Outcome Assessment in the Rheumatic Diseases*

James F. FRIES**

Diagnosis, prognosis, and therapy are the three great clinical sciences. Yet, prognosis suffers from a peculiar and profound neglect. Its principles are seldom taught. Major textbooks of medicine or rheumatology may not even list the term in the index.^{2,3)} The very term "prognosis" lacks definition; it is not entirely clear what one does when one "gives a prognosis."

A prognosis is a prediction. It is a prediction of an expected future clinical outcome. Like all predictions, it is inherently probabilistic and inherently uncertain. Prognosis takes place in the face of changing therapy and possibly in the face of a changing natural history. Despite these problems, the ability to estimate prognosis is essential to rational clinical decision-making. A rational clinical decision must be made by estimating the outcomes which would occur under two or more alternative types of management, and then selecting that alternative which is most likely to lead to the best patient outcome. If outcomes for different alternatives cannot be estimated accurately or are not forecast optimally, then the clinical decision is reduced to a whim.

Four Special Considerations

First, prognosis is primarily for the individual, not the group. While overall prognosis for a group of patients with, say, rheumatoid arthritis is interesting and important in many ways, it is not directly relevant to the clinical situation. The clinical decision has to be based on individual prognostic

factors. There is mild disease and there is severe; there is early disease and late; disease happens in young and old; men and women; and in different racial groups. Certain constellations of symptoms and outcomes cluster together. The presence of some findings nearly excludes certain future events, and may nearly ensure others. On the other hand, prognosis cannot be individualized to the level of the anecdote. We need experience with certain numbers of patients in order to be confident that particular outcomes are likely to follow. Thus, prognostic predictions require stratification by different variables, such as those above, and use of any other characteristics that might serve to define a patient as a member of a reasonably small and relatively homogeneous subgroup.

Second, prognosis must be quantified as much as possible to be useful. "Good prognosis" or a "poor prognosis" are terms with very limited clinical value. Thus, data from large patient populations stored in computer data banks or from relevant studies in the literature are needed to develop numerical estimates. Life table statistics are often particularly appropriate to this task.

Third, prognosis for a chronic illness must include estimation of distant event probabilities as well as proximate. Short-term (weeks or months) or even intermediate-term (1-5 years) predictions are not sufficient. With rheumatic diseases, or for that matter any chronic diseases, emphasis upon short-term outcomes can be misleading and often is ultimately detrimental to care. A typical course of rheumatoid arthritis, diagrammed in Fig. 1, serves as an example. The patient with rheumatoid arthritis has an average 21-year disease course⁴⁾. Death occurs typically some four to twelve years prematurely, disability increases slowly through the course with the most rapid rises in the early years, discomfort fluctuates throughout the period of disease, iatrogenic effects from medicine or surgery occur sporadically, and massive economic impact is

* This work was supported in part by a grant to ARAMIS (American Rheumatism Association Medical Information System), (AM 21393) and the Stanford Arthritis Center (AM 20610). Paper prepared for Prognosis in the Rheumatic Diseases, N. Bellamy (Ed.), MTP Press Ltd., London, 1986, and presented in slightly modified form here by permission.

** Department of Medicine, Stanford University School of Medicine, Stanford, California

Fig. 1

felt, with costs generally increasing over the course of the illness. The therapeutic goal, to improve outcome, consists of reducing the stippled areas in Fig. 1, being interested at once in the area under the curve and at the final point at the end of illness.⁵⁾ Presentation of prognosis in terms of "disability-years" or lifetime medical cost is thus appropriate.

Fourth, prognosis must proceed from a specific point in the course, represented by the parameters of the individual patient. For describing the "natural" history of a disease, an "inception cohort" of individuals followed from the onset of symptoms is appropriate. In the clinical setting, however, it is more usual for prognostic questions to arise later in the disease course; indeed, the need for prognostic estimates occurs with every clinical decision throughout the course of illness. For these questions the past is past, and it is the future from that point which must be predicted.

Fifth, prognosis must encompass all meaningful outcomes. With the rheumatic diseases in particular, one cannot focus solely on mortality rates. Indeed, many of the conditions discussed in this book have negligible effects upon mortality. Nor is it ethically or practically possible to derive a single number to represent prognosis. To do so would require setting a cost equal to a year of life, a year of life equal to a degree of disability, a degree of disability equal to an amount of pain, and so forth. In prediction of rheumatic disease outcomes there are frequently conflicts between outcomes in

different dimensions and erroneous decisions may be made if all dimensions are not considered. For example, a drug might reduce disability but carry a certain risk of death. A strong analgesic might be quite successful at reducing pain but might impede function.

In this paper, I define "outcome" as consisting of the five dimensions of Fig. 1 and attempt to define these dimensions so that they are mutually exclusive yet collectively exhaustive. Thus, discomfort includes concepts such as depression or patient dissatisfaction as well as physical pain or fatigue, iatrogenic effects can be "medical" or "surgical," and economic impacts "direct" or "indirect."

Assessment of Clinical Outcomes

The primary task of medicine is to make people better. Unfortunately, this straightforward directive contains hidden pitfalls, some discussed above. We must know what "better" is, and we must be able to tell when we have achieved it. With short-term illness such as appendicitis or pneumococcal pneumonia, the answers are almost self-evident. But with the rise of chronic diseases as the dominant health threats in developed nations, the answers have become far more complex.

Based on social science research techniques we and others have been able to develop instruments that reflect patient preferences: to be alive, to have full function, to be free of symptoms, to be free of side effects, and to remain solvent.⁶⁾ It has been possible to develop instruments that obtain direct quantitative assessments in patient-oriented terms, such as disability, rather than in doctor-oriented terms, such as reduction of sedimentation rate.⁷⁾ There has been gradual acceptance of a framework within which these assessments can be made. This framework, elaborated above and demonstrated in Fig. 1, encompasses five disease dimensions (mortality, and four measures of morbidity—disability, discomfort, iatrogenic adverse reactions, and economic impact).⁸⁾

Defining the medical task as that of improving long-term cumulative health outcomes implies a new emphasis in our investigative approaches, medical literature, and education processes. Changes are needed, conceptually and practically.

Process versus Outcome

An important first step has been to move from measures of "process" to measures of "outcome." "Process" measures include such variables as sedi-

mentation rate, latex titre, or level of antibody to DNA. Process measures have no intrinsic value to the patient; they are of clinical value only to the extent that they serve as accurate surrogates for "outcome" measurements. Thus, if the sedimentation rate accurately quantitated "fatigue," and "fatigue" is part of patient discomfort, and patient discomfort is an outcome, then the sedimentation rate would have value as a surrogate outcome measurement, conveniently ascertained. But, to the extent that these traditional tests are only loosely correlated with ultimate outcome, they represent false endpoints. The equivalent folk saying is that "the operation was a success, but the patient died." Good process, but unfortunate outcome.

Hard versus Soft Data

An important second step is to rethink our use of the terms "hard data" and "soft data." "Hard data" have generally been thought to be those obtained from a laboratory test, such as hemoglobin values. Hard data are easily quantitated, continuous on a long scale, relatively reproducible, and reassuringly numeric. Because variables such as "hemoglobin" have these attributes, we have come to believe that so do all numbers that emerge from the laboratory such as serum complement values, sedimentation rates, or levels of antibody to DNA. We have tended to use the word "hard" for data that are essentially "processes" variables as discussed above. More properly, we should have reserved use of the term "hard" to represent data with good measurement characteristics of reliability and validity.

A dilemma for the clinician in this old formulation was that the variables which were "hard" were largely irrelevant to patient status, while those which were clearly relevant (as with quality of life measures) were considered "soft."

Fortunately for the new science of outcome assessment, the distinction between "hard" and "soft" data is itself a "soft" one. The carefully developed new measures of patient outcome (such as the Health Assessment Questionnaire (HAQ) shown in Fig. 2 or 3^{5,6,8)} or the Arthritis Impact Measurement Scales (AIMS)⁷⁾ provide data that have far more robust measurement characteristics than many laboratory data.^{9,10)}

Had we applied tests of measurement characteristics to laboratory values, such as erythrocyte sedimentation rates, we would have become disillusion-

ed with these measures somewhat earlier. Essentially, data are not available for sedimentation rates with split samples sent to separate laboratories or for repeat measurements in the same individual on successive days; one suspects that we would not be happy with the results of such evaluations. It has seldom been asked whether the sedimentation rate is an important endpoint (face validity) or whether it correlates well with measures which are important (convergent validity). When tests such as level of antibody to DNA or serum complement have been subjected to such examination, their measurement characteristics have been found wanting.⁹⁾ Large laboratory errors, 20 per cent or more, are frequently encountered.⁹⁾ In contrast, test-retest correlations of the new disability scales, for example, may be as high as 0.98.¹¹⁻¹⁴⁾

Happily, these facts are increasingly recognized. A recent survey of clinical-trial experts revealed the five most valued endpoints for rheumatoid arthritis to be global assessment by the physician, joint count, self-report of pain, self-report of morning stiffness, and grip strength—in that order.¹⁵⁾ None of these is a "hard" measure in the traditional sense.

Thus, the terms "hard" and "soft" must now be used to refer to the actual measurement characteristics of a variable rather than to the point of origin of the data. On every one of the central characteristics that define "hard" data (precision, quantifiability, reproducibility, inter-observer variation, face validity, convergent validity), most laboratory test results used in evaluating an arthritis patient are far inferior to the questionnaire outcome assessment instruments.

Short- versus Long-Term Assessment

Long-term observations (over 5 years) are clearly needed for evaluation of long-term diseases. Health status at any one time represents a combination of present disease activity and accumulated past damage. Obviously, an optimal therapy must be one that minimizes the impact of the disease over time, neither disregarding the early symptoms nor neglecting the long-term difficulties. A clinician managing a patient with rheumatoid arthritis is frequently aided by a long-term flow sheet that displays sedimentation rate, grip strength, hemoglobin, latex fixation titer, and other values over time.

From the discussion above, it is clear that the physician will be even better served by flow sheets that include assessment of outcome status at inter-

HEALTH ASSESSMENT QUESTIONNAIRE					PHASE <u>XV</u>
Name _____					TOR.# _____
Date _____					QUESTDAT _____
<div style="border: 1px solid black; padding: 5px;"> In this section we are interested in learning how your illness affects your ability to function in daily life. Please feel free to add any comments on the back of this page. </div>					HAQADMIN _____
• Please check the one response which best describes your usual abilities OVER THE PAST WEEK:					QUESTYPE <u>2</u>
	Without ANY Difficulty	With SOME Difficulty	With MUCH Difficulty	UNABLE To Do	STUDSTAT _____
DRESSING & GROOMING					
Are you able to:					
-Dress yourself, including tying shoelaces and doing buttons? _____					
-Shampoo your hair? _____					
ARISING					
Are you able to:					
-Stand up from an armless straight chair? _____					
-Get in and out of bed? _____					
EATING					
Are you able to:					
-Cut your meat? _____					
-Lift a full cup or glass to your mouth? _____					
-Open a new milk carton? _____					
WALKING					
Are you able to:					
-Walk outdoors on flat ground? _____					
-Climb up five steps? _____					
• Please check any AIDS OR DEVICES that you usually use for any of these activities:					
_____ Cane		_____ Devices Used for Dressing (button hook, zipper pull, long-handled shoe horn, etc.)			
_____ Walker		_____ Built Up or Special Utensils			
_____ Crutches		_____ Special or Built Up Chair			
_____ Wheelchair		_____ Other (Specify: _____)			
• Please check any categories for which you usually need HELP FROM ANOTHER PERSON:					
_____ Dressing and Grooming		_____ Eating			
_____ Arising		_____ Walking			
(REV. 8/88)					RISE _____
					DRESSING _____
					EATING _____
					WALK _____
- 1 - Stanford Arthritis Center Disability and Discomfort Scales, 1981					

Fig. 2

vals, perhaps every six months, suitable for evaluation of either intermediate-term (1-5 years) and long-term (over 5 years) outcomes. If we wish to reduce the cumulative impact of disease, then we must be able to define and measure cumulative impact. Chart forms meeting these specifications

have been developed by the American Rheumatism Association Medical Information System (ARAMIS) and are available from ARAMIS offices at 701 Welch Road, Suite 3301, Palo Alto, CA 94304.

Univariate versus Multivariate Issues

A final basic consideration refers to the disturb-

● Please check the one response which best describes your usual abilities OVER THE PAST WEEK:

	Without ANY Difficulty	With SOME Difficulty	With MUCH Difficulty	UNABLE To Do	
HYGIENE					
Are you able to:					
—Wash and dry your entire body?	_____	_____	_____	_____	HYGIENE
—Take a tub bath?	_____	_____	_____	_____	
—Get on and off the toilet?	_____	_____	_____	_____	
REACH					
Are you able to:					
—Reach and get down a 5 pound object (such as a bag of sugar) from just above your head?	_____	_____	_____	_____	REACH
—Bend down to pick up clothing from the floor?	_____	_____	_____	_____	
GRIP					
Are you able to:					
—Open car doors?	_____	_____	_____	_____	GRIP_ARA
—Open jars which have been previously opened?	_____	_____	_____	_____	
—Turn faucets on and off?	_____	_____	_____	_____	
ACTIVITIES					
Are you able to:					
—Run errands and shop?	_____	_____	_____	_____	ACTIVITY
—Get in and out of a car?	_____	_____	_____	_____	
—Do chores such as vacuuming or yardwork?	_____	_____	_____	_____	
● Please check any AIDS OR DEVICES that you usually use for any of these activities:					
_____ Raised Toilet Seat	_____ Bathtub Bar				
_____ Bathtub Seat	_____ Long-Handled Appliances for Reach				
_____ Jar Opener (for jars previously opened)	_____ Long-Handled Appliances in Bathroom				
Other (Specify _____)					
● Please check any categories for which you usually need HELP FROM ANOTHER PERSON:					
_____ Hygiene	_____ Gripping and Opening Things				
_____ Reach	_____ Errands and Chores				
We are also interested in learning whether or not you are affected by pain because of your illness.					
● How much pain have you had because of your illness IN THE PAST WEEK ?					
PLACE A MARK ON THE LINE TO INDICATE THE SEVERITY OF THE PAIN					
NO PAIN					VERY SEVERE PAIN
0	_____				100

(REV. 8/88) — 2 — Stanford Arthritis Center Disability and Discomfort Scales, 1981

Fig. 3

ingly multivariate world of chronic illness. Non-biologic inputs clearly effect outcome. Yet, our medical system teaches biologic mechanisms almost exclusively. In the rheumatic diseases, patient motivation and compliance, educational level, socioeconomic status, availability of community resources,

payment mechanisms, public and patient education, family support, level of depression, and other factors may be more important determinants of disability than the biologic activity of the disease or the effect of medication. These multiple influences on patient outcome require a broader model of health.

Moreover, a multivariate model largely negates the traditional experimental setting. The univariate experiment that has formed the basis of biologic science and its clinical analog, the randomized controlled clinical trial, are of only limited use in approaching problems of this complexity. Experiments necessarily study single factors in isolation. Multivariate problems involving multiple "risk factors" and many interactions are not susceptible to such reductionist analysis. The broad concept of patient outcome requires that clinical knowledge be developed within a more complex framework.

The Health Assessment Questionnaire (HAQ)

The Stanford Arthritis Center Health Assessment Questionnaire (HAQ) is based on the above concepts. The five outcome dimensions are measured. Development and testing of these instruments has required more than seven years, and the resulting questionnaires have been administered in more than 100 settings, with computer-stored data available on over 100,000 administrations on 20,000 patients. The instrument is now a standard part of the ARAMIS (American Rheumatism Association Medical Information System) Outcome Assessment Protocols and has been modified for use in the United States Health and Nutrition survey (HANES I Follow-up, HANES II and HANES III). Administration of the two dimensions (disability and pain) shown in Figs. 2 and 3 takes five to eight minutes with manual scoring possible in one minute.

Over 30 validation studies for the Health Assessment Questionnaire have been reported (c.f. Refs. 5, 6, 8, 10-19). Responses are similar whether the instrument is self-completed, whether administered by a nurse-assessor or physician at interview, or whether values are assigned after performance evaluation in the patient's home. Rheumatoid arthritis patients show a mean score of approximately 1.2 on the scale of 0 to 3 for disability and an increase per year of disease of approximately 0.08 units.⁵⁾ Disability in rheumatoid arthritis has been shown to increase most rapidly in the first ten to twelve years of disease, then at a slower rate over the remainder of the course.¹¹⁾ Values are similar for very different populations, including those in such places as Stanford, Phoenix, Wichita, and Saskatoon.¹¹⁾

In contrast, osteoarthritis shows mean disability

values of only 0.4, and a rate of increase of disability that is less than one half that of the rheumatoid patient.⁵⁾ In fact, disability in the osteoarthritic patient is more highly correlated with overall age than with duration of disease. Initial values of the disability index predict four-year outcome more accurately than do measurements of morning stiffness, number of involved joints, sedimentation rate, or latex fixation titre.^{11, 17)} Medical costs for those in the fourth quartile of disability are more than nine times those in the first quartile.¹⁸⁾

Recent Progress in Outcome Assessment

Recent progress in assessing patient outcome has been made in six areas, illustrated by the HAQ and by other well-developed instruments subjected to similar trial (most particularly the AIMS scale from Boston University). First, the conceptual framework of multidimensional outcome has been established. Comparison of the HAQ and AIMS has shown natural clustering of items around dimensions such as disability. For example, physical disability scores on HAQ and AIMS have high inter-instrument correlation coefficients of 0.91. Correlations between components of the same dimensions are similarly high, usually ranging from 0.6 to 0.8. In contrast, correlations between dimensions, even in the same instrument, are much weaker, usually less than 0.3. Discomfort and disability are only weakly correlated, as are disability and cost, discomfort and approaching death, and other correlations between dimensions.¹⁰⁾ These results demonstrate that outcome may be described with a small number of discrete dimensions.

Second, the new measures are valid. In addition to validations discussed above, patient costs are abstracted from hospital and physician records and compared with services self-reported by patients. Computer tapes of insurance carriers can be matched with costs reported by patients. Validation studies by protocol are repeated in multiple locations to ensure that they can be generalized. A standard group of validation tests are repeated in every study to ensure that previous results can be confirmed with a random subsample of patients. The efforts required to perform these validations has been large, and the cost considerable, but the resulting instruments provide quality data worthy of the effort.

Third, efficiency has improved. An information-science aphorism notes that the quality of data ob-

tained is inversely related to the quantity requested. Certainly, in clinical practice, time limits the effort available. Procedures to reduce redundancy have allowed construction of shorter instruments, thus improving data-quality. Disability may now be assessed in a practical manner, at lower cost than that of a single laboratory test.⁸⁾

Fourth, relevance has been demonstrated. Until recently, evidence was scant that patient values actually were being captured by these instruments. In the comparative study between the AIMS and HAQ, rheumatoid arthritis patients were asked to indicate the overall impact of their arthritis on a scale of 0 to 100. The disability and discomfort dimensions of each instrument explained about two thirds of the total variance represented by patient values, and disability made up approximately two-thirds of the explained variance.¹⁰⁾ This is an extraordinarily strong result, and it indicates a high concordance between what is measured by these instruments and the personal values of patients with rheumatoid arthritis.

Fifth, normal and average values for particular conditions are now known.⁵⁾ It is difficult to use a variable clinically if you do not know what "normal" is and what various degrees of "abnormal" mean. Now, results can be compared to established multicenter figures.

Finally, these outcome measures have been used in randomized controlled clinical trials and have been found to be at least as sensitive than traditional measures in determining differences between groups of patients given active drug and placebo.¹⁹⁾ Thus, the instruments are sensitive even to small and short-term changes.

Recommendations

In clinical practice, patient records should include flow sheets delineating the trends and tempo of the underlying disease. Health outcome measurements are an essential part of such flow sheets and may well prove the most important part of such serial observations. Such data can optimally prepare the physician for use of the information which follows in this volume.

Controlled clinical trials should also include these measures of quality of life. They are both more meaningful and more sensitive than traditional measures.

The development of outcome assessment is at an exciting stage. The new concepts challenge tradi-

tional views about the role of medical care, the importance of strictly biologic inputs, and the role of the patient in deciding what should or should not be done. These developing methodologies broaden our conception of outcome and provide a substrate for development of a science of prognosis.

References

- 1) Fries, J. F., Ehrlich, G. E. (Eds.): *Prognosis: Contemporary Outcomes of Disease*, Addison-Wesley Publishing Company, Menlo Park, California (1985).
- 2) Kelley, W. N. et al.: *Textbook of Rheumatology*, 2nd ed, W. B. Saunders, Philadelphia (1985).
- 3) Wyngarten, J. B. and Smith, L. B. (Eds.): *Cecil. Textbook of Medicine*, W. B. Saunders, Philadelphia (1988).
- 4) Mitchell, D. M., Spitz, P. W., Young, D. Y. et al.: *Survival, Prognosis, and Causes of Death in Rheumatoid Arthritis. Arthritis Rheum.*, **29** (6) : 706-714 (1986).
- 5) Fries, J. F.: *The Assessment of Disability: From First to Future Principles. Br. J. Rheumatol.*, **22** (suppl): 43-48 (1983).
- 6) Fries, J. F., Spitz, P., Kraines, R. G. et al.: *Measurement of Patient Outcome in Arthritis. Arthritis Rheum.*, **23** : 137-145 (1980).
- 7) Meenan, R. F., Gertman, P. M. and Mason, J. H.: *Measuring Health Status in Arthritis: The Arthritis Impact Measurement Scales Arthritis Rheum.*, **23** : 146-152 (1980).
- 8) Fries, J. E.: *Toward an Understanding of Patient Outcome Measurement. Arthritis Rheum.*, **26** : 697-704 (1983).
- 9) Feigenbaum, P. A., Medsger, T. A., Jr., Kraines, R. G. et al.: *The Variability of Immunologic Laboratory Tests. J. Rheumatol.*, **9** : 408-414 (1982).
- 10) Fries, J. F., Spitz, P. W. and Young, D. Y.: *The Dimensions of Health Outcomes: The Health Assessment Questionnaire, Disability and Pain Scales. J. Rheumatol.*, **9** : 789-793 (1982).
- 11) Mitchell, D. M., Spitz, P. W., Young, D. Y. et al.: *Survival, Prognosis, and Causes of Death in Rheumatoid Arthritis. Arthritis Rheum.*, **29** (6) : 706-714 (1986).
- 12) Kirwan, J. R. and Reeback, J. S.: *Stanford Health Assessment Questionnaire modified to assess disability in British patients with rheumatoid arthritis. Br. J. Rheumatol.*, **25** : 206-209 (1986).
- 13) Pincus, T., Summey, J. A., Soraci, S. A. et al.: *Assessment of patient satisfaction in activities of daily living using a modified Stanford Health Assessment Questionnaire. Arthritis Rheum.*

- 26: 1346-1353 (1983).
- 14) Sullivan, F. M., Eagers, R. C., Lynch, K. et al.: Assessment of disability caused by rheumatic diseases in general practice. *Ann. Rheum. Dis.*, **46**: 598-600 (1987).
- 15) Bombardier, C., Tugwell, P., Sinclair, A. et al.: Preference for endpoint measures in clinical trials: results of structured workshops. *J. Rheumatol.*, **9**: 798-801 (1982).
- 16) Peck, J. R., Ward, J. R., Smith, T. W. et al.: Convergent/discriminant validity of the HAQ disability index in rheumatoid arthritis using a multitrait-multimethod matrix. *Arthritis Rheum.*, **30**: S193 (abstract) (1987).
- 17) Wolfe, F., Kleinheksel, S. M., Cathey, M. A. et al.: The clinical value of the HAQ disability index in patients with rheumatoid arthritis. *J. Rheumatol.* (in press).
- 18) Nevitt, M. C., Yelin, E. H., Henke, C. J. et al.: Risk factors for hospitalization and surgery in patients with rheumatoid arthritis: Implications for capitated medical payment. *Ann. Intern. Med.*, **105**: 421-428 (1986).
- 19) Bombardier, C., Ware, J., Russell, I. J. et al.: Auranofin therapy and quality of life in patients with arthritis, results of a multicenter trial. *Am. J. Med.*, **81**: 565-578 (1986).