, MD, VU University Hospital; D. Ramey, MSc, Department of Medicine, Stanford University; R.O. Day, MD, FRACP, Department of Clinical Pharmacology, St. Vincent's Hospital; P. Brooks, MD, FRACP, Head, Faculty of Health Sciences, University of Queensland; P. Tugwell, MD, MSc, FRCCP, Department of Medicine, Ottawa Hospital, B. Kristjansson, PhD, Department of Epidemiology and Community Medicine, University of Ottawa; J. Clinch, PhD, Ottawa Health Research Institute.

Address reprint requests to V. Welch, Loeb Health Research Institute, 1053 Carling Avenue, CEU F6, Ottawa, Ontario K1Y 4E9, Canada. E-mail: vwelch@lri.ca

ATTRIBUTES ASSESSED BY THE STANFORD INDEX

The STI captures 2 types of side effects: patient rated *symptoms* and side effects attributed to specific drugs[10]. Fries, *et al* conducted a validation study that compared 8 different methods of calculating a toxicity index. They found that the indices calculated on the basis of patient reported symptoms (independent of the drug) were less able to discriminate between patients taking different drugs[10]. In this study, clinician rated side effects had less interobserver variability within side effects and greater variability between side effects. Based on this validation study, the method proposed for calculating the STI was based on clinician rated side effects and importance.

The initial version of the STI measures 3 of the above 7 attributes of patient rated side effects: frequency, severity, and importance to the clinician (Table 1).

Ongoing discussions with the OMERACT Drug Safety Working Party have reviewed options to capture the remaining 4 additional attributes: patient importance, impact upon activities, economic resources, and integration of benefits and adverse events. These discussions have identified revisions necessary to capture all attributes except the impact on economic resources.

Patient importance

Patient importance is captured by the question: "How important was this side effect to you?" and recorded by a categorical scale from 0 to 10 with the anchors of 0, "not at all important" and 10, "very important." Early validation studies on patient assessed importance found that patient and physician weightings of the severity of each side effect were different for 55% of the side effects. However, the relative ranking of side effects was very similar whether patient or physician weights were used. Patient weightings had greater interobserver variability, but patient weightings showed less difference between different side effects. Patients had difficulty with rating the importance of laboratory side effects[10].

Discussions with the OMERACT Safety Working Party proposed revising the patient importance question to explicitly take into account the influence of the side effect on the

Table 1. Evaluation of the Stanford Toxicity Index (STI) according to 7 important patient-assessed attributes.

Attribute Assessed	STI Initial	STI Revised	Captured	Not Captured
Frequency	+	+	Patients select from a laundry list of symptoms and answer questions about each symptom: (Patients can also list any other symptom they want) 1. What was the symptom? 2. Which drug caused it? 3. Did you stop the drug because of it? 4. Severity (mild, moderate, severe) 5. Importance (not at all to very much)	Certain symptoms and side effects are absent- e.g., hirsutism, weight gain, appetite gain, hematuria, stool occult blood positivity No specific question re: frequency, but the Working Party decided frequency is captured by severity and importance
Severity	+	+	As above, patients rank severity (mild, moderate, severe)	
Importance to patient	–	+	The revised HAQ asks patients to rank "importance of side effect to you"	Not incorporated into STI However, this is an analysis issue, since data exist to calculate the STI based on patient's attribution of importance
Importance to clinician	+	+	Clinicians rate the importance of the side effects and symptoms recorded by patients	
Impact on activities	–	–	The STI captures ADL limitations at every visit, but does not attribute a change to disease activity or adverse events	The Stanford Group feel that since the HAQ is included in the ACR 20 and will be compared with a control group this impact of AE on activities will be captured - and that to sort out the separate effect of AE on activities is not easy
Impact on economic resources	–	+	Captures endoscopy and other outpatient procedures, hospitalization, emergency room, and surgery	Should await recommendations on reference cases from the OMERACT Health Economics Task Force
Integration of benefit with AE? Bipolar instrument	–	+	STI measures "patient satisfaction" Overall satisfaction: Please rate your satisfaction with each drug as a treatment for your arthritis on a scale of 0–10, where 0 = totally dissatisfied and 10 = extremely satisfied	Safety Working Party to decide if this captures this attribute and how this interfaces with the Patient Global Scale in the ACR 20

AE: adverse events; STI: Stanford Toxicity Index; HAQ: Health Assessment Questionnaire; ADL: activities of daily living.

patient's activities. This would capture the attribute of importance of activities in the same question.

Proposed modification. The question has been modified as follows: "Considering the frequency, severity, and impact on activities, how important was this side effect to you?"

Results from a small sample of patients of the Arthritis, Rheumatism and Aging Medical Information System database were presented at the PANLAR Congress in Montreal in June 1998[17] (Table 2). The results suggest that patient ranked importance is different from the clinician based STI ranking. For example, for side effects with data from more than 10 patients, clinicians ranked fatigue, heartburn, nausea, and indigestion as the most important side effects. In contrast, patients ranked mucosal ulcers, lower abdominal pain, tinnitus, and diarrhea as the top 4 side effects (Table 2).

This is quite different from the results of the hypothetical exercise performed during the initial validation studies of the STI. In the initial validation studies, patients and physicians were presented with cards describing the side effects and their implications and were asked to rank them according to importance. The results of these validation studies showed that the relative ranking of side effects was similar for patients and physicians[10].

Other empiric evidence on patient rated importance suggests that patient importance should be used for weightings. Talley, *et al*[9] evaluated a patient rated dyspepsia index that uses patient rated importance to weight the items. They found that factor analysis without the patient rated weighting resulted in quite different factor loading for the domains they considered to have less face validity.

The next step. The OMERACT Safety Working Party will be asked how best to proceed to include a quantitative estimate of patient importance. The major decision needed that will drive subsequent steps is whether to aim for: (a) a mean importance ranking across a group of patients, or (b) a score for the individual patient. The latter would be more consistent with the way in which benefits are assessed, where each patient's view of their pain and global score are used. How can this individual patient approach be applied? One option would be to develop a visual analog scale (VAS) that asks the patient to rate the importance of each adverse effect (such as multiplying the severity/frequency by the importance). Another way would be to review all the adverse effects and then ask the patient to indicate how important all of these are overall on a single VAS.

Impact on activities

The STI is part of the Health Assessment Questionnaire, the instrument most widely used to assess disability, as one of the 7 required components of the American College of Rheumatology improvement criteria, ACR20/50/70. Currently the patient assessed impact on activities is not measured separately for benefit and safety, and the Stanford group feel that it is not useful to attempt to separate out which aspects of disability are due to the disease and which aspects are due to the adverse events.

The next step. The Safety Working Group should determine if there is agreement that it is unnecessary to disaggregate to capture the effects of drug adverse events separate from the disease and the beneficial effect of the intervention.

Economic resources

The STI captures data on endoscopy, outpatient procedures, hospitalization, emergency visits, and surgery. However, it does not capture extra physician visits due to drug associated side effects.

Table 2. Sample of data presented at the PANLAR Congress 1998, showing difference in patient and clinician rankings of side effects sorted by number of patients.

	N	STI Weight	STI Weight Rank	Mean Patient Importance	Importance Rank
Side effect					
Insomnia	5	4	5	88	14
Indigestion	7	3.8	2.5	54	5.5
Rash	7	4.7	7	54	5.5
Tinnitus	8	4.3	6	50	3
Lower abdominal pain	8	5.1	10	38	1
Vertigo	9	4.8	8	78	12
Alopecia	9	5	9	81	13
Diarrhea	9	7.8	14	50	4
Fatigue	10	3	1	57	7
Mucosal ulcers	10	5.8	12	49	2
Heartburn	12	3.8	2.5	69	10
Purpura	13	6.5	13	67	9
Upper abdominal pain	17	5.7	11	69	11
Nausea	37	3.9	4	64	8

STI: Stanford Toxicity Index, rated by the clinician.

The next step. No recommendation will be made on this until after the OMERACT Economics Task Force develops recommendations on what costs should be collected for the Reference Case Project.

Integration of benefits and adverse events

The satisfaction question added to the STI in 1998 was revised in 1999 to ask the patient to explicitly take into account the benefits and side effects. The revised question is phrased: Considering both effectiveness and side effects, please rate your satisfaction with each drug on a scale of 0–10. 0: You were "totally dissatisfied" and 10: You were "extremely satisfied."

This proposed modification will be presented to the Safety Working Party to see if they agree this captures this attribute.

HOW TO ENSURE THAT STANFORD INDEX MEETS OMERACT FILTER?

The next step is to propose a method of incorporating these patient assessed attributes into the overall Stanford Toxicity Index. Currently, the STI is calculated using 3 scores: (1) the clinician importance of side effects (weighted by severity); (2) laboratory side effects (weighted by severity); and (3) hospital days (weighted by likelihood that they are attributed to drug)[18]. Scoring of the patient report component will need to be developed for each of the patient assessed attributes above.

The OMERACT filter involves assessing the truth, discrimination, and feasibility.

Truth

Truth is an indication of whether the STI measures what is intended.

Purpose, population, setting. The purpose, population, and setting are defined. Currently the STI has only been used in a postmarketing surveillance setting, not that of a clinical trial.

Content validity. The 7 different components discussed above (frequency, severity, importance to patient, importance to clinician, impact on activities, impact on economic resources, and integration of benefit with adverse events) reflect the content validity.

Construct validity. The revised patient rated STI should be positively related to the number of adverse event related discontinuations and negatively related to the overall efficacy of the treatment.

Face validity. For the current STI patients were surveyed as to understanding of symptoms, and confusing terms were removed. The same will be necessary for the additional items.

Discrimination

Sensitivity to differences, or responsiveness. This will need testing in clinical trial datasets; this could be done in the same manner used to determine the 7 items in the ACR20, by assessing the ability to discriminate between active and placebo groups.

Reliability. Repeat administration was used to refine the descriptions of symptoms given to patients. The new items will need to be tested in this manner.

Feasibility

The format and time required to administer is reported to be acceptable to patients. Data on completion rate of all questions are needed. Also, the ease of calculating and interpreting the toxicity index is important.

REFERENCES

1. Gazarian M, Tugwell P, Boers M, et al. Patient based methods for assessing adverse events in clinical trials in rheumatology. Progress report for the OMERACT Drug Toxicity Working Party. Outcome measures in rheumatology. J Rheumatol 1999;26:207-9.
2. Simon LS, Weaver AL, Graham DY, et al. Anti-inflammatory and upper gastrointestinal effects of celecoxib in rheumatoid arthritis. A randomized controlled trial. JAMA 1999;282:1921-8.
3. Langman MJ, Jensen DM, Watson DJ, et al. Adverse upper gastrointestinal effects of rofecoxib compared with NSAIDs. JAMA 1999;282:1929-33.
4. Moreland LW, Schiff MH, Baumgartner SW, et al. Etanercept therapy in rheumatoid arthritis. A randomized controlled trial. Ann Intern Med 1999;130:478-86.
5. Suarez-Almazor ME, Belseck E, Shea B, Wells G, Tugwell P. Methotrexate for rheumatoid arthritis (Cochrane Review). The Cochrane Library Issue 1. 1-1-2000. Oxford: Update Software.
6. Woodworth T, Furst D, Strand V, et al. Proposed rheumatology common toxicity criteria. 1-1-1998. http://www.ilar.org/omeract/toxintro.htm .
7. Calhoun EA, Bennett CL, Peeples PA, et al. Perceptions of cisplatin-related toxicity among ovarian cancer patients and gynecologic oncologists. Gynecol Oncol 1998;71:369-75.
8. Stephens RJ, Hopwood P, Girling DJ, Machin D. Randomized trials with quality of life endpoints: are doctors' ratings of patients' physical symptoms interchangeable with patients' self-ratings? Qual Life Res 1997;6:225-36.
9. Talley NJ, Verlinden M, Jones M. Validity of a new quality of life scale for functional dyspepsia: A United States multicenter trial of the Nepean Dyspepsia Index. Am J Gastroenterol 1999;94:2390-7.
10. Fries JF, Spitz PW, Williams CA, et al. A toxicity index for comparison of side effects among different drugs. Arthritis Rheum 1990;33:121-9.
11. Levine M. The patient oriented symptoms index: the POSI. Forthcoming.
12. Peloso PMJ. The validation of the Patient Oriented Symptoms Index for measuring patient side effects in rheumatology clinical trials. Toronto: University of Toronto; 1994.
13. Morgan SL, Baggott JE, Vaughn WH, et al. Supplementation with folic acid during methotrexate therapy for rheumatoid arthritis: a double-blind, placebo-controlled trial. Ann Intern Med 1994;121:833-41.
14. Duffy CM, Arsenault L, Watanabe-Duffy KN. Level of agreement between parents and children in rating dysfunction in juvenile arthritis and juvenile spondyloarthritides. J Rheumatol 1993;20:2134-9.
15. Duffy CM, Arsenault L, Watanabe-Duffy KN, Paquin JD. The juvenile arthritis quality of life questionnaire — development of a new responsive index for juvenile rheumatoid arthritis and juvenile spondyloarthritides. J Rheumatol 1997;24:738-46.
16. Boers M, Brooks P, Strand V, Tugwell P. The OMERACT filter for outcome measures in rheumatology. J Rheumatol 1998;25:198-9.
17. Singh G. PANLAR conference, Montreal, Canada, June 1998; ARAMIS database results on patient patient-rated severity.
18. Fries JF. ARAMIS and toxicity measurement. J Rheumatol 1995;22:995-7.