WORKING PAPER
ALFRED P. SLOAN SCHOOL OF MANAGEMENT

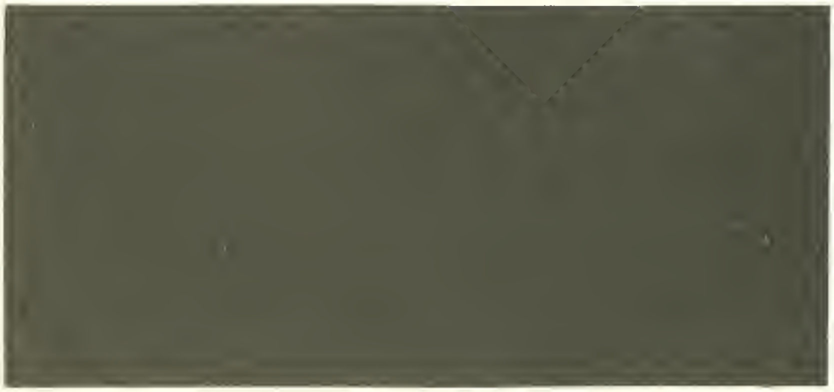
A METHOD OF DETERMINING WHICH MEDICAL RECORDS SHOULD BE STORED IN THE "ACTIVE FILE"*

by

John F. Rockart**

March 1972 593-72

MASSACHUSETTS
INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02139



A METHOD OF DETERMINING WHICH MEDICAL RECORDS SHOULD BE STORED IN THE "ACTIVE FILE"*

by

Francis
John F. Rockart**

March 1972 593-72

---

*This investigation was supported by PHS Research Grant No. 3 R18 HS00307-01 from the National Center for Health Service Research and Development, HSMHA. The assistance of Miss Deborah L. Parent and Mr. Bruce E. Stangle is gratefully acknowledged.

**Associate Professor of Management, Alfred P. Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts, and Research Associate, Lahey Clinic Foundation, Boston, Massachusetts.

Abstract

As the numbers of patients seen by medical facilities increase, the question of which medical records to keep "active" has become increasingly important. Limited storage space forces the development of a decision rule to be used to separate active from inactive records. With the advent of computer-stored records and the need to utilize most effectively costly direct-access storage, the question of which records should be considered to be active has become significant for an additional reason. This paper presents a method to define, more precisely than existing methods, those records which should be stored in the active file--either the physical medical record file or the computer-stored record file.

A METHOD OF DETERMINING WHICH MEDICAL RECORDS SHOULD BE STORED IN THE "ACTIVE FILE"

John F. Rockart, Ph.D.

In recent years increasing attention has been centered upon the medical record as a prime instrument of continuing patient care. Many problems concerning the record have been investigated extensively. Prime among these 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.

Given this concern about the medical record, noticeably little has been written concerning the problem of determining whether a particular record should be kept available or "discarded" --either completely or into a less accessible storage location. Stated differently, little is known about effective algorithms for separating those records which have a high probability of reuse (and therefore should be kept in an "active" file) from those which have a low probability of future utilization (the "inactive" file). The work reported here concerns this problem and notes the results of a study at the Lahey Clinic Foundation in Boston which was aimed at determining the size and most efficient composition of the "active" file of records.

One major aim of such a study is, of course, to determine which <u>physical</u> records should be kept in the active file. The question of physical record storage is important since the active file space is generally quickly accessible "prime space" located in the main building of a medical facility. Access time to a

particular record is usually very fast--but the cost of the storage space is high. The "inactive" file, on the other hand, is usually located in a less accessible, less expensive storage space which is farther away from the point of care and from which access time may be considerably longer.

Today, however, an additional need for an efficient solution to the problem of record storage has arisen. This is the need 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 storage of at least some data, medical or administrative, concerning patients on computer files is presently 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 physical medical record in active physical storage.

In summary, the problem of determining which records to store in the active file is now of increased interest. A more effective solution can provide benefits with regard to both the physical record storage problem and the computerized medical record storage problem.

Current Methods

The current, most used general solution to the problem of identifying and separating the active records from the inactive

records is a simple one. The solution is based on a single variable--the time since the patient's last visit to the clinic or hospital. At arbitrary intervals, all records are scrutinized to determine the length of time since the patient's last visit or, in some cases, the last use of the record. Those records for which there has been no activity in "n" months (where "n" may represent 12, 18, 24, and so forth months) are sent to the inactive file.* The records that have been used within this time period are left in active status.

Several methods have been devised to make this periodic purge simpler. Records may be date stamped in a clearly visible place at each time of use. Alternately, an imprinting of a series of years on the outside of the jacket allows a simple checkmark to be made opposite the appropriate year during the record's first use in that year. At purge time, all records not having a check showing utilization in year 19xx or afterward are consigned to the inactive file. This latter method gains in clerical ease and neatness what it loses in precision of the date of the last use.

As noted, the available literature is scant. The problem itself was identified in 1965 by Whitston[7] of the Kaiser Foundation. He suggested that "the scientific culling and control of inactive records is a generally overlooked but critical function" but 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 "designed to measure the utilization experienced by med-

---

*In some institutions, this may mean microfilming.

ical records retrieved from either active or inactive storage." The age of the record retrieved was noted as well as the purpose of the retrieval. Chelew concluded that the over-all use rate for records at Mount Zion dropped below 2 percent of all record retrievals for those records more than 12 years old and that, on the basis of his study, "little economic justification can be found for retaining medical records beyond the minimum legal requirements." This limit, of course, varies significantly from state to state.

Although noting that microfilming older, less active records has helped ease the storage problem, Chelew did not attempt to define, except by the above implication, when records become "less active." 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.

Basis for The Lahey Clinic Foundation Study

As noted earlier, however, the advent of computer-stored medical record data has increased the importance of a more precise understanding of a patient's (and therefore a record's) probability of activity. It was, in large part, the need to decide which patient records should be kept on a computer "active patient file" which initiated a study at the Lahey Clinic.

Founded by Dr. Frank H. Lahey in 1925, the Lahey Clinic currently has a staff of approximately 100 physicians, divided lust records obtained from ellipse [illegible]

The use of the [illegible] was not [illegible]

of the retrieval [illegible] "Ielo [illegible] conditions" [illegible]

for records [illegible] and [illegible]

[illegible]

the [illegible]

found [illegible]

requ[illegible]

from [illegible]

[illegible]

into a dozen major specialties and approximately 20 recognized subspecialties. Between 500 and 600 patients visit the clinic daily. Patients requiring hospital beds are admitted to several neighboring hospitals, primarily the New England Deaconess and the New England Baptist Hospitals.

The clinic is currently well along in the construction of an on-line system which will enable appointment scheduling personnel to serve more efficiently the needs of patients requesting appointments.[10,11] To service effectively the needs of old patients, who represent an overwhelming majority of calls received, the appointment secretaries need access to such items of information as (1) the patient's physicians--to ensure that patients are returned to those specialists most familiar with their case, (2) the length of time since the patient was last in the clinic--to determine the length of appointment necessary, (3) the physician's orders concerning tests to be scheduled at the next appointment, and (4) other similar data. All of this information is currently stored in the medical record which is called for and used to obtain the data for a significant number of patients. One ultimate aim of the Lahey appointment system is to keep on-line on the computer a file of scheduling pertinent information for those patients who are expected to have a high probability of calling for another appointment, that is, those patients who are deemed to be "active." Thus the problem of determining which "patient medical records" should be in the active file at Lahey is a significant one.

Methods

A four-step process was utilized in the study. These steps were:

1. Hypothesizing of variables affecting patient activity,
2. Data collection and processing,
3. Determination of significant explanatory variables, and
4. Development of a method for the determination of the size of the active file and the construction of the file.

Hypothesis of Variables. A four-man team, knowledgeable in the clinic's procedures (including one physician) hypothesized the following variables to be indicative of possible patient, and therefore record, activity:

1. The number of months since the patient's last visit,
2. Sex of patient,
3. Age of patient,
4. Location of the patient's residence--distance from the clinic,
5. Number of times the patient had previously returned to the clinic,
6. Specialty or specialties in which the patient has received care, and
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 increased. The other variables had previously been untested.

It should be noted that the computer active file was to be built on the basis of 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 record uses appear to be coincident with, or only slightly lagging, a patient visit (for example, third party payment use, diagnosis department use) and therefore fully covered by an active file based on patient appointment probability. The remaining uses (for example, research purposes, education, correspondence) do not call for immediate record retrieval in either the computer or physical record system. They therefore are not important in the determination of an active file and even may be neglected if the major criterion for the active file is the need for prompt access to the record--which it is in most outpatient settings.

Data Collection and Processing. Data were collected by extracting information from the medical records for a random group of patients from both the clinic's active and inactive patient files. The Lahey's active file contains patients who have been active within approximately the past 20 months and currently numbers some 50,000 records. The inactive file contains 650,000 records. Samples from each file were selected on the basis of common terminal digits. Terminal digit filing provided a simple randomizing technique and also greatly eased the record pulling problem. Some 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 is 3.9 to 1. Thus the more interesting active file was given

greater weight in the sample with a later correction for this ratio being made in the compilation of statistics.

For recording and analysis purposes, the definitions of variables (2), sex, and (5), number of previous return visits, are apparent. Variable (3), age at time of appointment visit, was recorded in deciles. All specialties (variable 6--approximately 20) in which the patient received care were recorded. Diagnosis (variable 7) was not recorded since it was evident that the sample would be too small to allow sufficient sample sizes.

Location of residence (variable 4) was recorded on a three-part scale. A "local" patient was defined as a patient whose home residence was within a one-hour drive of the clinic. A "semi-distant" patient was defined as having more than a one-hour automobile trip yet less than a three-hour trip (that is, a patient who was more likely to return home than stay in the clinic area during a two-day visit). All patients from outside this perimeter were considered as being "distant."

For purposes of determining the "number of months since last visit" it was necessary to define the end of one visit and the start of another. The rule utilized was that all appointments not separated by a two-week appointment-free period were considered as part of the same visit sequence. Thus, if a patient were seen on April 2, 3, 4, 15, and May 13, he was recorded as having made two separate "visits" to the clinic--with his return being within one month following the previous visit.

Using the data gathered, printouts in the form of Table 1 greater ... in the sample with a lower ...

ratio being ... (significantly ...

... (significantly ...

(tables ...

...

[illegible]

were produced for each variable. This enabled visual inspection of the data and formed the basis for the ultimate system printout--to be discussed later. Column 1 indicates two-month periods following the patient's previous visit. Column 2 shows the percent of patients who returned of those who could have returned during this period.* Column 3 is a cumulative percent of returning patients. Column 4 denotes the percentage of those who ultimately returned (TOT) who had returned by the end of this period (CUM/TOT). In essence the figure in column 4 states, for each time period, that if the records from this group are maintained on the active file until this number of months following the previous visit, "n" percent of the patients who return will be in the active file at the time of their return.

To illustrate the use of this table, if records for distant patients who have been to the clinic only once are maintained on the active file for 24 months, 22.22 percent of these patients will have returned by that date (column 3). Since the total percent whoever returned was 27.86 percent, some 79.75 percent (22.22/27.86) of the patients (column 4) would be in the active file at the time they inquired to make their next appointment.

Results

Data on patient returns were grouped for chi square testing in a matrix with the values of each variable on one axis and

---

*The percentage of patients who had returned was adjusted in this calculation to take into account those patients who could not have returned in this time period, that is, the denominator (initially the entire sample in the first period) was reduced in each subsequent time period by those patients who had been away from the clinic for less than the number of months denoted by that time period.

the number of returns within years one to nine (with ten or more and "no return" as final categories) on the other. Of the variables considered, three exhibited significantly different patient return tendencies. Two variables showed no significant differences, while there was insufficient data to test the two remaining variables.

Significant Variables. 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 juxtaposed with all others (that is, local and semidistant combined). Finally, it was found that the propensity to return of patients who had been to the clinic more than once was significantly different (0.01 level) than 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 can be viewed as particularly surprising. The time variable has been well established by common knowledge and practice. It is also logical to assume that patients from the local and near-local area might tend to return more quickly to the clinic (as they do) than those from a distance.

Perhaps the least intuitively obvious result is the tendency of patients who have returned to the clinic at least once to have

a significantly different rate 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 elicit 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 other 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 illustrated by having already made multiple visits that the clinic's "style" fits their needs. One would expect the same pattern in hospital outpatient departments caused by transients, patients who had a regular physician but who sought one-time semiemergency or specialty care, and other patients who did not choose to return.

Nonsignificant Variables. 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. (It was noted that those less than 30 and more than 60 years of age returned sooner than those between 30 and 60; this difference was not statistically significant, however.)

Additional Data Needed. The numbers of different patient diagnoses were too numerous to allow tests on this variable. In

addition, the numbers of different specialties and subspecialties at the Lahey Clinic were too great (22) to allow us to make strong statements concerning the significance of the data obtained in this study. As might be expected, there are definite indications that returning patterns do differ significantly by specialty utilized (allergy being a clear case of a specialty for which patients tend to return more quickly than for most other specialties). We plan to investigate this variable more fully in the immediate future by the automation of the current research methodology on the new computer system.

Utilizing only the two most significant variables (the distance of residence and the number of previous returns to the clinic), the Lahey patient population can be factored into four separate patient populations (Table 2) whose return patterns are unique. The patterns follow the tendencies just discussed with higher percentages of local patients returning more quickly than patients from a distance and with patients who have been to the clinic more than once clearly returning earlier and in greater numbers than the patient who had thus far been to the clinic only once.

A chi square test on these four groups (again with aggregations at years 1-9, 10+, and no return) shows the returning patterns of these sets of patients to be significantly different at the 0.01 level.

Use of Results

Table 3 illustrates a form in which the results of the study

additions the numbers of different specialties [illegible]
at the Lahey Clinic in one year (19[illegible]) [illegible]
statements concerning the significance of [illegible]
this study. [illegible]
that occurring [illegible]
utilized (a) [illegible]
patients [illegible]
tics). [illegible]
immediate [illegible]
ology [illegible]
[illegible]
tance [illegible]
the [illegible]
patient [illegible]
The [illegible]
pa[illegible]
[illegible]
mo[illegible]
[illegible]
[illegible]
pat[illegible]
[illegible]

and the associated data can be utilized to determine which records should be kept in the active file and which records should be transferred to a secondary storage area.

The first entry in each cell of the matrix for each of the four patient categories in Table 3 shows the expected percent return of patients by two-month periods (rows) for the group under consideration (columns). This figure is taken for each group from column 2 of the appropriate version of Table 1. The second entry is the percent of returning patients who had returned by the end of this period (Table 1, column 4). The third figure is the estimated file size for this group at this point (see the Appendix for the derivation of these file sizes).

Using this table, one can determine the composition of the active file by any of four possible criteria. These are:

1. Utilization level desired,
2. File size desired,
3. Percentage of patients desired to be found in the active file upon inquiry, and
4. Optimum cost trade-off point between active and inactive files.

Utilization Level. Constructing a file by utilization level is perhaps most appealing--it is logical to desire an equal probability of return for each marginal (low end) group of each of the file components--and easily illustrated. If, for example, one wished to keep in the active file all records that had at least a 4 percent chance of being utilized within the next two-month

period, the file would be constituted by keeping active all those records for patients fitting the grouping shown above the dashed line labeled "4 percent" at the right side of Table 3. If one wanted to maintain in the active file those records having a probability of greater than 1 percent of being utilized in the next two-month period, the file would be made up of the patient record in the cells above the "1 percent" line in Table 3. (At this 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 be developed can also be determined from Table 3. This is done by adding the "file size" figures in each of the four columns in the cell just above the service level line. These figures represent a cumulative total of the number of patients in each category. For example, at the one percent level, the file size would be 45,900--a sum of 19,800 from LM, 4,700 from DM, 7,500 from DS, and 13,900 from LS.

File Size. To build a file contained by the size of the file, the above procedure can be followed in reverse. (Simple additions at each utilization level allow a fit to any standard

---

*It would be aesthetically more pleasing to use a smooth curve rather than picking points from the discrete data. However, affected as it is by physician instructions concerning patients' returns (these instructed-returns cluster around 3, 6, 9, 12, 18, and 24 months), the data do not follow a standard generating function. Some smoothing could, of course, be done but the methodology would remain the same.

size of file, if file size is the binding constraint.) For example, if one were to use a disk file (or had an active medical records storage room) which held 68,000 records and wanted to maintain a constant percent utilization level then it can be determined that the appropriate utilization level would be 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 is possible to state that the file should be built so that, on average, 80 percent of patients inquiring for appointments will be found in the file. An adjustment is necessary here for the size of each group, but the 80 percent level can be seen to follow approximately the 1 percent service level line (with 87.4, 87.1, 75.3, and 78.4 in the four groups) discussed previously.

Cost Trade-off. Finally, one might wish to develop the size of the active file on the basis of the trade-off between the costs concerned with locating a record in the active file and the costs of locating a record in the inactive file. In a gross view, the costs can be stated quickly although, especially in the case of a computer system, there are significant calculations to be made. For both the active and inactive files, there are costs connected with storage and procuring the record (finding it, delivering it, and replacing it in the file). 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 latter cost

will be, of course, highly subjective.

Since, for each institution, some of the above costs are fixed, 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 3 (as in the file size case noted previously) to determine the service level and inquiry hit percentage which would be obtained. Should these parameters be unsatisfactory to management, a deviation from the optimum cost level can be made by increasing the size of the active file.* In any case a solid managerial understanding of the effects of any particular active-inactive storage policy on cost, service level, and inquiry hit level can be gained from this method.

Summary and Conclusions

It is initially necessary to divide the patient population into those subpopulations which have different patterns of return. The size and composition of the active file can then be determined by any of four possible criteria. Using any one of the criteria

---

*Although it would be interesting to develop further the mathematics of this cost trade-off, it has not been done here since it requires a rather extensive, although not difficult, formulation which is secondary to the purpose of this article. In addition, in practice, especially with computer systems, it is felt that most active files will probably be constructed by method (2) file size desired or, less often, by methods (1) utilization level desired or (3) "hit" percentage. The subjective cost estimates necessary for method (4) "cost trade-off" are difficult to make--and, in practice, it is felt will tend to force the choice of another method. The Lahey computer file is currently composed using method (2) desired file size.

as the primary constraint, the explicit or implicit (in the case of cost trade-offs) values of the other three criteria can be quickly obtained. The method that has been presented thus enables an administrator using the criteria most germane to his institution to develop the "optimal" active-inactive file trade-off for either a computer system or a medical record room.

The method has been illustrated in a case where three significant variables were found (two in addition to the 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. The analysis sheet (Table 3) would have to include each pool and therefore would assume additional columns. But the basic method would not change.

It is apparent that there are some variables which have not been adequately dealt with because of the sample size of only 1,156. Among these variables, as previously noted, are diagnosis and specialty. A fully automated appointment system, such as is 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 is thus possible to investigate additional variables, improve the sample size, and further refine the output of the method illustrated.

Although it would be difficult to select records to be moved to the inactive file from the active physical medical record storage file based on 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 noting the decision rule,

the computer can both automatically move the computer-stored records from the computer active to the computer inactive file and also produce a listing of those records to be "purged" from the active file in the medical record room. Obviously the listing can be in any order desired by the medical records purging team.

Appendix

Size Computation for Each Group

The file sizes shown in Table 3 for each new group are based on figures of 4,000 new patients for each period of which 3,000 are local.

A. Single Visit Patients. The file sizes are determined by subtracting from the initial patient population in each period the sum of the number of patients who have returned in previous periods. The percentage of patients who return in each period is given in the first column of Table 3 for each group.

Let a_i = percentage of patients returning in time period i, and c = the number of new patients for each period. In any period n the file size will equal

$$\sum_{i=1}^{n} (1 - \sum_{k=1}^{i} a_k)c.$$

For example, the number of local single visit patients in the file if this group were kept active for only one period would be

$$(1 - 0.107)\ 3{,}000 = 2{,}700.$$

After period 2 the file size would be

$$(1 - 0.107)\ 3{,}000 + (1 - (0.107 - 0.086))\ 3{,}000 = 5{,}100.$$

To state this in levels, there are 3,000 local new patients who can be in period 2. Of these 3,000 new patients, 10.7 percent will have returned by the end of period 2 and will have been put in the multiple visit file. Of the 3,000 patients who came in period 1, 10.7 percent will have returned in time period 1 and another

Size composition [illegible]

The life [illegible] on Elasmo[illegible] are long).

[illegible] by subtract[illegible] the sum of [illegible] periods [illegible] is clos[illegible]

[illegible] and [illegible] n the [illegible]

8.6 percent in period 2. These patients will have been transferred to the multiple visit group.

B. Multiple Visit Patients. The file for multiple visit patients receives inputs from two sources. The single visit file (new patients returning) is the first source, while the second source is the multiple visit file itself.

The computation of the number of returning new patients assumes a constant number of new patients each time period (which can be shown mathematically to occur after approximately 100 periods have passed). At that point the number of patients "transferring" in any period from the single visit system has stabilized. To determine the number of returning new patients each period, take the sum over 100 periods of the percentage of new patients who return in each period and multiply it by the number of new patients in each period. For example, to determine the number of patients who will enter the local multiple file each period, calculate

$$\sum_{i=1}^{100} a_i$$

where a_i = percentage of local new patients returning in period i. In this case the cumulative percentage is 37.07 percent. Since it was initially assumed that 3,000 local new patients entered the system each period, there are 0.3707 x 3,000 or approximately 1,100 patients who will enter the local multiple visit file in each time period.

In the second case, the second input, returning patients who return again, occurs because each time a multiple visit

patient returns he re-enters the multiple visit file in time period 1. To calculate the multiple visit file size it is necessary to perform a series of recursive operations on an m X m matrix.

DEFINE

b_n as the percentage of multiple visit patients who will return in time period N (n = 1,2,3.....M).

c as the input to the multiple visit file from the single visit file in each period.

$x_{i,j}$ as a cell in the m X m matrix.

ASSUME

$$x_{1,1} = c$$

Then calculate for each $m \geq 2$

$$(1) \quad x_{m,1} = c + \sum_{N=1}^{m-1} b_N \, x(m-n),n$$

(2) for each n < m

$$x_{(m-n),\,(n+1)} = (1 - b_n) \; x(m-n),n$$

This system stabilizes after 70 periods (that is, when m = 70). The diagonal formed by the cells $X_{70,1}$ $X_{69,2}$,..., $X_{2,69}$, and $X_{1,70}$ provides the increase to the file in periods 1 through 70 respectively. For example, the file size at the end of the period n where 1 n 70 $=_{i=1} X(71-i,i)$.

Table 1

Summary of Patient Returns

Distant

Visit Followed by Another Visit

Months After Previous Visit	Percent Returning This Period	Cumulative Percent Returning (CUM)	CUM/TOT*
1 to 2	5.79	5.79	20.79
3 to 4	5.01	10.81	38.78
5 to 6	3.14	13.95	50.07
7 to 8	0.79	14.74	52.91
9 to 10	0.97	15.71	56.37
11 to 12	1.69	17.40	62.44
13 to 14	2.12	19.52	70.05
15 to 16	0.43	19.95	71.61
17 to 18	1.04	20.99	75.34
19 to 20	0.0	20.99	75.34
21 to 22	0.18	21.17	75.99
23 to 24	1.05	22.22	79.75
25 to 26	0.19	22.41	80.41
27 to 28	0.0	22.41	80.41
29 to 30	0.0	22.41	80.41
241 to 242	0.0	27.86	100.00
243 to 244	0.0	27.86	100.00

27.86 = (TOT) = Final percent of this patient group who ultimately returned.

SHARE = 0.36 OF TOTAL 2,172

*CUM/TOT = Percentage of those who ultimately returned (TOT) who had returned by the end of the period shown in column 1.

Table 2

Separate Patient Populations

Number Clinic Visits	Residence	
	Local	Distant
One	Local single visit (LS)	Distant single visit (DS)
Two or more	Local multiple visit (LM)	Distant multiple visit (DM)

Table 3

File Estimator Table*

onths Away	Local Multiple Visit (LM) 1	2	3	Distant Multiple Visit (DM) 1	2	3	Distant Single Visit (DS) 1	2	3	Local Single Visit (LS) 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	Utilization
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	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	

*This table shows the differing return patterns for the four groups of patients defined in Table 2. For each group, three columns of data are presented for each two-month period following the prior visit. These columns, numbered 1, 2, and 3 represent:

Column (1)--The percentage of patients in this group returning in the two-month period (equivalent to column 2 of Table 1).

Column (2)--The percentage of those patients who ultimately returned who had returned by the end of this period (equivalent to column 4 of Table 1).

Column (3)--The estimated file size in thousands for this group at this point in time (see Appendix for derivation).

References

1. Bennet, A.E., and Holland, W.W.: The medical record and the computer. I. The selection of input. In McLaghlan, G., and Shegog, R.A. (eds.): Computers in the Service of Medicine, 2nd ed., p. 122. London: Oxford University Press, 1968.

2. Greenes, R.A., Barnett, G.O., Klein, S.W., et al.: Recording retrieval and review of medical data by physician-computer interaction. N. Engl. J. Med. 282:307-315 (Feb. 5) 1970.

3. Kiely, J.M., Juergens, J.L., Hisey, B.L., et al.: A computer-based medical record. Entry of data from the history and physical examination by the physician. J.A.M.A. 205:571-576 (Aug. 19) 1968.

4. Korein, J., Tick, L.J., Woodbury, M.A., et al.: Computer processing of medical data by variable-field-length format. J.A.M.A. 186:132-138 (Oct. 12) 1963.

5. MacLeod, V., and Vaughan, R.: Medical records: to microfilm or not? How to make the decision. Hospitals 43:71-76 (March 16) 1969.

6. Martin, S.F.: An On-Line, Real-Time Record Management System for a Medical Clinic. Unpublished S. M. Thesis, Massachusetts Institute of Technology, Cambridge, 1968, 90 pp.

7. Whitson, C.W.: An analysis of the problems of medical records storage. Hosp. Management 100:27-36 (Oct.) 1965.

8. Chelew, P.G.: Most medical records are kept too long. Mod. Hosp. 107:105-107 (Aug.) 1966.

9. Lennox, B., and White, K.J.: Selective culling of hospital medical records. Lancet 1:489-490 (March 6) 1971.

10. Rockart, J.F.: Model-based systems analysis: a methodology and case study. Industrial Management Rev. 11:1-14 (Winter) 1970.

11. Rockart, J.F.: A new look at clinical schedules. Technology Review 72:35-39 (Dec.) 1970.