ARAMIS and Toxicity Measurement

JAMES F. FRIES

ABSTRACT. Side effects of medications make up an important part of adverse outcomes experienced by patients with rheumatic diseases. Quantitative measures to assess toxicity, however, have not been available, and this lack has limited estimates of the magnitude of effects and of differences in side effects among different drugs. This paper describes the development of the Arthritis, Rheumatism and Aging Medical Information System (ARAMIS) Toxicity Index and the issues arising in construction of such an index, and reviews early results in comparing toxicities of antirheumatic drugs. Findings have had major value in revising therapeutic strategies for rheumatic diseases, particularly rheumatoid arthritis, and have set the stage for development of toxicity–therapeutic ratios for different drugs. (*J Rheumatol 1995;22:995–7*)

Key Indexing Terms:
DRUG TOXICITY TOXICITY INDEX ARAMIS POSTMARKETING SURVEILLANCE

Rational clinical decision making, perhaps almost an oxymoron, requires quantification of the unquantifiable. If the typical decision requires selection of drug A or drug B, then the decision maker requires, at a minimum, an accurate knowledge of all the good things that can be expected to happen with each drug, and similar knowledge of all of the bad things that can be expected to follow each treatment. In the most profound sense this knowledge is required not just for the average of a group of individuals but for the specific patient for whom the decision looms.

To approach the issue of comparative effectiveness or of comparative toxicity of 2 agents, indexes must be constructed, and these indexes must attempt an aggregation of dissimilar items. Indexes always raise serious questions of intuitive (face) validity, as well as issues of measurement reliability and validity. They are usually controversial.

The Arthritis, Rheumatism and Aging Medical Information System (ARAMIS) outcome assessment paradigm has always included a major dimension of treatment toxicity[1]. It was inevitable, therefore, that when we began systematic approaches to pharmacoepidemiology[2] we would need to use a toxicity index, since we sought to compare different drugs, to look at effects over time, to examine effects of patient covariates, and so forth.

However, our initial literature search, in 1989, came up empty. No one had proposed methods of presenting a single index number representing drug toxicity. The Food and Drug Administration could not explain how it determined that a drug was unacceptably toxic, although case study suggested that discovery of a rare serious toxicity might lead to drug recall or that an increase in specific toxicity (e.g., transaminitis) over alternatives might lead to nonapproval. Thus, a drug with greater overall toxicity might sometimes be preferred to one of lesser overall toxicity. Perhaps attempts at overall quantitation represented a fool's errand, an attempt to quantify the unquantifiable. Or, perhaps, no one had used good datasets to approach the issue.

We believe strongly that the issues underlying rational clinical choice are too important to leave fallow simply because they are difficult or because they admit only to imperfect solutions. We have large, prospective, longitudinal datasets including patient reported toxicity, physician recorded toxicity, laboratory values, and with examination of all hospitalizations, and all deaths.

For the past several years, therefore, the ARAMIS Post-Marketing Surveillance Program has been developing methodology to approach a central question of therapeutics: Which drugs are most toxic, and by how much? We have developed and reported on a Toxicity Index that counts and weights symptoms, laboratory side effects, and hospitalizations both by side effect type and severity, and computes these into a single index number representing the toxicity of a particular medicine; this number is adjusted statistically for differences in length of time on different drugs and for differences in characteristics of patients receiving different drugs[3].

MATERIALS AND METHODS

Initial studies have consisted of analysis of many thousands of courses of disease modifying antirheumatic drugs (DMARD), nonsteroidal antiinflammatory drugs (NSAID), and prednisone therapy in 2747 patients with rheumatoid arthritis (RA) over about 8000 patient years of observation. Patients are studied prospectively and longitudinally. All patients have RA, and 5 databank centers were initially involved. Two centers (Santa Clara County and Saskatoon) represent community based populations, 2 (Wichita and Phoenix) represent private rheumatologic practices, and one center (Stanford) represents a university referral practice. Data consist of routine clinical information, demographics, diagnoses, symptoms, physical signs, laboratory findings, therapy employed, and data each 6 months from the Health Assessment Questionnaire (HAQ) detailing disability, symptoms, side effects, and economic effects. Development, validation, and methodology of the

From the Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305.

Supported by grant AM21393 from the National Institutes of Health to ARAMIS (Arthritis, Rheumatism and Aging Medical Information System).

J.F. Fries, MD, Associate Professor of Medicine, Stanford University School of Medicine.

Address reprint requests to Dr. J. Fries, 1000 Welch Road, Suite 203, Palo Alto, CA 94304-1808.

Toxicity Index have been reported. Hospitalizations are individually abstracted from discharge summaries, and deaths from discharge summaries and death certificates. Deaths in patients lost to followup are identified by use of the National Death Index[4,5].

The Toxicity Index includes components of symptoms, laboratory tests, and hospitalizations. Events resulting in hospitalization or death are attributed to a drug if the attending physician made that attribution. If no such attribution was made, they are fractionally attributed on the basis of the relative risk of the event while using this class of medications compared with the relative risk of the event while not using this class of medications. If more than one potential contributing drug is present, as for example with the upper gastrointestinal (GI) hemorrhage associated with both prednisone and an NSAID, attribution is further fractionally allocated between the potential offending agents. Symptoms are additionally categorized as mild, moderate, or severe, and scores weighted accordingly. Units for the Toxicity Index are scored per year of exposure to the agent.

Hospitalization records and death certificates for all events occurring during each 6 month reporting period are reviewed to determine the relationship of hospitalizations and death to drug use. Discharge summaries are supplemented by ARAMIS clinical data and HAQ data. To be attributed, events have to be either directly attributed by a physician to a specific drug, to a specific drug and a clinical problem, or likely to be drug related based upon a known relationship (previously reported in the literature) between a class of drugs and a particular set of symptoms or complications. Thus, any toxic event, no matter how rare, can be attributed if the physician attending the patient believes the drug to be responsible, while in the absence of physician notation, a probable drug relation can be counted as well, but only if the relationship between drug and event is established. The algorithms may be readily applied by trained abstractors in most instances, and in ambiguous situations records are further reviewed by an ARAMIS physician who makes the final judgment[3].

The Toxicity Index is computed as the sum of symptom side effects, laboratory side effects, and hospital days resulting from adverse reactions. Symptom side effect weights range from 0 to 10 and are multiplied by the severity factor, 0.5 for mild, 1.0 for moderate, and 1.5 for severe. The same procedure is employed for laboratory side effects. The number of hospital days is multiplied by the weighting factor for a hospital day (8.4) and fractionated according to the rules described above.

Since some side effects tend to occur early in the course of treatment (e.g., rash), others late (e.g., osteoporosis), and others relatively evenly over time (GI hemorrhage), toxicity index values and standard errors are computed for 6-month periods of exposure (for the first 6 month period, second and third 6 month periods, etc.) for each drug and then combined. Raw scores are statistically adjusted by standardization as described below, using age, race, sex, duration, disease severity, and new start versus continuing therapy, comorbidity, presence of concurrent therapy, and other variables. Sensitivity analyses using several alternative weighting systems have been performed, and yield similar results.

Regression trees were used to develop strata for standardization, following the procedures of Bloch and Segal and using the computer program CART[6]. Regression trees were constructed for the first period, the second and third periods, periods 4 through 8, and periods 9 and above. A similar list of classifying variables was obtained for each set of periods. Each regression tree first split either on disability or comorbidity, and these variables appeared in all trees. Additional variables selected included age, duration, and number of swollen joints. Since we are not able to distinguish patients who were first prescribed a drug in the first observation period from those who had begun taking the drug before our observation, we repeated the standardized analyses on those who were not taking the drug at first observation period but subsequently began treatment. The variables appearing in these regression trees were closely similar to those already described. Strata developed by these techniques were used to develop standardized toxicity index scores. The hypothesis for analysis is that there is no difference in standardized toxicity index scores between drugs; thus, the alternative hypotheses are 2-sided. The test statistic is based on the normal approximation to the distribution of differences between the standardized toxicity index values.

RESULTS

Table 1 presents the toxicity indices for NSAID[4,7]. There is a range of 3 to 4 fold in toxicity between the most toxic and least toxic NSAID. Of prescription NSAID, when employed for RA, salsalate, ibuprofen, and naproxen are the least toxic, and ketoprofen, tolmetin, meclofenamate, and indomethacin are the most toxic.

Table 2 shows a similar display of relative toxicity for the DMARD[5,7]. A wide range of toxicities is seen here, ranging from a benign profile and score for hydroxychloroquine to higher values for methotrexate, azathioprine, auranofin, and the reference drug prednisone. Comparing data for NSAID (Table 1) with DMARD (Table 2), interesting conclusions emerge. All data were obtained from the same patients over the same time period, with statistical adjustment for time taking the drug and for characteristics of patients receiving specific therapies. The overlap of toxicities between NSAID and DMARD is much more pronounced than are the differences, both before and after statistical adjustment. Hydroxychloroquine would be a very nontoxic NSAID, while the most toxic NSAID exhibit similar overall toxicity as DMARD such as methotrexate and azathioprine. Each of the results reported above holds when the overall drug experience or only new starts are analyzed, and holds at each of 5 clinical centers with greatly differing

Table 1. *Relative toxicity of NSAID: data from 5 ARAMIS databank centers; standardized toxicity index scores*

Drug	Number of Courses	Mean ± SE	Rank
Salsalate	121	1.28 ± 0.34	1
Ibuprofen	503	1.94 ± 0.43	2
Naproxen	939	2.17 ± 0.23	3
Sulindac	511	2.24 ± 0.39	4
Piroxicam	790	2.52 ± 0.23	5
Fenoprofen	161	2.95 ± 0.77	6
Ketoprofen	190	3.45 ± 1.07	7
Meclofenamate	157	3.86 ± 0.66	8
Tolmetin	215	3.96 ± 0.74	9
Indomethacin	386	3.99 ± 0.58	10

Table 2. *Relative toxicity of DMARD: data from 5 ARAMIS databank centers; standardized toxicity index scores*

Drug	Number of Courses	Mean ± SE	Rank	Hospitalization Component
OH-Chloroquine	639	1.38 ± 0.15	1	0.00
Intramuscular gold	659	2.27 ± 0.17	2	0.15
D-penicillamine	496	3.38 ± 0.36	3	0.99
Methotrexate	660	3.82 ± 0.35	4	1.02
Azathioprine	190	3.92 ± 0.39	5	0.83
Auranofin	409	5.25 ± 0.32	6	0.00

patient selection characteristics. Rank orders also persist after extreme changes in weighting are employed, from totally unweighted to using squared weights.

In the comparative DMARD experience reported above, 2 drugs, hydroxychloroquine and auranofin, were not associated with serious toxicity requiring hospitalization or resulting in death, although auranofin had a very frequent occurrence of annoying but reversible side effects, principally diarrhea. Additionally, when a toxicity index for NSAID is constructed from only GI toxicity, rank orders are essentially identical.

DISCUSSION

Toxicity index information is important and provides unique insights, in part because of the methodologic difficulties. A recent Newcastle conference on NSAID toxicity brought together 12 groups with data on comparative toxicity (Henry, unpublished data). Despite differences in datasets, drugs studied, methods, and assumptions, the data were congruent. For example, all groups had studied ibuprofen, naproxen, and piroxicam, and in every study ibuprofen was least toxic, naproxen middle, and piroxicam most toxic. Moreover, the range of toxicity, about 3 fold over all drugs, was similar. Thus, data appear valid.

On the other hand, aspirin was unexpectedly nontoxic in our data, leading to further study. The relative benignity of aspirin was found to be accounted for by a low relative dose (2665 mgm/day) and by frequent use of coated aspirin preparations in actual clinical experience[8]. Thus, actual experience may be different from that predicted in preapproval clinical trials.

Remaining caveats and requirements are numerous. To avoid hidden effects of underlying assumptions, raw data must be presented when data are published so that others may recompute with different assumptions. Better weighting systems are desirable; although they are unlikely to change results very much they may better satisfy purists. We have now changed our statistical adjustment procedures for differing patient characteristics when taking different drugs, using general linear models for analysis of covariance; while numbers change, actual results are closely similar.

The standard procedure calculates the crude toxicity index for each drug as the total toxicity units accrued for that drug divided by the total years of exposure to the drug, then statistically adjusts the crude index for patient characteristics including age, sex, time on drug, previous side effects, disease duration, educational level, disability level, concurrent medications, and others, using analysis of covariance with SAS software (SAS Inc, Cary, NC). The new procedure calculates a toxicity index score for each patient consisting of the toxicity units for a drug divided by the length of time taking drug, averages these scores across patients, then statistically adjusts scores similarly for differing patient characteristics.

Nevertheless, most toxicity studies will be observational if data are to be sufficient, and if patients receiving different drugs are quite different, statistical adjustment is likely to be incomplete. This does not appear a problem with NSAID, but may be with DMARD. Death is the ultimate side effect, but occurs too seldom to be included in an index even if the weighting problems could be solved. Compliance and relative dosage present additional problems, and dose/side effect relationships are needed for all drugs studied.

Yet the rewards are many. The stage is set for quantitative study of toxic–therapeutic ratios (e.g., toxicity index/disability index changes) that can further inform clinical choice. From a health policy perspective, increased use of less toxic NSAID and increased avoidance of the most toxic can importantly reduce aggregate NSAID toxicity, perhaps by one-half or more[4,7,9-11].

REFERENCES

1. Fries JF: Toward an understanding of patient outcome measurement. *Arthritis Rheum 1983;*26:697–704.
2. Fries JF, Singh G, Lenert L, Furst DE: Aspirin, hydroxychloroquine, and hepatic enzyme abnormalities with methotrexate in rheumatoid arthritis. *Arthritis Rheum 1990;*33:1611–9.
3. 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.
4. Fries JF, Williams CA, Bloch DA: The relative toxicity of nonsteroidal antiinflammatory drugs. *Arthritis Rheum 1991;*34:1353–60.
5. Fries JF, Williams CA, Bloch DA: The relative toxicity of disease modifying antiinflammatory drugs. *Arthritis Rheum 1991;*34:1353–60.
6. Breiman L, Freidman J, Olshen R, Stone C: *Classification and Regression Trees.* Belmont, CA: Wasworth, 1984:93–129.
7. Fries JF, Williams CA, Ramey DR, Bloch DA: The relative toxicity of alternative therapies for rheumatoid arthritis: Implications for the therapeutic progression. *Semin Arthritis Rheum 1993;*23:68–73.
8. Fries JF, Ramey DR, Singh G, Morfeld D, Bloch DA, Raynauld JP: A reevaluation of aspirin therapy in rheumatoid arthritis (RA). *Arch Intern Med 1993;*153:2465–71.
9. Gabriel SE, Bombardier C: NSAID induced ulcers: An emerging epidemic? *J Rheumatol 1990;*17:1–4.
10. Fries JF, Miller SR, Spitz PW, Williams CA, Hubert HB, Bloch DA: Toward an epidemiology of gastropathy associated with nonsteroidal antiinflammatory drug use. *Gastroenterology 1989;*96:647–55.
11. Fries JF, Williams CA, Bloch DA, Michel BA: Nonsteroidal antiinflammatory drug-associated gastropathy: Incidence and risk factor models. *Am J Med 1991;*91:213–22.