

THE  
UNIVERSITY  
OF RHODE ISLAND

University of Rhode Island  
DigitalCommons@URI

---

Psychology Faculty Publications

Psychology

---

1986

# Systems for multivariate monitoring of behavioral status over time

Robert L. Stout

John Stevenson

*University of Rhode Island*, [jsteve.uri@gmail.com](mailto:jsteve.uri@gmail.com)

*See next page for additional authors*

Follow this and additional works at: [https://digitalcommons.uri.edu/psy\\_facpubs](https://digitalcommons.uri.edu/psy_facpubs)

Terms of Use

All rights reserved under copyright.

---

## Citation/Publisher Attribution

Stout, R.L., Stevenson, J., Faraone, S. et al. Behavior Research Methods, Instruments, & Computers (1986) 18: 141. <https://doi.org/10.3758/BF03201013>

Available at: <http://dx.doi.org/10.3758/BF03201013>

This Article is brought to you for free and open access by the Psychology at DigitalCommons@URI. It has been accepted for inclusion in Psychology Faculty Publications by an authorized administrator of DigitalCommons@URI. For more information, please contact [digitalcommons@etal.uri.edu](mailto:digitalcommons@etal.uri.edu).

---

**Authors**

Robert L. Stout, John Stevenson, Steve Faraone, and John Simpson

# Systems for multivariate monitoring of behavioral status over time

ROBERT L. STOUT

*Brown University, Providence, Rhode Island  
and Butler Hospital, Providence, Rhode Island*

JOHN STEVENSON

*University of Rhode Island, Kingston, Rhode Island*

STEVE FARAONE and JOHN SIMPSON

*Harvard University, Cambridge, Massachusetts  
and Brockton/West Roxbury VA Medical Center, Brockton, Massachusetts*

Decision-theoretic criteria are presented for optimizing the information gathered from a series of interviews over time. It is shown that the optimum interviewing strategy depends strongly on assumptions about the covariation of behavior over time. Standard interviewing strategies, including the major-problem/target-complaints approach, are optimal only under extreme assumptions about behavior. An interviewing strategy based on dynamic programming is presented that will provide optimal information return from a series of interviews under assumptions that are realistic for mental health applications. A system using this approach can tailor its interviewing strategy to adapt to differences in interview content, item importance, and individual response patterns, selecting the optimally informative questions to ask each subject at each point in time. Simulation results show that this approach achieves a 34% reduction in the false negatives obtained with the major-problem/target-complaints method, and, depending on the acceptable error rate, a reduction of 47% or more in the questions that are needed in standard interviewing.

Recent research on psychopathology suggests that many questions about the classification of psychiatric disorders, and their prognoses, can be answered only by monitoring changes in symptom patterns over time (Grossman, Harrow, Fudala, & Meltzer, 1984; Stout, 1984). The course of mental illnesses is usually far from uniform; patients' symptoms can worsen and remit within a few days, and phenomena such as symptom substitution can complicate the picture considerably (Cahoon, 1968; Reider, 1976).

Assessing the status of a patient even at a single point in time is a considerable task. Mental status interviews vary in their content and size from such short instruments as the Brief Psychiatric Rating Scale (21 items; Overall & Gorman, 1962) to extensive diagnostic interviews, such as the DIS (Robins, Helzer, Croughan, & Ratcliff, 1981), containing many hundreds of items. The complexities of

psychiatric assessment make monitoring *changes* in a patient's symptom status over time a formidable technical challenge; not only are there many variables to be assessed, but these variables must be measured at many points in time. Researchers who wish to do repeated assessments have found themselves forced to make painful compromises with respect to the number of interviews per patient, the number of questions per interview, and/or the quality of the data gathered because of the limitations imposed by time, cost, and human tolerance.

The usual practice in follow-up interviewing today is to ask each subject the same questions in each interview. Repeated interviewing by this technique imposes severe burdens on subjects, who must answer repetitious questions, many of which are not relevant to them personally, interview after interview. For example, it is often important to know whether a person is abusing narcotics, but if a person has never abused narcotics, he/she rapidly wearies of being asked about narcotic abuse month after month. Patients are also frustrated by the fact that even though the interviewers ask many irrelevant questions, they often fail to ask about crucial aspects, important in the patients' individual lives, because these topics have been omitted to keep the interviews tolerably short.

Researchers are aware that a burdensome interview schedule will lead to a reduction in subject compliance and stereotyped, inappropriate responses. To reduce the burden on patients, researchers have been forced to make

---

The research on which this article is based was supported in part by Biomedical Research Support Grant RR05817 to Butler Hospital. The PAS data used in this research were produced under National Institute of Mental Health Grant MH 26012, "Problems as Predictors of Treatment and Outcome," to Richard Longabaugh and Robert Stout. R. L. Stout is associated with the Department of Psychiatry and Human Behavior at Brown University; J. Stevenson is in the Department of Psychology at the University of Rhode Island; and S. Faraone and J. Simpson are in the Department of Psychiatry at Harvard University. Requests for reprints should be sent to R. L. Stout at Butler Hospital, 345 Blackstone Boulevard, Providence, RI 02906.

many compromises in research design. Among these compromises are the elimination of desirable but nonessential questions, a reduction in the number and/or increase in the spacing between interviews, and the acceptance of less reliable information. Reliability is often compromised by a reliance on relatively general, abstract questions when specific, concrete questions would provide data of higher validity (Angle, Ellinwood, & Carroll, 1978).

### THE PRESENT ROLE OF COMPUTER TECHNOLOGY

Computer interviews and tests are becoming increasingly accepted as tools for gathering clinical research data (Greist & Klein, 1980; Johnson, 1984). As research instruments, computer interviewing systems have several important advantages, including standardized presentation and reliable data recording and management. Furthermore, they make the subject's task easier by omitting logically irrelevant questions. Advanced systems may provide even more benefits (Stout, 1981; Stout, Kriebel, & McCullough, 1981). Nonetheless, current techniques provide only a partial answer to the problem of optimally managing information-gathering in follow-up interviewing.

Like standard human interviews or paper questionnaires, present computer interviewing systems treat each interview as an isolated entity. Branching may be employed to eliminate logically unnecessary questions, but the same questionnaire is used for all subjects at all time points.

#### The Longitudinal Perspective

Many of the limitations of standard methods can be overcome by considering what can be learned from a whole series of interviews rather than focusing on one interview at a time. Thus, what is needed is a system that can manage not just a single interview, as present interviewing programs do, but a schedule of interviews as a whole. This problem requires an approach substantially different from that used in present systems.

The problem of selecting the questions that should be asked of a particular subject at a particular point in time, given that subject's responses to earlier follow-ups, is a multistage decision problem. A decision-theoretic approach known as "dynamic programming" is the normative method for solving problems of this kind (Bellman & Dreyfus, 1962; Whittle, 1982). Dynamic programming is a branch of decision science devoted to determining optimum decision policies in situations in which decisions are made sequentially and the state of the world and/or information about the state of the world changes over time. We present below a method for applying this approach to the problem of managing follow-up interviews.

#### FORMAL STATEMENT OF THE PROBLEM

The purpose of a repeated-follow-up research study is to provide estimates of the state of an individual on  $k$  be-

havioral measures at  $p$  points in time. That is, the result of the follow-up study is a  $k \times p$  matrix,  $\mathbf{B}^*$ , of estimates of the true behavioral state of the individual,  $\mathbf{B}$ . For simplicity of exposition, we will assume below that the behavioral measures are all dichotomous (i.e.,  $\mathbf{B}$  and  $\mathbf{B}^*$  are binary matrices), and that the  $p$  time points are equally spaced. These are assumptions of convenience, and are not necessary for any of the mathematical arguments. We also assume, for convenience, that a given behavioral measure corresponds to a single interview question.

A less trivial assumption is that when an interviewee is asked a question about his/her status on a given measure, his/her answer provides certain knowledge about his/her status (i.e., the answer is perfectly informative). Thus, response errors (bias, evasion, and so on) are outside the purview of this article.

We define the information loss of an interview schedule to be:

$$U(\mathbf{B}, \mathbf{B}^*) = \sum_{i=1}^k \sum_{j=1}^p u[i, (b(i,j) - b^*(i,j))], \quad (1)$$

where  $u[i, x]$  is a nonnegative disutility function for behavioral measure  $i$ . We assume that  $u[i, x]$  is a monotonic function of  $|x|$ , but  $u$  will not be a symmetric function of  $x$  unless false negatives and false positives have equal weight. The fact that each  $u$  is a function of  $i$  as well as  $x = [b(i,j) - b^*(i,j)]$  allows us to weight some items as more important than others.

In general, the actual value of  $U$  is unknown for any given interview, since that would entail knowing with certainty the answers to questions that are not asked; however, under appropriate assumptions, we can calculate and minimize its expected value.

An interview schedule  $\mathbf{S}$  is defined to be a  $k \times p$  binary matrix such that  $s(i,j) = 1$  if question  $i$  is asked at time  $j$ , and  $s(i,j) = 0$  otherwise.  $\mathbf{S}$  may be either predetermined or, as we suggest below, determined in a stepwise manner with the questions asked at time point  $j$  determined by the answers to questions at points 1 through  $j-1$ .

The cost of an interview schedule is designated by  $C(\mathbf{S})$ .  $C(\mathbf{S})$  will depend on the number of questions asked at each interview point (it is likely to be an accelerating function of the number of questions in each interview), but it may also depend on the questions asked if some questions have larger explicit or implicit costs than others. For example, personally intrusive questions may cost no more than other questions in terms of time and interviewer effort, but they may be "expensive" in other ways.

The problem of designing an optimal interview schedule can then be formulated as the problem of finding an interview schedule  $\mathbf{S-OPT}$  such that the total loss function

$$L = E(U) + C \quad (2)$$

is minimized. In most follow-up research, external constraints, such as a fixed number of interviews and a limited duration per interview, imply that  $C(\mathbf{S})$  is bounded by a constant. Furthermore, it is almost always desirable to acquire as much information as possible, given the exist-

ing cost limits. Under these conditions, the problem of determining **S-OPT** reduces to the problem of deciding which are the most informative questions that can be asked, given the cost constraint. Hence, in the discussion below, we will concentrate on how **S-OPT** depends on  $U$ .

**DEPENDENCE OF S-OPT ON THE MODEL FOR B**

The performance of any algorithm for selecting an interview schedule depends critically on how behavior is expected to vary over time. There is no one strategy for selecting **S** that will perform optimally under all possible conditions; the problem is to find a strategy that will perform acceptably under assumptions that are consistent with empirical knowledge of human behavior. Perfect performance is impossible, but acceptable error rates may be achievable with a relatively simple algorithm.

Below, we discuss three sets of assumptions about the covariance of behavior over time and the implications of these assumptions for interviewing strategies. In discussing follow-up techniques, we assume that the behavioral state of the subjects at baseline is known with certainty through an exhaustive baseline assessment. The adequacy of these assumptions and the performance of the algorithms corresponding to each set will be assessed on follow-up data from a study of acute inpatient care.

**Case 1: Covariance Across Time is Zero**

If **B** is completely unpredictable across time (that is, that behavior at time  $t+1$  is independent of behavior at time  $t, t-1, \dots$ ), the problem of selecting **S-OPT** reduces to the problem of finding a subset of the  $k$  questions such that

$$E(U) = \sum_{i=1}^k u[i, (b(i) - b^*(i))] \quad (3)$$

is minimized, where  $b^*(i)$  is the base rate for item  $i$ , which by assumption is constant over time. In this case, **S-OPT** is a matrix in which each row is either all zeros or all ones, with the rows of ones corresponding to the optimally informative questions it is possible to present given the cost constraint and the zero rows to questions it is not possible to present. Also, **S-OPT** is the same for all subjects.

Thus, the optimum strategy is to ask every subject the same questions at each point—tailoring the interviews is impossible. This, of course, is a limiting case, but it seems to be the principal case under which the strategy of asking the same questions of every subject at every point is optimal.

**Case 2: Major-Problem/Target-Complaints Situation**

The assumption that all the behaviors that need to be assessed at follow-up can be identified at baseline is the foundation of several popular follow-up strategies. The earliest of these, Target Complaints (Battle et al., 1966) and Goal Attainment Scaling (Kiresuk & Sherman, 1968),

used clinical procedures to select a subset of behaviors, typically two or three, to be assessed at follow-up. More recently, the computerized Major Problem Rating System (McCullough, 1981; Stevenson, McCullough, Longabaugh, & Stout, 1984) has employed a standardized computer interview to identify each person's major problems at intake to treatment.

The assumption underlying these techniques is that **B** is a matrix in which any row that begins with a zero is all zeros; that is, a symptom not present at baseline can be assumed not to occur in the future (or, if it does occur, its presence is not relevant for evaluating the treatment directed at the original problems). Under this assumption, **S-OPT** is a matrix with rows of ones corresponding to the behaviors present at baseline and rows of zeros for all other behaviors. If cost constraints prevent follow-up of all the behaviors present at baseline, then a subset of the behaviors is selected. Thus, follow-up interviews under the major problem approach are individually tailored in that each subject receives an individualized follow-up interview, but the interview is the same at all follow-up points.

The primary drawback of this approach from our point of view is the likelihood of a high false negative rate. That is, if behavior  $i$  was not recorded at baseline, then  $b^*(i,j) = 0$  for all  $j$ , which could lead to a high false negative rate if new problems or symptoms appear after treatment at a significant rate.

**Case 3: Order 1 Markov Process**

If we assume that past behavior is a good but not perfect predictor of future behavior, a more realistic approach is possible. Suppose that **B** is governed by an order 1 Markov process such that the different behaviors are independent but  $b(i,j+1)$  depends on  $b(i,j)$  through a known order 1 Markov matrix **M**( $i$ ). Then for each  $j$  we can form a preliminary estimate of the behavioral state of the individual as follows.

Let **A**( $i,j$ ) be a vector of the probabilities of Markov states for item  $i$  and time  $j$ . If the subject's state is known with certainty, as we assume it is at baseline from an exhaustive interview, and as it is at any other time  $j$  when the subject has been asked item  $i$ , then **A**( $i,j$ ) is all zeros except for a one indicating the appropriate state. The number of Markov states need not be the same as the number of response states; see below for an example. If item  $i$  was last asked at time  $j_1$ , then for  $j_2 > j_1$ , the probability distribution for the subject's Markov state at time  $j_2$  is given by:

$$\hat{\mathbf{A}}(i,j_2) = \mathbf{A}'(i,j_1)\mathbf{M}^{n}(i), \quad (4)$$

where  $n=j_2-j_1$ . Thus, prior to the interview at time  $j$ , we can compute the predicted probability that the subject will respond affirmatively to item  $i$ :

$$\hat{b}(i,j) = E(b(i,j) | \hat{\mathbf{A}}(i,j)). \quad (5)$$

Using the assumptions that the  $b(i,j)$  are mutually independent and the Markov process is of order 1, the globally

optimum interviewing procedure is to select those items for which the expected information loss,

$$U(i) = E(u[i, (b(i, j) - \hat{b}(i, j))]), \quad (6)$$

is largest. Items with large  $U(i)$  values are selected until the cost constraint is exceeded.

If we are relatively certain about how the subject would answer item  $i$  at time  $j$ , then  $E(b(i, j) - \hat{b}(i, j))$  will be small and the item will probably not be asked. The longer it has been since item  $i$  was asked, the larger this term becomes and the more likely it is that the item will be scheduled. If the probable answers to two items is equally uncertain, the item with the higher value to the researcher will be scheduled.

### Simulation Results

Given a data base of questionnaires, where the same questionnaire was given to all subjects at all points, it is possible to assess how well a system using dynamic programming would have performed in selecting a tailored subset of questions for each subject. The selection of questions from the standard interview can be simulated, and actual error rates calculated.

For this purpose, we used data from a follow-up study of 695 psychiatric inpatients and day hospital patients (see Longabaugh, Stout, Kriebel, McCullough, & Bishop, in press). These patients were a random sample of admissions, covering the entire range of acute psychiatric cases except for severe organic brain disorders.

The instrument for this analysis was the Problem Appraisal Scale (PAS), which is a revised version of Endicott and Spitzer's (1972) Psychiatric Evaluation Form, a widely used instrument in psychiatric outcome research. The PAS comprises 38 items covering major psychiatric symptoms, problem behaviors, and role functioning; thus, the content is similar to that covered by most psychiatric outcome instruments. The PAS items were originally rated on a 5-point scale. To simplify the analysis, the scale was dichotomized, with *none* and *slight* in one category and *mild* to *marked* in the other. The PAS was administered three times: upon admission, upon discharge, and 1 year after admission. The study was restricted to the 322 cases that had nonmissing data at all three points.

To simulate the performance of the major problem approach (Case 2), it was assumed that all behaviors that were problematic at baseline, regardless of how many, would be asked about at 1-year follow-up, and that no other behaviors would be followed.

For the dynamic programming approach (Case 3), we estimated Markov probabilities for changes from discharge to 1 year as follows. At discharge, a person was categorized in the unaffected (U) state with respect to a PAS item if the behavior corresponding to the item was absent at both admission and discharge, in the inactive (I) state if the behavior had been present at admission but not at discharge, and in the active (A) state if the behavior was present at discharge. The rationale for using this set

of states is that for some symptoms a history of the symptom is associated with elevated risk for its later display even if it is not active at discharge. Probabilities for transitions from discharge to 1 year for 3 of the 38 items are given in Table 1.

The first of the items in Table 1, depression, is a common concomitant of many disorders, and psychiatric patients are at high risk for it after discharge even if they show no signs of it during the index admission. Thus, this is an item that the algorithm would select for frequent presentation for all cases, but especially frequently for those who have shown recent signs of it. The second symptom in Table 1, delusions, provides an example of an instance in which a history of the symptom clearly elevates the risk at follow-up (compare the Inactive with the Unaffected rows for delusions with the same rows for depression). The dynamic programming algorithm would rarely present the delusions item to Unaffected cases, would test Inactive cases more frequently, and would test Active cases most frequently. For the third symptom, bizarre behavior, the algorithm would rarely present the item to any subjects for whom the behavior was not active at the time of the preceding interview.

In this simulation, we did not use differential item utility weights, so the dynamic programming algorithm selected items solely on the basis of uncertainty.

The dependent measures in the simulation were false-positive and false-negative error rates for the dynamic programming algorithm, and the false-negative error rate for the major-problems algorithm. (By definition, there are no false positives in the major-problems approach.) For the major-problems algorithm, the number of questions asked depends on the number of items positive at baseline and, hence, is a random variable. For dynamic programming, the error rates are functions of the number of questions the (simulated) interviewing program is allowed to ask, which in actual research would be determined by the cost versus error trade-off. For example, if the interviewing program were allowed to ask one question, it would have to impute the answers to the other 37 and would be at risk for a false choice on those 37 items.

**Table 1**  
Markov Matrices for Predicting State at 1 Year  
for Selected Symptoms

Symptom	Initial State	Outcome State		
		Unaffected	Inactive	Active
Depression	Unaffected	.66	.00	.34
	Inactive	.00	.64	.36
	Active	.00	.32	.68
Delusions	Unaffected	.97	.00	.03
	Inactive	.00	.90	.10
	Active	.00	.71	.29
Bizarre behavior	Unaffected	.99	.00	.01
	Inactive	.00	.97	.03
	Active	.00	.75	.25

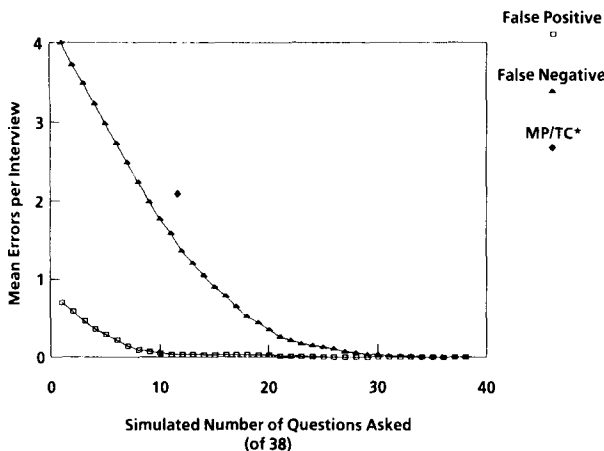


Figure 1. Error rates for simulated dynamic programming PAS interviews. \*Major-problem/target-complaints method.

The results of the simulation are displayed in Figure 1. For the major-problem/target-complaints approach, the result of the study is a single point (the diamond in the figure). The abscissa of this point is the mean number of major-problem items per patient (11.59, SEM = .26); the ordinate is the mean number of false negatives (2.09, SEM = .13).

The two curves show the false-positive/false-negative error rates for the dynamic programming algorithm as a function of  $n_q$ , the number of questions asked. The false-positive rate is quite low, less than 0.1 for  $n_q > 7$ . For false negatives, the rate drops from a peak of 3.99 for  $n_q = 1$  to below 1.0 for  $n_q > 14$ . An overall error rate of 1% (false negatives plus false positives adding to 0.38 or less) is achieved at  $n_q = 20$ . Thus, a 1% error rate is achieved by the dynamic programming algorithm with a 47% reduction in the number of questions asked relative to the standard complete interview. Lower error rates can be achieved by increasing  $n_q$  further; greater reductions in the number of questions per interview are achieved by accepting a higher error rate.

The dynamic programming algorithm achieves a lower false-negative rate than does the major-problems approach [1.37 at  $n_q = 12$  vs. 2.09 for an average of 11.59 questions  $t(321) = 4.380, p < .0001$ ] because it is more likely to inquire about symptoms, such as depression, which have a high rate at follow-up even though they are not present at baseline.

The figures from this simulation study are likely to understate the performance of the dynamic programming algorithm for many circumstances. The prediction interval, almost 1 year, is longer than the 1- to 6-month intervals used in many mental health studies. Furthermore, better performance is to be expected when the pool of items is larger (e.g., 100-300), since the items are likely to be more specific and thus more highly correlated over time.

#### Possible Adjustments to the Algorithm

There is no fundamental reason why the dynamic programming algorithm could not be generalized to take

advantage of the correlations among different items as well as the correlations over time within items. There are undoubtedly circumstances under which this would be appropriate; however, as a practical matter, in mental health outcome questionnaires, the interitem covariance typically carries much less information than the within-item correlations over time. For the PAS data, the median absolute interitem correlation at baseline is .070, whereas the median absolute correlation within items from baseline to 1 year later is .225. This situation arises, in part, by construction; to shorten the instruments, strongly overlapping items are often eliminated.

Many interviews employ broad screening questions to reduce average interview length. Questions such as, "Have you experienced any difficulty in your relationship with your spouse in the last month?" are followed, if answered positively, by more specific questions. Several studies show, however, that, with psychiatric populations, general screening questions are frequently answered negatively when more specific probes turn up positive answers (see, e.g., Hay, Hay, Angle, & Ellinwood, 1977). It should be possible, however, to follow a dynamic programming interview as described above with screening items covering the topics omitted by the algorithm.

#### CONCLUSION

Dynamic programming allows us to use rational, quantitative techniques in gathering data, and still to individualize the interview process in a way that previously has been characteristic only of human interviewers. Our results show that this technique is capable of producing data of high quality under live conditions, with a population of acute psychiatric patients interviewed about topics common to most follow-up instruments. In addition, dynamic programming offers flexibility not found in other approaches, including differential item utilities and adaptation to variations in the volatility of items.

The potential uses of dynamic programming interviewing systems clearly extend beyond mental health research. Any survey research in which repeated interviewing is done could potentially make use of this approach to maximize information return while controlling costs. Also, repeated follow-ups are important in some areas of treatment as well as in research. Systems for ongoing monitoring of behavioral status could be used as minimally intrusive early warning systems in the long-term treatment of chronic disorders.

Furthermore, we have shown how follow-up interview design and management can be based on rational, objective principles. The described techniques should further the development of scientifically designed follow-up methods.

#### REFERENCES

- ANGLE, H. V., ELLINWOOD, E. H., & CARROLL, J. (1978). Computer interview problem assessment of psychiatric patients. In F. H. Orthner (Ed.), *Proceedings of the Second Annual Symposium on Com-*

- puter Applications in Medical Care (pp. 137-148). New York: Institute of Electrical and Electronics Engineers.
- BATTLE, C., IMBER, S., HOEHN-SARIC, R., STONE, A., NASH, E., & FRANK, J. (1966). Target complaints as criteria of improvement. *American Journal of Psychotherapy*, **20**, 184-192.
- BELLMAN, R. E., & DREYFUS, S. E. (1962). *Applied dynamic programming*. Princeton, NJ: Princeton University Press.
- CAHOON, D. D. (1968). Symptom substitution and the behavioral therapies: A reappraisal. *Psychological Bulletin*, **69**, 149-156.
- ENDICOTT, J., & SPITZER, R. L. (1972). What! Another rating scale? The psychiatric evaluation form. *Journal of Nervous & Mental Disease*, **154**, 88-104.
- GREIST, J. H., & KLEIN, M. H. (1980). Computer programs for patients, clinicians, and researchers in psychiatry. In J. B. Sidowski, J. H. Johnson, & T. A. Williams (Eds.), *Technology in mental health care delivery systems* (pp. 161-181). Norwood, NJ: Ablex Publishing.
- GROSSMAN, L. S., HARROW, M., FUDALA, J. L., & MELTZER, H. Y. (1984). The longitudinal course of schizoaffective disorders. *Journal of Nervous & Mental Disease*, **172**, 140-149.
- HAY, W. M., HAY, L. R., ANGLE, H. V., & ELLINWOOD, E. H. (1977). Computerized behavioral assessment and the problem-oriented record. *International Journal of Mental Health*, **6**, 49-63.
- JOHNSON, J. H. (1984). An overview of computerized testing. In M. D. Schwartz (Ed.), *Using computers in clinical practice* (pp. 131-133). New York: Haworth Press.
- KIRESUK, T. J., & SHERMAN, R. E. (1968). Goal attainment scaling: A general method for evaluating comprehensive community mental health programs. *Community Mental Health Journal*, **6**, 443-453.
- LONGABAUGH, R., STOUT, R., KRIEBEL, G., McCULLOUGH, L., & BISHOP, D. (in press). DSM-III and clinically identified problems as a guide to treatment. *Archives of General Psychiatry*.
- McCULLOUGH, L. (1981). Systematic evaluation of the impact of computer-acquired data on psychiatric care. In H. G. Heffernan (Ed.), *Proceedings of the Fifth Annual Symposium on Computer Applications in Medical Care* (pp. 426-430). New York: IEEE Computer Society Press.
- OVERALL, J. E., & GORMAN, D. R. (1962). The Brief Psychiatric Rating Scale. *Psychological Reports*, **10**, 799-812.
- REIDER, N. (1976). Symptom substitution. *Bulletin of the Menninger Clinic*, **40**, 629-640.
- ROBINS, L. N., HELZER, J. E., CROUGHAN, J., & RATCLIFF, K. S. (1981). National Institute of Mental Health Diagnostic Interview Schedule: Its history, characteristics, and validity. *Archives of General Psychiatry*, **38**, 381-389.
- STEVENSON, J. F., McCULLOUGH, L., LONGABAUGH, R., & STOUT, R. (1984). *The development of an individualized, problem-oriented psychiatric outcome measure*. Paper presented at the Evaluation '84 Conference, San Francisco.
- STOUT, R. L. (1981). New approaches to the design of computerized interviewing and testing systems. *Behavior Research Methods & Instrumentation*, **13**, 436-442.
- STOUT, R. L. (1984, June). *Multidimensional scalogram models for changes in psychotic symptoms over time*. Paper presented at the Joint Meeting of the Classification Society and the Psychometric Society, Santa Barbara, CA.
- STOUT, R. L., KRIEBEL, G. K., & McCULLOUGH, L. (1981). Advanced methods for interactive data entry by patients. In H. G. Heffernan (Ed.), *Proceedings of the Fifth Annual Symposium on Computer Applications in Medical Care* (pp. 387-391). New York: IEEE Computer Society Press.
- WHITTLE, P. (1982). *Optimization over time: Dynamic programming and stochastic control* (2 volumes). New York: Wiley.