Medical Record Storage: A Method for Determining Level of Activity

by John F. Rockart

On the basis of return-visit patterns, a method is presented for determining which medical records should be kept in the active file and how long such records should be kept active for criteria of desired file size, desired probability of record use, desired percentage of successful file searches, and relative cost of storage in active and inactive files. The method is relevant to both physical and computerized medical record files and can incorporate any additional sets of variables that define patient groups with distinct return-visit patterns. Implementation of the system in the Lahey Clinic is described, and possibilities for further refinement of the method are discussed.

In recent years increasing attention has been centered on the medical record as a prime instrument of continuing patient care. Among many problems investigated have been the feasibility of automating portions of the medical record [1–4], the economics of microfilming the record [5], and means of keeping track of record location. Little has been written, however, on the problem of determining whether a particular record should be kept available, stored, or discarded. An effective algorithm for sorting those records with a high probability of reuse into an "active" file and those with a low probability of future use into an "inactive" file is much needed.

One major aim of such an algorithm is, of course, to determine which *physical* records should be kept in the active file. The question of physical record storage is important, since the active file space is generally prime space located in the main building of a medical facility. Access is usually fast, but the cost of the storage space is high. The inactive file is usually located in a less expensive storage space farther away from the point of care, and access time may be considerably longer. An additional need today is to determine which *computer-stored* medical records should be kept active ("on-line") for clinical or administrative use out of the many hundreds of thousands of records currently kept on file at most large clinics or hospitals [6].

This second need is perhaps more pressing than the first, because computer

---

Supported by PHS Research Grant No. 3 R18 HS00307-01 from the National Center for Health Services Research and Development, HSMHA.

Address communications and requests for reprints to John F. Rockart, Ph.D., Division of Research, Lahey Clinic Foundation, 605 Commonwealth Ave., Boston, Mass. 02215.

storage of at least some medical and administrative data on patients is increasing at a rapid pace. Moreover, the relative cost of storing a record in prime computer space—in an on-line direct-access file—is potentially far greater than the cost of storing a record in active physical storage. An effective solution to this problem will provide benefits with regard to both physical and computerized medical record storage.

The available literature is scant. The problem was identified as critical in 1965 by Whitston [7] of the Kaiser Foundation, but he provided no discussion of possible solutions. A year later Chelew [8], working at the Mount Zion Hospital and Medical Center in San Francisco, reported the results of a study of the utilization of active and inactive records; the age of the record and the purpose of the retrieval was noted. Records more than 12 years old made up less than 2 percent of all record retrieval at Mount Zion, and on this basis Chelew concluded that there was "little economic justification . . . for retaining medical records beyond the minimum legal requirements." Chelew did not attempt to define, except by this implication, when records become "less active"; the legal limit, of course, varies significantly from state to state. Recently Lennox [9] has suggested that records should be culled and only summary documents (in most cases the discharge letter with a few other items) be kept available. In this system the bulk of the culled record would be kept in quite inaccessible storage.

The method currently most used to identify and separate active from inactive records is based on a single variable—the elapsed time since the patient's last visit to the clinic or hospital. At arbitrary intervals all records are examined for the length of time since the patient's last visit or the last use of the record. Records not used during some previous number of months are sent to the inactive file or, in some hospitals, microfilmed; records that have been used within this time period are left in active status.

Lahey Clinic Foundation Study

The Lahey Clinic in Boston currently has a staff of approximately 100 physicians, divided into a dozen major specialties and approximately 20 recognized subspecialties. Between 500 and 600 patients visit the clinic daily. The clinic is currently well along in the construction of an on-line appointment-scheduling system [10,11]. For patients making repeat visits—who represent an overwhelming majority of calls received—the appointment secretaries need access to such items of information as the patient's physicians, the length of time since the patient was last in the clinic, the physician's orders concerning tests to be scheduled at the next appointment, and other data. Currently this information must be retrieved from the medical record for a significant number of patients. One aim of the Lahey appointment system is ultimately to keep on-line on the computer a file of pertinent scheduling information for those patients who have a high probability of calling for another appointment. The study reported in this article was motivated in large part by the consequent need to determine which patient records should be in the active file at Lahey.

It should be noted that the computer active file is to be based on patient appointment activity. Many authors rightly note other reasons to maintain a record in an active file [7,8], but a significant number of these other reasons coincide with or closely follow a patient visit (for example, diagnosis or third-party payment) and therefore are covered by a file based on appointment probability. Other uses such as research, education, and correspondence do not call for *immediate* record retrieval from either the computer or a physical record system and therefore are not important in establishing an active file; they may even be disregarded if the major criterion is the need for prompt access to the record, as it is in most outpatient settings.

Methodology

Selection of Variables

A four-man team (including one physician), all familiar with the clinic's procedures, hypothesized that the following variables should be predictive of patient, and therefore record, activity:

1. Number of months since patient's last visit
2. Sex
3. Age
4. Location of residence—distance from the clinic
5. Number of times patient had previously returned to the clinic
6. Specialty or specialties in which patient had received care
7. Patient's previous diagnosis

The first variable is the traditional one, with the probability of return expected to decrease as the number of months since the last appointment increases. The other variables were untested.

Data Collection and Processing

Information was extracted from the medical records for a random group of patients from the clinic's active and inactive patient files. The Lahey Clinic's active file covers about 50 000 patients who have been active within approximately the past 20 months; the inactive file contains about 650 000 records. Samples from each file were selected on the basis of common terminal digits, which provided a simple randomizing technique; 236 records were chosen from the active file and 920 from the inactive. The ratio of inactive to active records in the population is 13 to 1, while the ratio in the sample was 3.9 to 1. Thus the more interesting active file was given greater weight in the sample. A correction for this ratio was made later in the compilation of statistics.

For purposes of determining Variable 1, number of months since last visit, it was necessary to define the end of one visit and the start of another. The rule used was that all appointments not separated by a two-week appointment-free period were considered part of a single visit sequence. Thus a patient seen on

Table 1. Percent of Distant Patients Returning during Two-month Periods Following Single Previous Visit

Period (months)	Percent returning	Cumulative percent of group	Cumulative percent of ultimate returns
1– 2	5.79	5.79	20.79
3– 4	5.01	10.81	38.78
5– 6	3.14	13.95	50.07
7– 8	0.79	14.74	52.91
9– 10	0.97	15.71	56.37
11– 12	1.69	17.40	62.44
13– 14	2.12	19.52	70.05
15– 16	0.43	19.95	71.61
17– 18	1.04	20.99	75.34
19– 20	0.0	20.99	75.34
21– 22	0.18	21.17	75.99
23– 24	1.05	22.22	79.75
25– 26	0.19	22.41	80.41
27– 28	0	22.41	80.41
...	...	...	...
241–242	0	27.86	100.00

Apr. 2, 3, 4, and 15 and May 13 was recorded as having made two separate "visits" to the clinic, with the return visit within one month of the previous visit. Variable 3, age at time of appointment visit, was recorded in deciles. Location of residence (Variable 4) was recorded on a three-part scale. A "local" patient was defined as one whose home was within a one-hour drive of the clinic. A "semi-distant" patient was defined as having more than a one-hour but less than a three-hour automobile trip; these patients were judged more likely to return home than to stay in the clinic area during a two-day visit. Patients from outside this three-hour perimeter were classified as "distant." For Variable 6, all specialties in which the patient received care were specified. Diagnosis (Variable 7) was not recorded, since the categories were too numerous for adequate sample sizes.

Using the data gathered, summary printouts in the form of Table 1 showing patient return patterns were produced for each variable. Column 1 in Table 1 indicates two-month periods following the patient's previous visit. Patients who returned within two months following their previous visits are shown in the first entry in column 2 as a percentage of the total sample; their number was then subtracted from the total sample to yield the number of patients who could potentially return during the second two-month period after a previous visit. Thus the second entry in column 2 shows patients who actually returned during the third and fourth months following their previous visits, expressed as a percentage of all patients who did not return earlier. Subsequent entries in column 2 are based on similarly reduced denominators. In consequence the third column shows the cumulative percentage of *the total sample* who had returned by the end of any given two-month period, and the fourth column shows the cumulative percentage of *all patients who ultimately returned.*

To illustrate the use of this table: if records for distant patients who have been to the clinic only once are maintained in the active file for 24 months, 22.22

percent of these patients will have returned by that date (column 3). Since from this group 27.86 percent ultimately returned, 79.75 percent (22.22/27.86) of the returning distant patients will be in the active file at the time they make their next appointments.

Testing the Variables

Data on patient returns were grouped for chi-square testing in a matrix with the values of each variable on one axis and the number of returns within each of years 1–9 (with "10 or more" and "no return" as final categories) on the other.

Three of the variables considered exhibited significantly different patient return tendencies. As was expected, the time since the last visit was a significant factor at the 0.001 level when tested against a bogey of a hypothesized uniform return rate. In addition, the distance of the patient's home from the clinic was significant at the 0.01 level. By far the most significant difference was found when the distant patients were compared with local and semidistant patients combined. Finally, it was found that the likelihood of return of patients who had been to the clinic more than once was significantly different (0.01 level) from that of those who had been to the clinic only once. However, the return rate did not appear to vary significantly from visit to visit after patients had returned for their second visit.

None of these results is particularly surprising. The time variable has been well established by common knowledge and practice. It is also logical that patients from the local and near-local area tend to return sooner to the clinic than those from a distance. Perhaps the least intuitively obvious result is the tendency of patients who have made at least one return visit to the clinic to have a significantly different rate of subsequent return from patients who have been to the clinic only once. Yet this too appears logical. Many patients come to the Lahey Clinic for a "one-shot" reason; for example, to obtain a second opinion as to whether a specific surgical operation should be undertaken or to have specific therapy. Other patients use the clinic in an emergency when their own physician is unavailable. Still others may find on their initial visit that their care, their bill, the clinic location, or some administrative service was less desirable than they had expected. It is logical to expect these one-time patients to have a different and lower propensity to return to the clinic than patients who have shown by having already made multiple visits that the clinic's style fits their needs. The same pattern would be expected in hospital outpatient departments for transients, for patients who had a regular physician but who sought one-time semi-emergency or specialty care, and for other patients who did not choose to return.

There was no significant difference in return patterns between men and women. Similarly, age differences (tested by deciles) were not significant with regard to pattern of return. While those under thirty and over sixty returned sooner than those between thirty and sixty, this difference was not statistically significant.

As mentioned earlier, data were not collected for diagnoses because the categories were too numerous to allow tests in a sample of the size used here. The specialties and subspecialties at the Lahey Clinic (22) were also too numerous to justify any strong statements concerning the significance of this variable, although there were definite indications that return patterns do differ by specialty (allergy being a clear case of a specialty for which patients tend to return soon). This variable will be investigated more fully in the immediate future, by the automation of the current research methodology on the new computer system.

Determining File Components and Size

Utilizing only the two most significant variables, distance of residence and number of previous returns to the clinic, the Lahey patient population can be factored into four patient populations, local single visit (LS), distant single visit (DS), local multiple visit (LM), and distant multiple visit (DM), whose return patterns are unique. The patterns follow the tendencies described earlier, with higher percentages of local patients than of patients from a distance returning sooner and with patients who had been to the clinic more than once returning earlier and in greater proportion than patients who had been to the clinic only once. A chi-square test on these four groups, with the same return categories as in the earlier tests of variables, showed the return patterns of these sets of patients to differ significantly at the 0.01 level.

In order to estimate the overall size of the active file, it is necessary to calculate the contribution to file size for each of the four groups of patients separately. These may then be combined in a matrix as shown in Table 2 to determine the overall file composition. Each cell of the matrix for each patient type contains three figures. The first figure is taken from column 2 of the appropriate version of Table 1 and shows the expected return of those patients in the given time period as a percentage of those who have not returned earlier. The second figure in each cell is from column 4 of Table 1 and shows the cumulative return in the period as a percentage of all those patients who ultimately returned. The third figure is the estimated file size for that patient group alone; this figure must be calculated first for the single-visit groups, because some of these will ultimately feed into the multiple-visit groups. The file sizes are determined by subtracting from the initial patient population in each period the number of patients who have returned in previous periods.

Let a_k = percentage of patients returning in any two-month period k, and c = the number of new patients for each period. In any period n the file size S may be found from

$$S = \sum_{i=1}^{n} \left[\left(1 - \sum_{k=1}^{i} a_k \right) c \right]$$

The Lahey Clinic receives about 4000 new patients every two months, of which 3000 are local. Thus 3000 new local single-visit patients are available in each period for potential return; as shown in the appropriate column of Table 2, 10.7 percent of these will return within two months in period $i = 1$, and the file

Table 2. File Estimator Matrix*

Period (months)	Local multiple visit (LM)			Distant multiple visit (DM)			Distant single visit (DS)			Local single visit (LS)			
	1	2	3	1	2	3	1	2	3	1	2	3	
2	8.1	12.7	2.1	4.9	7.8	0.5	5.8	20.8	0.9	10.7	28.9	2.7	
4	14.7	35.8	4.0	12.1	27.0	1.0	5.0	38.8	1.8	8.6	52.1	5.1	
6	10.8	52.8	5.7	6.1	36.7	1.4	3.1	50.1	2.7	4.4	63.8	7.4	4%
8	5.0	60.5	7.2	4.8	44.4	1.8	0.8	53.0	3.5	2.9	71.5	9.6	Utili-
10	3.3	65.8	8.6	3.8	50.4	2.2	1.0	56.4	4.3	1.5	75.6	11.8	zation level
12	4.6	73.0	10.0	9.8	66.0	2.6	1.7	62.4	5.1	1.0	78.4	13.9	1%
14	1.8	75.8	11.3	4.0	72.3	2.9	2.1	70.1	5.9	0.5	79.8	16.0	
16	1.5	78.2	12.6	0.9	73.7	3.2	0.4	71.6	6.7	0.4	80.9	18.1	
18	1.0	79.8	13.8	2.2	77.2	3.5	1.0	75.3	7.5	0.6	82.7	20.2	0.5%
20	1.1	81.6	15.0	1.9	80.2	3.8	0.0	75.3	8.3	0.3	83.4	22.3	
22	0.9	82.9	16.2	0.3	80.6	4.1	0.2	76.0	9.1	0.1	83.7	24.4	
24	1.1	84.6	17.4	2.2	84.1	4.4	1.1	79.8	9.9	0.4	84.8	26.5	
26	0.8	86.0	18.6	1.9	87.1	4.7	0.2	80.4	10.7	0.0	84.8	28.6	
28	0.9	87.4	19.8	0.6	88.1	5.0	0.0	80.4	11.5	0.3	85.5	30.7	
30	0.4	88.0	21.0	0.6	89.1	5.3	0.0	80.4	12.3	0.3	86.3	32.7	
32	0.5	88.7	22.2	0.6	90.1	5.6	0.0	80.4	13.1	0.2	86.8	34.7	
34	0.1	89.0	23.4	0.6	91.1	5.9	0.0	80.4	13.9	0.4	87.8	36.7	
36	0.4	89.6	24.6	0.0	91.1	6.2	0.0	80.4	14.7	0.2	88.3	38.7	
38	0.3	90.1	25.8	0.6	92.1	6.5	0.0	80.4	15.5	0.1	88.7	40.7	
40	0.4	90.6	27.0	0.3	92.5	6.8	0.0	80.4	16.3	0.3	89.4	42.7	
42	0.4	91.2	28.1	0.3	93.0	7.1	0.4	82.0	17.1	0.1	89.8	44.7	
44	0.3	91.6	29.2	0.0	93.0	7.4	0.4	83.5	17.9	0.4	90.9	46.7	
46	0.5	92.4	30.3	0.0	93.0	7.7	0.0	83.5	18.7	0.1	91.2	48.7	
48	0.5	93.2	31.4	0.0	93.0	8.0	0.4	85.0	19.5	0.3	91.9	50.7	

* For each patient group, Col. 1 shows percent returns in each period; Col. 2, cumulative percent of ultimate returns; Col. 3, estimated file size in thousands of records at end of each period.

size S will be $(1-0.107)\,3000 = 2679$. The 10.7 percent who returned will now be transferred to the multiple-visit file, but in period $i = 2$ an additional 8.6 percent of the first 3000 new patients will return, and also a second pool of 3000 new patients will make their initial visits, again with a 10.7 percent rate of return within two months. Thus at the end of period $i = 2$, the local single-visit file will contain $(1-0.107)3000 + (1-0.107-0.086)3000 = 5100$ records. This process generates the third columns under the single-visit patient groups in Table 2; in each period those who have returned are transferred to the multiple-visit file.

The single-visit system will approximate a limit after about 100 periods; given the constant input of 3000 patients per period and the observed cumulative rate of 37.07 percent of new local patients ultimately returning, after 100 periods have elapsed a constant number of about 1100 patients will be transferred to the multiple-visit file in every period. Thus new patients returning for the first time are the first source of inputs to the multiple-visit file; a second source is the multiple-visit file itself, since each return visit results in the patient's reentering the file in time period 1; that is, the patient reenters the pool of those potentially able to return within two months of a previous (noninitial) visit. Therefore the multiple-visit file for any period m must be calculated by a series of recursive operations on an $m \times m$ matrix.

Defining $x_{i,j}$ as any cell in this matrix, assume that the first cell, $x_{1,1}$, is equal to C, the constant periodic input from the single-visit file. Let b_n denote the percentage of multiple-visit patients who will return in any time period n $(n = 1, \ldots, m)$. Then the other cells of the matrix are defined by

$$x_{m,1} = C + \sum_{n=1}^{m-1} b_n \, x_{(m-n),n} \qquad (m \geqslant 2)$$

$$x_{(m-n),(n+1)} = (1 - b_n) x_{(m-n),n} \qquad (n < m)$$

This system stabilizes after 70 periods (that is, when $m = 70$). The diagonal formed by the cells $x_{70,1}, x_{69,2}, \ldots, x_{2,69}, x_{1,70}$ provides the increase to the file in periods 1 through 70 respectively, and with this information the third entry in each multiple-visit cell of Table 2 may be generated.

Criteria for the Active File

Using Table 2 one can determine the composition of the active file by any of four possible criteria: the utilization probability desired, the file size desired, the percentage of patient records found in the active file on inquiry, or the optimum cost trade-off point between active and inactive files.

Utilization Probability

Constructing a file by utilization probability is perhaps most appealing—it is logical to desire an equal marginal probability of use for each group of the file components. If, for example, one wished to keep in the active file all records that had a probability of 4 percent or more of being used within the next two-month period, the file would be constituted by keeping active all those records represented by cells in Table 2 that show at least 4 percent patient return in that period. These are the cells in Table 2 appearing above the dashed line labeled "4 percent."

It would be esthetically more pleasing to use a smooth curve for this purpose rather than picking points from the discrete data, but the data do not follow any standard generating function because they are influenced by physician's instructions concerning patient returns, which cluster around 3, 6, 9, 12, 18, and 24 months. Some smoothing could be done, but the procedure would remain the same.

If one wanted to maintain in the active file those records having a probability of greater than 1 percent of being used in the next two-month period, the file would be made up of patient records in the cells above the "1 percent" line in Table 2. At the 1 percent level, local multiple-visit (LM) patients would be kept on the file for 28 months, distant multiple-visit (DM) patients for 26 months, distant single-visit (DS) patients for 18 months, and local single-visit (LS) patients for 12 months.

For any utilization level, the size of the file that would result also can be determined from Table 2 by adding the "file size" figures in the cell from each column showing the desired percent utilization. These figures represent a cumulative total of the number of patients in each category. For example, at the 1 percent level the file size would be 45 900—19 800 from LM, 4700 from DM, 7500 from DS, and 13 900 from LS.

File Size

To build a file constrained by size, the above procedure can be followed in reverse. Additions across the columns at each patient return level allow a fit to any standard file size. For example, if one were to use a disk file (or had an active medical records storage room) that held 68 000 records and wanted to maintain a constant probability of record use, it can be determined that the appropriate patient return level would be 0.5 percent and that 48, 38, 24, and 18 months of records would be stored from the four categories.

Inquiry Percentage Found

It is also possible to determine the file size and composition by stating a decision rule with regard to the percentage of "hits" one desires in the active file upon inquiring into that file. For example, it might be desirable that, on average, the records of 80 percent of patients inquiring for appointments would be found in the active file. The second figure in each cell of Table 2 is useful for this purpose. While an adjustment is necessary here for the size of each group, the 80 percent level can be seen to follow approximately the 1 percent patient return line; 87.4 percent, 87.1 percent, 75.3 percent, and 78.4 percent of the four respective groups will have returned by the end of the periods indicated.

Cost Trade-off

Finally, one might wish to determine the size of the active file on the basis of the trade-off between the cost connected with retrieving a record from the active file and the cost of retrieving a record from the inactive file. In a gross view, these are the costs connected with finding, delivering, and replacing a record. In addition, a further cost can be assigned to each record in the inactive file —the cost of not having it immediately available when needed. This cost will, of course, be highly subjective.

Since some of the above costs are fixed for each institution, the costs-per-record of each type of storage will vary as different record volumes are assigned to active or inactive storage. In theory, at least, one would attempt to minimize the sum of all costs. At the active file size indicated by this minimum total cost level, the number of records in the active file can be used to enter into Table 2 to determine the percentages of record use and successful file inquiry that would be obtained. Should these parameters be unsatisfactory to management, the necessary deviation from the minimum cost could be estimated. The subjective costs necessary for the cost trade-off method are difficult to estimate and in practice may force the choice of another method. Nevertheless this procedure could yield a solid managerial understanding of the relation of storage policy, cost, record use, and successful inquiry level.

Implementation at the Lahey Clinic

In implementing the system in November 1971, the Lahey Clinic management chose file size as the predominant criterion for two reasons. First, limited disk storage space on the computer system dictated an upper limit of less than 75 000 records; second, it was desired to equate the computer file size with the physical active file capacity so that "purging lists" could be prepared for the medical record room by the computer system. This dictated a still smaller file size, on the order of 55 000 records. The "1 percent" line in Table 2 was therefore chosen as the initial standard, since it approximated the desired file size. Some minor adjustments were made to this standard to increase the actual number of records in the active file and to make allowances for some initial patient data limitations.

The results of the system are within expected limits. The active file as constituted by this method has remained at approximately 50 000 patient records, and while the size varies somewhat with the actual patient flow each month, the variation is not significant. For the adjusted 1 percent record-use level selected it would be expected that 82.4 percent of the returning patients would be on the file. In an initial sample of 300 patients, 84.1 percent were found to be on the file when they called for appointments. This difference is not significant at the 0.05 level on a chi-square test, which suggests that the method has good predictive value along the inquiry-percentage-found dimension.

Although it would be difficult to select records to be moved to inactive storage from the active record file on the basis of these rules with only a manual system, it is a simple matter to do so when the requisite patient variables are recorded on a computer file. By use of the decision rule, the Lahey computer system is now automatically moving computer-stored records from the computer active file to the computer inactive file and also producing a listing of those records to be purged from the active file in the medical record room. The listing is in the terminal digit order desired by the medical record purging team and facilitates faster and more frequent active file purges.

In sum, early experience with the active file suggests that it will fulfill expectations. A final evaluation will not be possible until the entire computer system, of which the file is a part, has been stabilized; then cost-benefit and other data will eventually become available.

Discussion

The method has been illustrated in a case where three significant variables were found (two in addition to the traditional time variable). There is, however, no reason why additional variables cannot be added to the system if they are found to define other pools of patients who return in significantly different patterns. While unique sets of patients are apparent here, it is important in adding other variables to take into account the likelihood of a high covariance between some (for example, between "age" and "specialties in which patient has received care") and to use only the best relatively independent predictors. The analysis matrix of Table 2 would have to include each pool and would therefore assume additional columns, but the basic method would not change.

It appears likely that the variables noted in this study will be among those found significant in other settings. The same variables were found to be significant in work performed in 1967–69 in the outpatient departments of Massachusetts General Hospital by the author and members of the Massachusetts General Hospital Laboratory of Computer Science (before the present method of utilizing this information had been developed).

As indicated, some of the variables—among them diagnosis and specialty—were not adequately dealt with because of the sample size of only 1156. A fully automated appointment system such as that being developed at the Lahey Clinic will allow an automatic trace of patient activity on multiple variables at a low marginal cost in computer time. It will thus be possible to investigate additional variables, improve the sample size, and further refine the output of the method described.

Acknowledgments. The author is grateful to Miss Deborah L. Parent, Lahey Clinic head programmer, and Mr. Bruce E. Stangle, research assistant, for their help in programming and editing.

References

1. Bennet, A. E. and W. W. Holland. The medical record and the computer. I. The selection of input. In McLachlan, G. and R. A. Shegog (eds.), *Computers in the Service of Medicine*, 2d ed. London: Oxford University Press, 1968.
2. Greenes, R. A. et al. Recording retrieval and review of medical data by physician-computer interaction. *New Eng. J. Med.* 282:307 Feb. 5, 1970.
3. Kiely, J. M. et al. A computer-based medical record: Entry of data from the history and physical examination by the physician. *J.A.M.A.* 205:271 Aug. 19, 1968.
4. Korein, J. et al. Computer processing of medical data by variable-field-length format. *J.A.M.A.* 186:132 Oct. 12, 1963.
5. MacLeod, V. and R. Vaughan. Medical records: To microfilm or not? How to make the decision. *Hospitals* 43:71 Mar. 16, 1969.

6. Martin, S. F. *An On-line, Real-time Record Management System for a Medical Clinic.* Master's thesis, Massachusetts Institute of Technology, 1968.
7. Whitston, C. W. An analysis of the problems of medical records storage. *Hosp. Mgmt* 100:27 October 1965.
8. Chelew, P. G. Most medical records are kept too long. *Mod. Hosp.* 107:105 August 1966.
9. Lennox, B. and K. J. White. Selective culling of hospital medical records. *Lancet* 1:489 Mar. 6, 1971.
10. Rockart, J. F. Model-based systems analysis: A methodology and case study. *Indust. Mgmt Rev.* 11:1 Winter 1970.
11. Rockart, J. F. A new look at clinical schedules. *Technol. Rev.* 72:35 December 1970.