DOCUMENT RESUME

ED 046 446 LI 002 463

AUTHOR	King, Donald W.; Caldwell, Nancy W.
TITLE	Study of the Cost-Effectiveness of Retrospective Search Systems.
INSTITUTION	Westat Research, Inc., Rockville, Md.
SPONS AGENCY	American Psychological Association, Washington, D.C.
PUB DATE	Jun 70
NOTE	44p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	*Cost Effectiveness, *Costs, Decision Making, Electronic Data Processing, Information Retrieval, *Information Systems, *Models
IDENTIFIERS	*American Psychological Association, APA

ABSTRACT

The purpose of this study was to explore cost-effectiveness factors that affect the choice among alternative systems. A cost-effectiveness model that may be used to evaluate potential systems was derived and a statement of the general magnitude of costs that the American Psychological Association (APA) can expect in implementing and operating alternate systems is given. These "typical" costs may be used to establish the cost-effectiveness of general classes of systems, such as mechanized search on-line by titles or by batch processing on index terms. A bibliography used for the literature survey is appended. (Author/AB)

STUDY OF THE COST-EFFECTIVENESS OF RETROSPECTIVE SEARCH SYSTEMS

A Study Conducted for the
AMERICAN PSYCHOLOGICAL ASSOCIATION

By

Donald W. King
Nancy W. Caldwell

Westat Research, Inc.
7979 Old Georgetown Road
Bethesda, Maryland 20014

PREFACE

During the period from June 1968 to April 1970, APA's Office of Communication Management and Development, working under the direction of the ad hoc Committee on Communications, prepared a plan for the development of a National Information System for Psychology. During the course of developing that plan, it became evident that a variety of decisions would have to be made with respect to the development of a retrospective search capability for the system. How could such a system be configured? Of the number of potential users, how many would actually use it? What would the system cost, and what would be the costs per search for each use?

In an attempt to obtain a model whereby questions of this type could be answered, Mr. Donald W. King, Executive Vice President of Westat Research, Inc., Betheada, Maryland, was selected to conduct a study of the cost-effectiveness of retrospective search systems. Mr. King is well known in the information science field for his work in evaluation and economics of information handling. This paper, by Mr. King and Nancy W. Caldwell, does not necessarily reflect the opinion or interpretations of the American Psychological Association.

Harold P. Van Cott
Director, Office of Communication
Management and Development

TABLE OF CONTENTS

Section		Page
	PREFACE.	i
I	INTRODUCTION	1
II	RETROSPECTIVE SEARCH SYSTEMS.	3
III	COST MODEL.	10
	Fixed Cost	10
	Variable Costs.	11
IV	COSTS OF ALTERNATIVE SYSTEMS	13
	BIBLIOGRAPHY	39
	Reports and Articles.	39
	Memoranda and Informal Reports.	41

I. INTRODUCTION

The American Psychological Association's Office of Communication Management and Development is considering alternative system designs for retrospective searching services. The purpose of this study was to explore cost-effectiveness factors that affect the choice among alternative systems. A cost-effectiveness model that may be used to evaluate potential systems was derived, and a statement of the general magnitude of costs that APA can expect in implementing and operating alternate systems is given. These "typical" costs may be used to establish the cost-effectiveness of general classes of systems, such as mechanized search on-line by titles or by batch processing on index terms.

As part of this study, a thorough literature search was conducted. To complement the sketchy information obtained from the search, we contacted a number of persons who manage operating information systems; however, much of their information was proprietary. Many of them were able to provide information that we feel to be more realistic than that found in the literature, and our estimates of "typical" costs are based heavily on these discussions. We emphasize that, while we consider these cost estimates to be reasonably adequate for broad system comparisons, APA must consider their own operation and each of the alternate subsystems in greater detail, using the cost-effectiveness model as a guideline.

One of the chief problems encountered in both the literature survey and the personal discussions was that cost information was almost always presented in a gross manner and could not be applied to other systems. In order to adapt cost information from one system environment to another, we must be able to isolate fixed costs from costs that vary by size of file, number of searches, number of terms, and average number of items retrieved. Furthermore, we must be able to distinguish the differences in cost of alternative systems and subsystems (e.g., user/system interface, input, hardware, search modes, output screening modes, etc.). The

cost-effectiveness model was derived to accommodate all of these factors. To emphasize the relative importance of these factors and to consider the system in its entirety, we estimated "typical" costs of 36 system and subsystem alternatives.

Finally, if APA decides to charge for retrospective search services, it is necessary to have some knowledge of their economic and marketing implications. Demand for the services can be influenced by price, advertising, promotion, and the quality of the system. Although it is difficult to reliably estimate demand and a price/demand relationship, the model above, plus the examples given, will yield an estimate of the cost/demand relationship. Some inferences are also drawn concerning frequency of use of the system.

II. RETROSPECTIVE SEARCH SYSTEMS

One of the difficulties with deriving a cost model for the design of retrospective search systems is that each system consists of several subsystems that can accomplish different functions by a number of alternative processes. The most important of these subsystems are user/system interface, input, search mode, output screening, and form of presentation to the user. The numerous possible alternative processes are determined by the hardware, software, and general procedures used to accomplish specific functions. System designers must choose among the alternatives with regard to the cost and/or effectiveness of each, but this is sometimes difficult because the systems are so interrelated. Below we have applied a mathematical model that yields measures of cost/effectiveness for several combinations of system alternatives.*

The effectiveness of search system performance can be measured in part by search accuracy. Perfect search accuracy implies that searches yield all relevant documents and no nonrelevant materials. All major retrospective search systems evaluated to date have operated with far from perfect accuracy. In order to apply the model mentioned above, we measure accuracy by recall and fallout. Recall is the proportion of relevant documents retrieved, and fallout is the proportion of nonrelevant documents retrieved.

A large fallout implies high costs, since each retrieved document adds to processing, output, screening, and mailing costs. High recall also involves substantial costs, as shown in the relationship in Figure 1. In most systems it costs substantially more to increase recall an incremental amount at high levels of recall than at low levels of recall. This fact is borne out in the next section.

*Briefly, the model is a finite Markoff chain that treats recall and fallout ratios observed for each subsystem as transition probabilities. The model is described in references 13 and 14.

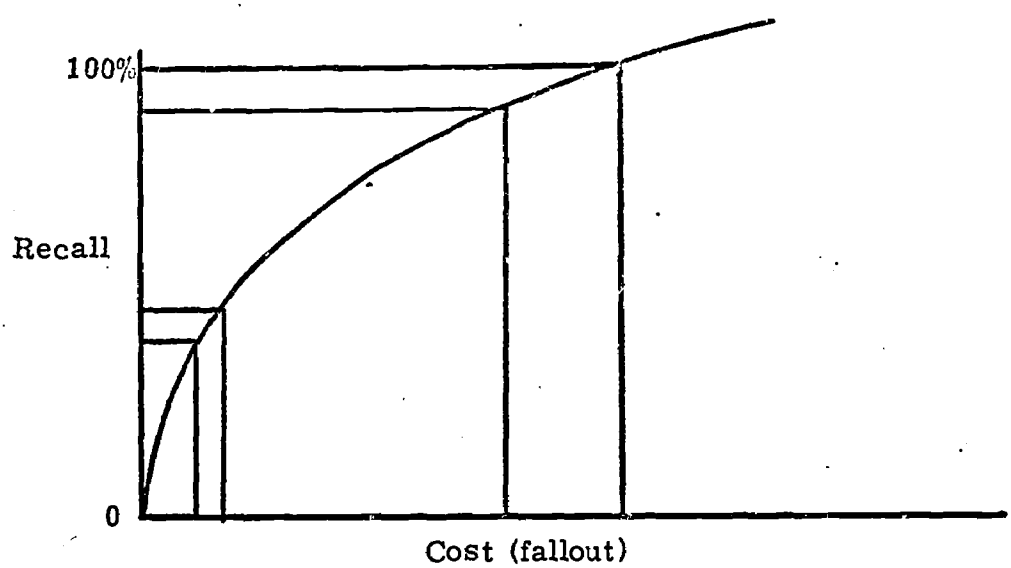


Figure 1. Cost vs. recall for typical systems

Since all subsystems contribute to the cost/effectiveness relationship, all of them should be included when considering the cost/effectiveness of retrospective search systems. Several common alternatives are discussed below.

Most systems use an intermediary to interpret a user's search request and to conduct the search. There are several ways in which the intermediary can communicate with the user to form a user/system interface. Two common ways are by written correspondence or personal contact, as by telephone. A study of the MEDLARS system,[6] which used written request forms, indicated that a number of relevant documents were missed and nonrelevant documents retrieved due to an intermediary not fully understanding the user's search requirements. The proportions of missed relevant and retrieved nonrelevant documents observed in this evaluation were applied in the analysis below. As a second alternative we assumed that personal contact could be made with users, increasing costs but reducing failures by one-half. (This one-half is conjectured, since we could find no actual value for this number in the literature.)

Most large system retrospective searching is performed by computer batch processing or on-line, and input is by indexing from a controlled vocabulary or from full-text of an abstract. Thus, we have essentially three search systems in our analysis: index input searched by batch processing, index input searched by on-line terminal from a controlled vocabulary, and full-text abstract input searched by on-line terminal from free language search queries. Any level of accuracy (recall) can be achieved by the three systems, but with a different number of documents retrieved and, hence, different cost. In order to analyze each system, we established typical retrieval necessary to obtain recall levels of 40%, 60%, 80%, and 100%.[12] These values were incorporated with our analytical models for illustrative purposes.

Many systems with a large number of documents yield a correspondingly large number of documents retrieved so that intermediaries must be used to screen out nonrelevant material. The screening is usually performed on titles or titles and abstracts together. However, even though nonrelevant retrieval is reduced, many relevant documents are also likely to be screened out by mistake. We have taken the results of two experiments to estimate recall and fallout. Results from the first experiment, involving tight screening, indicates that a high proportion of nonrelevant material is screened out, but many relevant documents are also screened out. Evidence from the second experiment, involving loose screening, suggests poorer screening with correspondingly fewer missed relevant documents.

Another system alternative is to send identified documents to the users in the form of titles or of titles and abstracts. Cost versus user convenience is the principal implication here.

In order to show the cost implications of potential APA systems, we applied the analytical model to all combinations (36) of two user/system interfaces, three search modes at four levels of recall, and six screening alternatives. The analysis assumed a total file size of 100,000 documents

(i.e., 25,000 annual input for four years) and fifty total relevant documents in the file. The results of this analysis are given in Table 1 for total retrieved per search and number of relevant documents retrieved per search. It is clear that the number of items retrieved necessary to increase levels of recall becomes extremely large as recall approaches 100% (note unscreened retrieval). For example, total retrieval from batch processing (with written request) doubles from 40% to 60% recall (32 to 64), doubles again from 60% to 80% recall (64 to 126) and triples (126 to 376) when increased from 80% to 100%. Thus, one must retrieve over ten times more documents (32 to 376) in order to increase recall from 40% to 100%. This, of course, has significant cost implications, which are discussed in the next section.

We note that, at 100% search recall, the total number of relevant items retrieved, using written requests, is 48 — two relevant documents are lost. When personal contact is made, the total number of relevant items retrieved is 49 — only one relevant document is missed. One might say that we should screen out the 328 nonrelevant documents (376 minus 48). However, we find that the best (and most expensive) screening process (screening on abstracts) screens out 22 of the relevant documents (48 minus 26) while culling out all but two of the nonrelevant documents (28 minus 26).

The on-line abstracts, on-line index, and batch processing systems are ranked, in that order, with regard to total retrieval necessary to achieve equivalent recall levels. This relationship becomes more pronounced as the level of recall increases. For instance, with a written request and no screening, the total retrieval at 40% recall is 25, 27, and 32 documents, respectively. However, at 100% recall, total retrieval is 175, 213, and 376 documents, respectively. Before one can really decide among the 36 system combinations, costs should be carefully considered, which is the topic of the next two sections.

Table 1. Total retrieval and number of relevant items retrieved by alternative system designs and levels of recall from searching

A. No Screening

Recall Level		Written Request			Telephone Request		
		Batch Processing	On-line Index	On-line Abstract	Batch Processing	On-line Index	On-line Abstract
.40	Retrieved	32	27	25	32	27	26
	Relevant	19	19	19	20	20	20
.60	Retrieved	64	51	42	65	52	43
	Relevant	29	29	29	29	29	29
.80	Retrieved	126	101	80	127	102	81
	Relevant	38	38	38	39	39	39
1.00	Retrieved	376	213	175	378	215	177
	Relevant	48	48	48	49	49	49

Table 1. Total retrieval and number of relevant items retrieved by alternative system designs and levels of recall from searching

B. Tight Screen on Titles

Recall Level		Written Request			Telephone Request		
		Batch Processing	On-line Index	On-line Abstract	Batch Processing	On-line Index	On-line Abstract
.40	Retrieved	3	3	3	3	3	3
	Relevant	3	3	3	3	3	3
.60	Retrieved	5	5	5	5	5	5
	Relevant	5	5	5	5	5	5
.80	Retrieved	7	7	7	7	7	7
	Relevant	6	6	6	7	7	7
1.00	Retrieved	10	9	9	9	9	9
	Relevant	8	8	8	8	8	8

C. Tight Screen on Abstracts

Recall Level		Written Request			Telephone Request		
		Batch Processing	On-line Index	On-line Abstract	Batch Processing	On-line Index	On-line Abstract
.40	Retrieved	10	10	10	11	11	11
	Relevant	10	10	10	11	11	11
.60	Retrieved	16	16	15	16	16	16
	Relevant	15	15	15	16	16	16
.80	Retrieved	21	21	21	22	21	21
	Relevant	21	21	21	21	21	21
1.00	Retrieved	28	27	26	28	27	27
	Relevant	26	26	26	26	26	26

Table 1. Total retrieval and number of relevant items retrieved by alternative system designs and levels of recall from searching

D. Loose Screen on Titles

Recall Level		Written Request			Telephone Request		
		Batch Processing	On-line Index	On-line Abstract	Batch Processing	On-line Index	On-line Abstract
.40	Retrieved	25	21	20	24	21	21
	Relevant	17	17	17	17	17	17
.60	Retrieved	45	38	32	45	38	33
	Relevant	25	25	25	26	26	26
.80	Retrieved	82	68	56	83	69	57
	Relevant	33	33	33	34	34	34
1.00	Retrieved	224	134	113	225	135	114
	Relevant	42	42	42	43	43	43

E. Loose Screen on Abstracts

Recall Level		Written Request			Telephone Request		
		Batch Processing	On-line Index	On-line Abstract	Batch Processing	On-line Index	On-line Abstract
.40	Retrieved	22	20	19	23	21	20
	Relevant	16	16	16	17	17	17
.60	Retrieved	41	35	31	42	36	32
	Relevant	25	25	25	25	25	25
.80	Retrieved	75	63	53	76	63	54
	Relevant	33	33	33	34	34	34
1.00	Retrieved	198	120	102	199	121	104
	Relevant	41	41	41	42	42	42

III. COST MODEL

After a careful examination of many different retrospective search systems, a generalized cost model was developed, which will permit APA to evaluate future, as well as present, systems.

The total cost of any given retrospective search system is composed of three types of costs:

1. fixed costs associated with each subsystem,
2. variable costs dependent on the number of items input to the system, and
3. variable costs dependent on the number of searches conducted.

Simply stated

$$C = C' + C''X_1 + C'''X_2$$

Taking these three separately, APA can analyze component costs of any system or subsystem with regard to size of the file (e.g., APA primary journals versus world's psychological literature), range of demand or system usage, and size of term list.

Fixed Costs

There are fixed costs associated with each subsystem. These include such items as staff, space rental, computer rental, and fixed computer storage changes (C_1). Once a commitment is made for a specific computerized search system, these costs will be incurred, even if the system is not used at all.

Another fixed cost element is the rent, staff, and screening devices (C_2) that may be used to display the full-text of an abstract or document representation in screening search output. Fixed costs associated with input (C_3) include such items as thesaurus development, staff, tape conversion, and update costs. Other fixed costs are staff, rent, and sundry items associated

with user/system interface (C_4) and mailing search output to the users (C_5) The fixed cost element is then

$$C' = C_1 + C_2 + C_3 + C_4 + C_5$$

Variable Costs

The variable costs that are dependent on file size or number of items input to the system (X_1) are composed of a cost per item of indexing, abstracting, keyboarding, and any other input processing. These costs can be allocated among various services that also use the input products (C_6), i.e., Psychological Abstracts, current awareness announcement, retrospective search, recurring bibliographies, and so forth. File loading costs (C_7) include costs that vary with number of terms (X_5). The entire cost component can be expressed as

$$C'' = C_6 + X_5 C_7$$

Another type of variable cost is the cost dependent on the number of searches conducted per year (X_2), or the demand for the retrospective search system. This is the most complicated of the elements of the model, being composed of three parts: fixed costs per search, costs dependent on the number of items retrieved (X_3), and costs dependent on the number of items sent to the user (X_4). The fixed elements of the cost are the set-up costs for mailing titles to users (C_8) and the cost of the user/system interface, i.e., the intermediary (C_9). For our purposes we consider two alternative methods of user/system communication: written and oral requests.

There are three costs dependent on the number of items retrieved in any search (X_3): the computer costs of retrieving (C_{10}) and printing out the item (C_{11}) and the costs of screening each item retrieved (C_{12}). The cost dependent on the number of items mailed per search (X_4) is the cost of actually mailing the titles or abstracts to the user (C_{13}). The entire component can be expressed as

$$C''' = C_8 + C_9 + X_3(C_{10} + C_{11} + C_{12}) + X_4 C_{13}$$

Combining the elements of the cost model, we have

$$C = C_1+C_2+C_3+C_4+C_5+X_1(C_6+X_5C_7)+X_2[C_8+C_9+X_3(C_{10}+C_{11}+C_{12})+X_4C_{13}]$$

where:

X_1 = number of items input
X_2 = number of searches conducted
X_3 = number of items retrieved per search
X_4 = number of items mailed per search
X_5 = number of terms in authority list
C_1 = fixed cost associated with computing
C_2 = fixed cost associated with screening
C_3 = fixed cost associated with input
C_4 = fixed cost associated with user/system interface
C_5 = fixed cost associated with mailing results
C_6 = total input cost per item
C_7 = total file loading cost per item per term
C_8 = fixed cost of mailing per search
C_9 = fixed cost of user/system interface per search
C_{10} = computer retrieval cost per item retrieved
C_{11} = computer printing cost per item retrieved
C_{12} = screening cost per item retrieved
C_{13} = mailing cost per item mailed
C = total annual cost

This general equation can be used to estimate costs of potential NISP search systems as well as to compare the cost/effectiveness trade-off of system alternatives.

IV. COSTS OF ALTERNATIVE SYSTEMS

Nearly all published literature on costs of information retrieval systems has emphasized a lack of good cost-accounting procedures and, therefore, reliable costing information. In all discussions with persons closely involved in many different retrieval systems, the need for accurate, complete, consistent, and available data on costs was acknowledged. From these interviews and literature, we derived "reasonable" cost estimates that are used in the cost model above to estimate the general magnitude of total costs of various alternative retrospective search systems and subsystems.

Examples of typical values are given below for cost items C_1, ..., C_{13} for three general systems: (1) batch processing, index input, controlled vocabulary search, (2) on-line processing, index input, controlled vocabulary search, and (3) on-line processing, abstract input, free language search. It must be emphasized that, while the estimated costs presented are reasonably adequate to establish the general magnitude of system costs, each subsystem should be carefully evaluated with regard to actual cost. These costs can then be compared with those given in this report to see if they are in line. Each cost of a potential system can also be incorporated with the model to determine its effect on total costs over a range of demand and in view of the number of items input.

The costs are broken down into input costs, search costs, screening costs, and mailing costs. The three general search systems costs are directly related to some input costs and most search costs. The system costs are also related to screening and mailing costs, since each of the three systems yields a different number of items retrieved (for a given recall level). The costs associated with the three general systems are given on the following pages.

	System		
	Batch Index	On-line Index	On-line Abstract
INPUT			
C_3-Thesaurus**	$15,000	$15,000	
Staff, other OH**	2,000	2,000	$ 8,000
Update	2,000	5,700	11,400
Tape conversion	3,000	5,000	10,000
C_6-Index & abstract*	.30/item	.30/item	.80/item
Keyboarding	.25/item	.25/item	.875/item
Other processing	.025/item	.0875/item	
C_7-File loading	.00015/item/term	.00015/item/term	.00003/item/term
SEARCH			
C_1-Computer rental	$65,000	$11,000***	$15,000***
Staff, other OH	66,000	12,000	16,500
Terminal rent		1,440	1,440
C_{10}-Computer processing	.05/item retrieved	***	***
C_{11}-Printout costs	.008/item received	.008/item retrieved	.008/item retrieved

*Two-thirds allocated to other processes and amortized over four years

**One-half allocated to other processes

*** The search costs (C_1, C_{10}, C_{11}) for on-line index and on-line abstracts are not given in the table since they depend on additional variables. The costs used for this study came from prices quoted from firms that provided on-line search systems.

Other costs depend on the user/system interface (written requests versus telephone requests). Typical costs for these are given below.

	Written request	Telephone request
SEARCH		
C_4 - staff, rent, other overhead	$2,000	$2,000
C_9 - staff, equipment	$11.25/search	$15.00/search

Screening can be performed on titles or abstracts which yield substantially different costs. Typical costs for these are given below.

	Titles	Abstracts
SCREEN		
C_2 - staff, rent, other overhead	-	$16,000
equipment	-	$16,000*
C_{12} - staff	$.04/item retrieved	$.125/item retrieved

*Amortized over four years

Finally, mailing costs also depend on whether titles or abstracts are sent to the user. Typical mailing costs are given below.

	Titles	Abstracts
MAILING		
C_5 - rent, equipment, staff, other overhead	$1,000	$1,000
C_8 - preparation, staff materials	$.20/search	-
C_{13} - postage, materials	$.002/item sent	$.10/item sent

An example is given to illustrate how costs are derived. Assume the following:

1. System processes include batch processing, telephone request, tight screening on abstracts, and abstracts mailed to users.
2. Number of items input (X_1) is 100,000 over a four year period.
3. Number of searches (X_2) is 4,000 per year.
4. Number of items retrieved per search (X_3) is 126; i.e., the recall level is 80 percent prior to screening (see Table 1).
5. Number of items mailed per search (X_4) is 22; i.e., the recall level at 80 percent and after screening (see Table 1).
6. Number of terms in authority list (X_5) is 1,000.

The model is:

$$C = C_1+C_2+C_3+C_4+C_5+X_1(C_6+X_5 C_7)+X_2[C_8+C_9+X_3(C_{10}+C_{11}+C_{12})+X_4 C_{13}]$$

$$= \$131,000+\$32,000+\$5,500+\$500+\$250$$
$$+100,000[\$.575+(1000)(\$0.0000375)]$$
$$+4000[0+\$15.00+126(\$.05+\$.008+\$1.25)+(22)(\$.10)]$$

$$= \$169,250+\$61,250+\$161,032$$

$$= \$391,532$$

The average cost per search is then \$97.90 (\$391,532/4000), cost per retrieved document is \$4.40 (\$97.90/22), and cost per relevant item retrieved is \$4.60 (\$97.90/21).

Using the cost model, the following costs were calculated as described in the example above for demand levels of 1,000 and 4,000. It is emphasized again that these cost estimates are only approximations derived from scanty information available from a variety of sources. However, these estimates should be very useful for gross design comparisons, for determining the effect of such factors as demand on the design comparisons, and to serve as a benchmark to assess potential new systems.

Table 2 below gives the cost per search, cost per item retrieved and cost per relevant-item retrieved for 1,000 requests for six alternative systems. Table 3 gives the same information for 4,000 requests. It appears that on-line index is consistently the best search system with regards to all three cost measures.* As the search recall level increases, the batch processing system becomes more competitive. A recall level of 80% is probably the most feasible. Of course 100% recall is more desirable, but the incremental costs necessary to achieve the level may be too great.

When comparing written request against telephone request, we find that in some instances written requests may cost less for the search but cost more for each relevant item retrieved (e.g., see Table 2A at .80 recall). However, the differences are so small that the distinction may not be important. The same holds true for mailing titles or abstracts to users so that one may choose the latter merely to provide a better service to the users.

It is noted that tight screening does not cost much more per search than loose screening or no screening but that it does cost much more per item retrieved and per relevant item retrieved. Also, the cost per search of tight screening on abstracts appears to be considerably greater than tight screening on titles while the cost per item retrieved and per relevant item retrieved is less. This is due to the fact that tight screening greatly reduces the number of relevant as well as nonrelevant documents retrieved. The cost of loose screening on titles is consistently less than loose screening on abstracts.

As might be expected, loose screening on titles costs very little more than no screening while loose screening on abstracts does cost some more.

*One should not generalize this comparison. For example, batch processing search costs can be reduced considerably if fixed costs (C_1) are allocated to a broader range of other services or reduced by using only off-shift time.

Table 2. Cost per search, per item retrieved and per relevant item retrieved for alternative retrospective search systems, subsystems, and recall levels (Demand = 1000)

A. No Screening, Mail Titles to User

Recall Level	Cost ($) For Each	Written Request			Telephone Request		
		Batch Processing	On-line Index	On-line Abstract	Batch Processing	On-line Index	On-line Abstract
.40	Search	212.00	117.00	252.00	216.00	121.00	256.00
	Item Retrieved	6.60	4.30	1.10	6.70	4.50	9.80
	Relevant Retrieved	11.20	6.2	13.30	10.80	6.00	12.80
.60	Search	213.00	117.00	252.00	217.00	121.00	256.00
	Item Retrieved	3.30	2.30	6.00	3.30	2.30	6.00
	Relevant Retrieved	7.40	4.00	8.70	7.50	4.20	8.80
.80	Search	216.00	135.00	252.00	220.00	138.00	256.00
	Item Retrieved	1.70	1.30	3.20	1.70	1.40	3.20
	Relevant Retrieved	5.70	3.50	6.60	5.60	3.50	6.60
1.00	Search	230.00	152.00	279.00	233.00	156.00	282.00
	Item Retrieved	.60	.70	1.60	.60	.70	1.60
	Relevant Retrieved	4.80	3.20	5.80	4.80	3.20	5.80

Table 2. Cost per search, per item retrieved and per relevant item retrieved for alternative retrospective search systems, subsystems, and recall levels (Demand = 1000)

B. No Screening, Mail Abstracts to User

Recall Level	Cost ($) For Each	Written Request			Telephone Request		
		Batch Processing	On-line Index	On-line Abstract	Batch Processing	On-line Index	On-line Abstract
.40	Search	215.00	119.00	254.00	219.00	123.00	258.00
	Item Retrieved	6.70	4.40	10.20	6.80	4.60	9.90
	Relevant Retrieved	11.30	6.20	13.40	11.50	6.20	12.90
.60	Search	219.00	122.00	256.00	223.00	125.00	260.00
	Item Retrieved	3.40	2.40	6.10	3.40	2.40	6.00
	Relevant Retrieved	7.60	4.20	8.80	7.70	4.30	9.00
.80	Search	228.00	144.00	260.00	232.00	148.00	264.00
	Item Retrieved	1.80	1.40	3.30	1.80	1.50	3.30
	Relevant Retrieved	6.00	3.80	6.80	6.00	3.80	6.80
1.00	Search	266.00	173.00	296.00	237.00	177.00	299.00
	Item Retrieved	.70	.80	1.70	.60	.80	1.70
	Relevant Retrieved	5.60	3.60	6.20	4.80	3.60	6.10

Table 2. Cost per search, per item retrieved and per relevant item retrieved for alternative retrospective search systems, subsystems, and recall levels (Demand = 1000)

C. Tight Screen on Titles, Mail Titles to User

Recall Level	Cost ($) For Each	Written Request			Telephone Request		
		Batch Processing	On-line Index	On-line Abstract	Batch Processing	On-line Index	On-line Abstract
.40	Search	214.00	119.00	255.00	217.00	123.00	258.00
	Item Retrieved	71.20	39.90	84.90	72.40	41.10	86.20
	Relevant Retrieved	71.20	39.90	84.90	72.40	41.10	86.20
.60	Search	216.00	122.00	257.00	220.00	126.00	260.00
	Item Retrieved	43.20	24.50	51.30	43.90	25.20	52.10
	Relevant Retrieved	43.20	24.50	51.30	43.90	25.20	52.10
.80	Search	222.00	145.00	261.00	226.00	149.00	265.00
	Item Retrieved	31.70	20.80	37.30	32.20	21.30	37.80
	Relevant Retrieved	37.00	24.20	43.50	32.20	21.30	37.80
1.00	Search	247.00	175.00	310.00	251.00	179.00	314.00
	Item Retrieved	24.70	19.40	34.50	27.90	19.90	34.90
	Relevant Retrieved	30.90	21.90	38.80	31.40	22.40	39.30

Table 2. Cost per search, per item retrieved and per relevant item retrieved for alternative retrospective search systems, subsystems, and recall levels (Demand = 1000)

D. Tight Screen on Abstracts, Mail Abstracts to User

Recall Level	Cost ($) For Each	Written Request			Telephone Request		
		Batch Processing	On-line Index	On-line Abstract	Batch Processing	On-line Index	On-line Abstract
.40	Search	249.00	154.00	289.00	253.00	158.00	293.00
	Item Retrieved	24.90	15.40	29.00	23.00	14.40	26.70
	Relevant Retrieved	24.90	15.40	29.00	23.00	14.40	26.70
.60	Search	254.00	159.00	293.00	258.00	164.00	297.00
	Item Retrieved	15.90	10.00	19.50	16.10	10.20	18.60
	Relevant Retrieved	17.00	10.60	19.50	16.10	10.20	18.60
.80	Search	266.00	188.00	302.00	270.00	191.00	306.00
	Item Retrieved	12.70	8.90	14.40	12.30	9.10	14.60
	Relevant Retrieved	12.70	8.90	14.40	12.80	9.10	14.60
1.00	Search	312.00	227.00	361.00	315.00	231.00	365.00
	Item Retrieved	11.10	8.40	13.90	11.30	8.60	13.50
	Relevant Retrieved	12.00	8.70	13.90	12.10	8.90	14.00

Table 2. Cost per search, per item retrieved and per relevant item retrieved for alternative retrospective search systems, subsystems, and recall levels (Demand = 1000)

E. Loose Screen on Titles, Mail Titles to User

Recall Level	Cost ($) For Each	Written Request			Telephone Request		
		Batch Processing	On-line Index	On-line Abstract	Batch Processing	On-line Index	On-line Abstract
.40	Search	213.00	119.00	255.00	217.00	123.00	258.00
	Item Retrieved	8.90	5.70	12.70	9.10	5.90	12.30
	Relevant Retrieved	12.80	7.20	15.30	12.80	7.20	15.20
.60	Search	216.00	122.00	257.00	220.00	126.00	260.00
	Item Retrieved	4.80	3.20	8.00	4.90	3.30	7.90
	Relevant Retrieved	8.70	4.90	10.30	8.90	4.90	10.20
.80	Search	222.00	145.00	261.00	226.00	149.00	265.00
	Item Retrieved	2.70	2.10	4.70	2.70	2.20	4.60
	Relevant Retrieved	6.70	4.40	7.90	6.60	4.40	7.80
1.00	Search	247.00	176.00	310.00	251.00	179.00	314.00
	Item Retrieved	1.10	1.30	2.80	1.10	1.30	2.80
	Relevant Retrieved	6.00	4.20	7.50	5.90	8.80	7.40

Table 2. Cost per search, per item retrieved and per relevant item retrieved for alternative retrospective search systems, subsystems, and recall levels (Demand = 1000)

F. Loose Screen on Abstracts, Mail Abstracts to User

Recall Level	Cost ($) For Each	Written Request: Batch Processing	Written Request: On-line Index	Written Request: On-line Abstract	Telephone Request: Batch Processing	Telephone Request: On-line Index	Telephone Request: On-line Abstract
.40	Search	250.00	155.00	290.00	254.00	159.00	294.00
	Item Retrieved	11.40	7.80	15.30	11.00	7.60	14.70
	Relevant Retrieved	15.20	9.50	17.70	15.10	9.50	17.50
.60	Search	257.00	162.00	295.00	261.00	166.00	299.00
	Item Retrieved	6.30	4.60	9.50	6.20	4.60	9.30
	Relevant Retrieved	10.40	6.60	12.00	10.30	6.60	11.80
.80	Search	271.00	192.00	305.00	275.00	196.00	309.00
	Item Retrieved	3.60	3.10	5.80	3.60	3.10	5.70
	Relevant Retrieved	8.30	5.90	9.30	8.00	5.80	9.20
1.00	Search	329.00	237.00	368.00	332.00	240.00	372.00
	Item Retrieved	1.70	2.00	3.60	1.70	2.00	3.60
	Relevant Retrieved	8.00	5.80	9.00	7.90	5.70	8.90

Table 3. Cost per search, per item retrieved and per relevant item retrieved for alternative retrospective search systems, subsystems, and recall levels (Demand = 4000)

A. No Screening, Mail Titles to User

Recall Level	Cost ($) For Each	Written Request			Telephone Request		
		Batch Processing	On-line Index	On-line Abstract	Batch Processing	On-line Index	On-line Abstract
.40	Search	63.00	42.00	78.00	67.00	46.00	82.00
	Item Retrieved	2.00	1.60	3.10	2.00	1.70	3.20
	Relevant Retrieved	3.30	2.20	4.10	3.30	2.30	4.10
.60	Search	64.00	47.00	78.00	68.00	50.00	82.00
	Item Retrieved	1.00	.90	1.90	1.10	1.00	1.90
	Relevant Retrieved	2.30	1.70	2.70	2.40	1.70	2.80
.80	Search	68.00	55.00	92.00	71.00	59.00	95.00
	Item Retrieved	.50	.50	1.20	.60	.60	1.20
	Relevant Retrieved	1.80	1.50	2.40	1.80	1.50	2.50
1.00	Search	81.00	78.00	118.00	84.00	82.00	122.00
	Item Retrieved	.20	.40	.70	.20	.40	.70
	Relevant Retrieved	1.70	1.60	2.50	1.70	1.70	2.50

Table 3. Cost per search, per item retrieved and per relevant item retrieved for alternative retrospective search systems, subsystems, and recall levels (Demand = 4000)

B. No Screening, Mail Abstracts to User

Recall Level	Cost ($) For Each	Written Request			Telephone Request		
		Batch Processing	On-line Index	On-line Abstract	Batch Processing	On-line Index	On-line Abstract
.40	Search	66.00	45.00	81.00	70.00	48.00	85.00
	Item Retrieved	2.00	1.70	3.20	2.20	1.80	3.30
	Relevant Retrieved	3.50	2.40	4.30	3.50	2.40	4.20
.60	Search	71.00	51.00	83.00	74.00	55.00	86.00
	Item Retrieved	1.10	1.00	2.00	1.10	1.10	2.00
	Relevant Retrieved	2.50	1.80	2.90	2.60	1.90	3.00
.80	Search	80.00	65.00	99.00	84.00	69.00	103.00
	Item Retrieved	.60	.60	1.20	.70	.70	1.30
	Relevant Retrieved	2.10	1.70	2.60	2.10	1.80	2.70
1.00	Search	117.00	99.00	136.00	121.00	103.00	139.00
	Item Retrieved	.30	.50	.80	.30	.50	.80
	Relevant Retrieved	2.40	2.10	2.80	2.50	2.10	2.90

Table 3. Cost per search, per item retrieved and per relevant item retrieved for alternative retrospective search systems, subsystems, and recall levels (Demand = 4000)

C. Tight Screen on Titles, Mail Titles to User

Recall Level	Cost ($) For Each	Written Request			Telephone Request		
		Batch Processing	On-line Index	On-line Abstract	Batch Processing	On-line Index	On-line Abstract
.40	Search	64.00	45.00	81.00	68.00	49.00	85.00
	Item Retrieved	21.50	15.00	27.00	22.70	16.20	28.30
	Relevant Retrieved	21.50	15.00	27.00	22.70	16.20	28.30
.60	Search	67.00	52.00	83.00	71.00	56.00	87.00
	Item Retrieved	13.50	10.40	16.60	14.20	11.10	17.30
	Relevant Retrieved	13.50	10.40	16.60	14.20	11.10	17.30
.80	Search	73.00	66.00	100.00	77.00	70.00	104.00
	Item Retrieved	10.50	9.50	14.30	11.00	10.00	14.90
	Relevant Retrieved	12.20	11.70	16.70	11.00	10.00	14.90
1.00	Search	98.00	101.00	150.00	102.00	105.00	154.00
	Item Retrieved	9.80	12.60	16.70	11.30	11.60	17.10
	Relevant Retrieved	12.20	12.60	18.80	12.70	13.10	19.20

Table 3. Cost per search, per item retrieved and per relevant item retrieved for alternative retrospective search systems, subsystems, and recall levels (Demand = 4000)

D. Tight Screen on Abstracts, Mail Abstracts to User

Recall Level	Cost ($) For Each	Written Request			Telephone Request		
		Batch Processing	On-line Index	On-line Abstract	Batch Processing	On-line Index	On-line Abstract
.40	Search	76.00	56.00	92.00	80.00	60.00	96.00
	Item Retrieved	7.60	5.60	9.20	7.20	5.40	8.70
	Relevant Retrieved	7.60	5.60	9.20	7.20	5.40	8.70
.60	Search	82.00	65.00	95.00	86.00	85.00	99.00
	Item Retrieved	5.10	4.10	6.40	5.30	5.30	6.20
	Relevant Retrieved	5.50	4.40	6.40	5.30	5.30	6.20
.80	Search	93.00	85.00	117.00	98.00	89.00	121.00
	Item Retrieved	4.40	4.00	5.60	4.40	4.20	5.80
	Relevant Retrieved	4.40	4.00	5.60	4.70	4.20	5.80
1.00	Search	138.00	129.00	162.00	142.00	133.00	166.00
	Item Retrieved	5.00	4.80	6.20	5.10	4.90	6.20
	Relevant Retrieved	5.00	5.00	6.20	5.50	5.10	6.40

Table 3. Cost per search, per item retrieved and per relevant item retrieved for alternative retrospective search systems, subsystems, and recall levels (Demand = 4000)

E. Loose Screen on Titles, Mail Titles to User

Recall Level	Cost ($) For Each	Written Request			Telephone Request		
		Batch Processing	On-line Index	On-line Abstract	Batch Processing	On-line Index	On-line Abstract
.40	Search	64.00	45.00	81.00	68.00	49.00	85.00
	Item Retrieved	2.70	2.10	4.10	2.80	2.30	4.00
	Relevant Retrieved	3.90	2.70	4.90	4.00	2.90	5.00
.60	Search	67.00	52.00	83.00	71.00	56.00	87.00
	Item Retrieved	1.50	1.40	2.60	1.60	1.50	2.60
	Relevant Retrieved	2.70	2.10	3.30	2.80	2.20	3.40
.80	Search	74.00	66.00	100.00	77.00	70.00	104.00
	Item Retrieved	.90	1.00	1.80	.90	1.00	1.80
	Relevant Retrieved	2.20	2.00	3.00	2.30	2.10	3.10
1.00	Search	98.00	100.00	150.00	102.00	105.00	154.00
	Item Retrieved	.40	.80	1.30	.50	.80	1.40
	Relevant Retrieved	2.40	2.40	3.60	2.40	2.50	3.60

Table 3. Cost per search, per item retrieved and per relevant item retrieved for alternative retrospective search systems, subsystems, and recall levels (Demand = 4000)

F. Loose Screen on Abstracts, Mail Abstracts to User

Recall Level	Cost ($) For Each	Written Request			Telephone Request		
		Batch Processing	On-line Index	On-line Abstract	Batch Processing	On-line Index	On-line Abstract
.40	Search	77.00	57.00	93.00	81.00	61.00	97.00
	Item Retrieved	3.50	2.80	4.90	3.50	2.90	4.80
	Relevant Retrieved	4.70	3.50	5.70	4.80	3.60	5.70
.60	Search	84.00	67.00	97.00	88.00	71.00	101.00
	Item Retrieved	2.10	1.90	3.10	2.10	2.00	3.20
	Relevant Retrieved	3.40	2.70	4.00	3.50	2.80	4.00
.80	Search	99.00	89.00	120.00	103.00	93.00	124.00
	Item Retrieved	1.30	1.40	2.30	1.40	1.50	2.30
	Relevant Retrieved	3.00	2.70	3.70	3.10	2.80	3.70
1.00	Search	155.00	138.00	170.00	159.00	142.00	174.00
	Item Retrieved	.80	1.20	1.70	.80	1.20	1.70
	Relevant Retrieved	3.80	3.40	4.10	3.80	3.40	4.10

One important consideration is whether or not the cost relationships among different systems hold over different levels of demand. For example, if on-line index/written request/screen on titles/mail titles is better than on-line abstract/written request/screen on abstracts/mail abstracts at 1,000 requests, is it also superior at 4,000 requests? Secondly, if it is better with regard to total search cost, cost per item retrieved or cost per relevant item retrieved at 1,000 requests, is it also better with regard to these measures at 4,000 requests? Even though there are small differences between systems at 1,000 and 4,000 requests, the differences are not appreciable enough to be of real concern. Also, the relationships observed for total retrieval, number of items retrieved and number of relevant items retrieved remain essentially the same for 1,000 and 4,000 requests.

To determine a range of level of demand that APA might be able to expect for a retrospective search system, we investigated similar systems. Two systems that do not charge for use of the retrospective search system (DDC and MEDLARS) report demands of 24,000 and 12,000 respectively. National Technical Information Service (formerly CFSTI), with a demand of 8,000, charges a nominal amount for its search service. SIE discovered that, when it began to charge $35.00 for an unscreened search, demand dropped from 8,000 to 4,000.

As shown in Table 4, cost is relatively insensitive to demand when demand is greater than 4,000. Table 4 also shows that the cost to APA of the retrospective search system is such that, if APA wants a self-supporting system, it will have to charge a price which, according to the previously cited experience, will keep the demand within a range of 1,000 to 4,000. In this demand range, the least inexpensive system — on-line index with written requests, no screening, and mailing of titles — remains the least expensive.

As described in the example above, some of the input costs were allocated to various other services. For the costs presented in Tables 2 and 3, 1/3 of the indexing costs and 1/2 of the thesaurus developmental costs were allocated to the retrospective search system. To investigate the sensitivity of the costs to

Table 4. Cost per search with an allocation of 1/3 of the indexing costs and 1/2 of the thesaurus development costs to the retrospective search system

Number of Searches	Fixed Cost C_1+C_2	Fixed Cost (allocated) $C_3+C_4+C_5$	Variable Costs		Total Cost	Total Cost per Search
			C''	C'''		
500	163,000	6250	61,250	20,129	250,629	501.26
1,000	163,000	6250	61,250	40,258	270,758	270.76
2,000	163,000	6250	61,250	80,516	311,016	155.51
3,000	163,000	6250	61,250	120,774	351,274	117.09
4,000	163,000	6250	61,250	161,032	391,532	97.88
5,000	163,000	6250	61,250	201,290	431,790	86.36
6,000	163,000	6250	61,250	241,548	472,048	78.67
10,000	163,000	6250	61,250	402,580	633,080	63.31

Cost per search with an allocation of all of the indexing and thesaurus development costs to the retrospective search system

Number of Searches	Fixed Cost C_1+C_2	Fixed Cost (allocated) $C_3+C_4+C_5$	Variable Costs		Total Cost	Total Cost per Search
			C''	C'''		
500	163,000	9750	121,250	20,129	314,129	628.26
1,000	163,000	9750	121,250	40,258	334,258	334.26
2,000	163,000	9750	121,250	80,516	374,516	187.26
3,000	163,000	9750	121,250	120,774	414,774	138.26
4,000	163,000	9750	121,250	161,032	455,032	113.76
5,000	163,000	9750	121,250	201,290	495,290	99.06
6,000	163,000	9750	121,250	241,548	535,548	89.26
10,000	163,000	9750	121,250	402,580	696,580	69.66

different allocations of these fixed cost elements, the cost per search for this system* at various levels of demand was calculated for two different allocations of the fixed costs. The first allocation is that described above and used in Tables 2 and 3 for comparisons with other system alternatives. In the second, all of the indexing and thesaurus developmental costs were allocated to the retrospective search system. As shown in Table 4, the costs per search are relatively insensitive to this change in allocation.

It is clear that APA information system resides in a marketlike environment and that all of the economic and marketing implications of this environment must be considered. It is also clear, however, that the distribution and sale of information products and services is not like most marketing environments in that these products and services are interrelated and the functions involved in article transfer may be performed in many ways. APA will be faced with a number of decisions concerning marketing of new services and modification of the old. These decisions include questions of pricing, promotion and advertising policies, and channels of distribution; and they must be based on considerations of cost, income, demand, and the effect of the decisions on other components of the system (this discussion is excerpted from reference 15).

The schema in Figure 2 depicts the functions and processes of the APA system in a marketing environment. It is shown that processes necessary to accomplish the composition, reproduction, acquisition and storage, identification, location, and presentation functions lead to improvements in such things as accessibility, quality, accuracy, speed, and timeliness of article transfer from authors to users. These improvements are made in order to increase user satisfaction, which in turn motivates the psychologists to use the system. This motivation, however, is also partially determined by the price one must pay to use the system and by promotion, sales, and advertising procedures. The price the psychologist must pay involves not only APA's charges for its products and services but also what he must pay in his own time. For example, if a retrospective search results in 5,000 identified titles, he is not likely to be

*tch processing, telephone request, tight screening on abstracts and abstracts mailed to users.

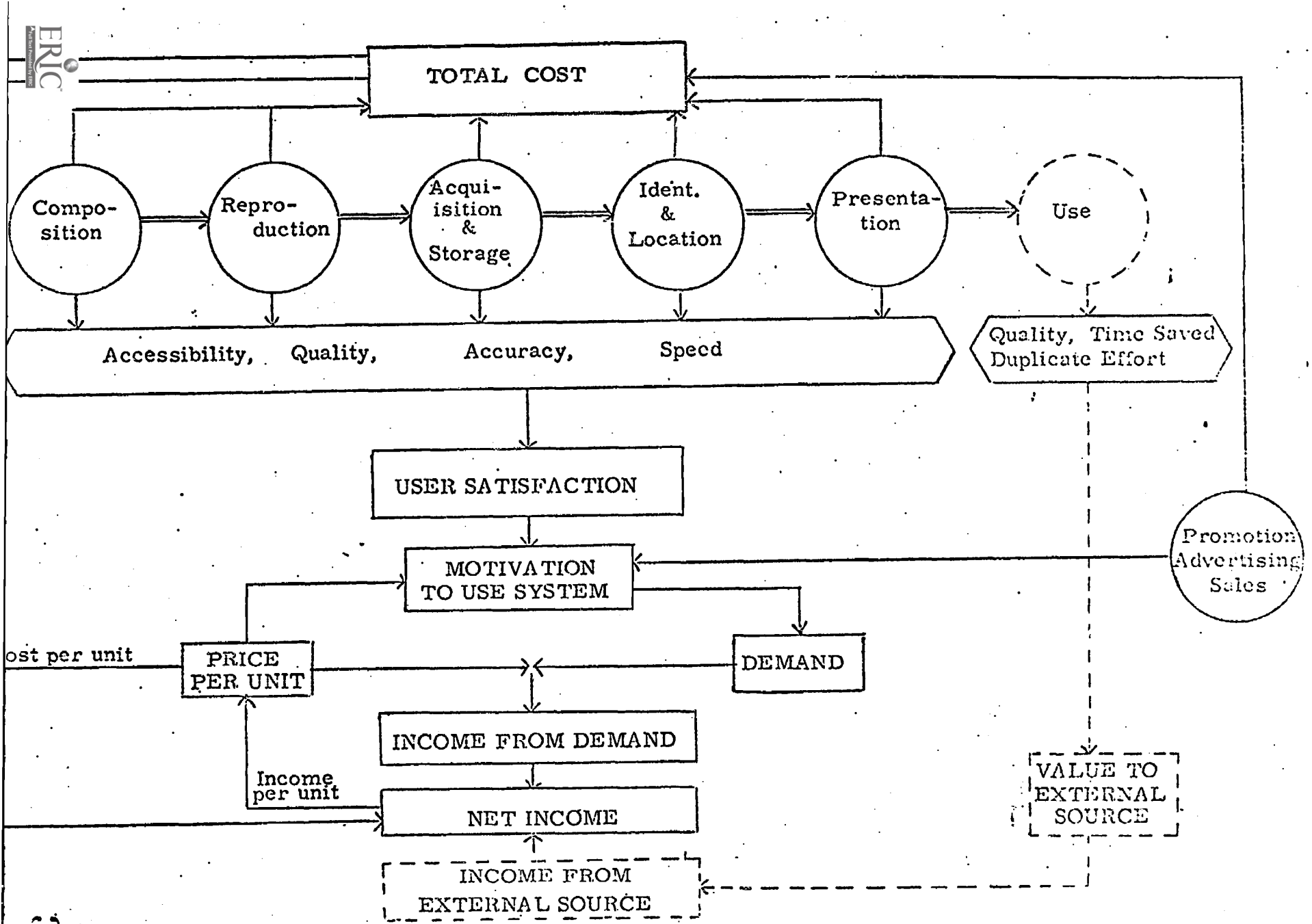


Figure 2. Relationships among functions and measures when users are not a part of the system being evaluated

satisfied since he must pay such a high price in his own time to screen out those documents which do not interest him.

In order to make decisions concerning marketing factors, APA should design and implement an internal costing system, an example of which is given by Helmkamp.[3] This system must be able to identify unit costs that can be subdivided into fixed and variable costs. Information products and services typically have a high fixed cost and relatively low marginal cost. For example, the fixed cost for producing a journal article may be as illustrated in Figure 3.

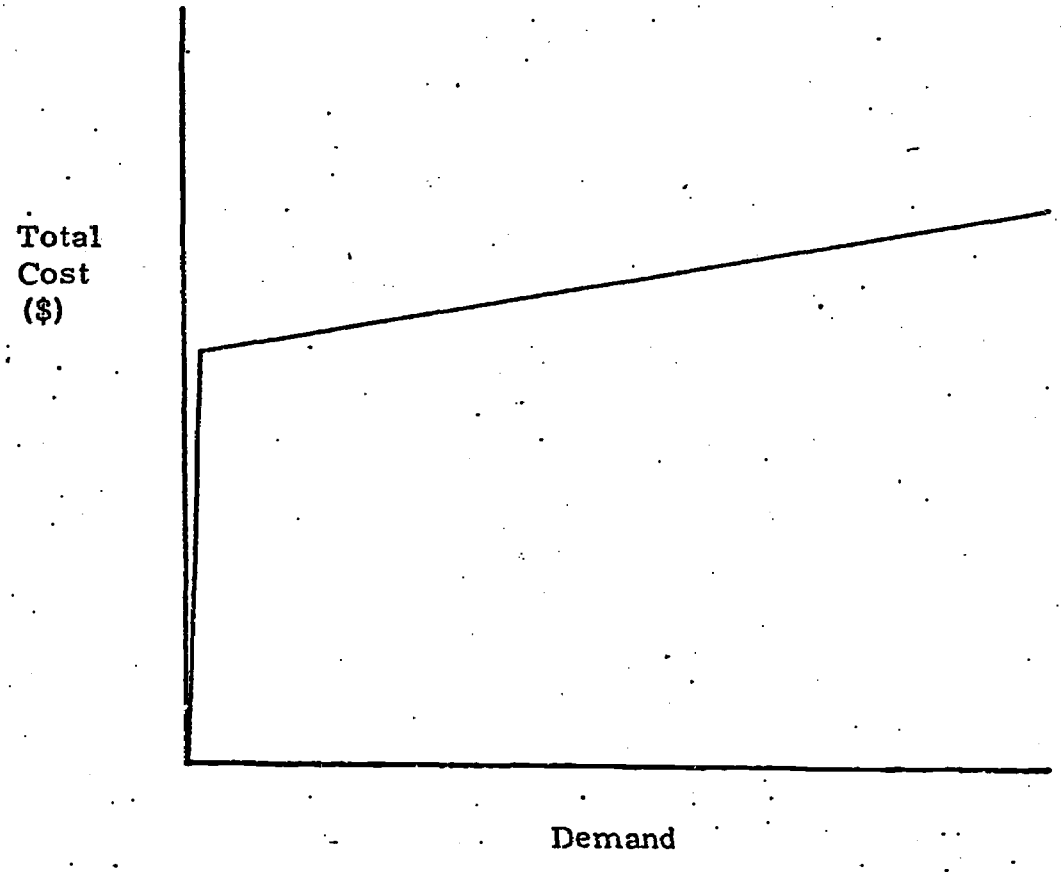


Figure 3. Typical cost curve for information products and services

When one plots the marginal cost against quantity or usage, the curve drops as shown in Figure 4.

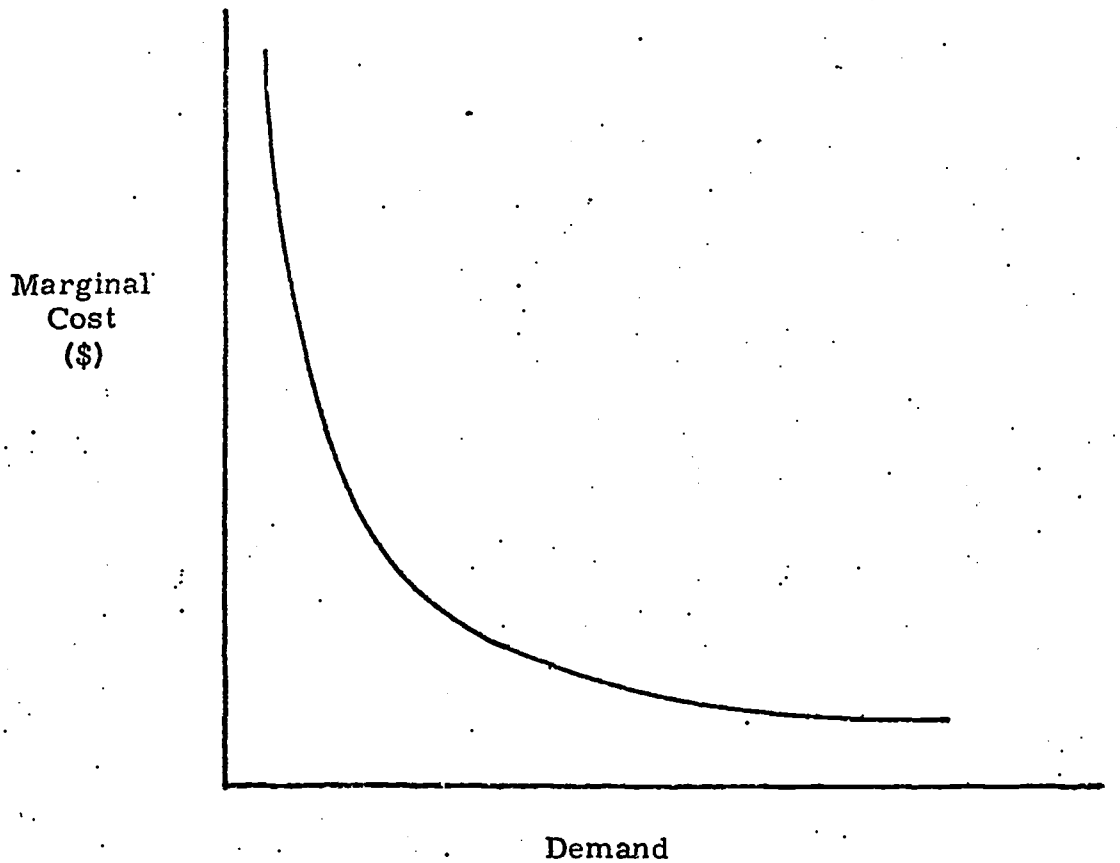


Figure 4. Typical marginal cost curve for information products and services

It is important to establish marginal cost over a likely range of demand for each of the products and services.

APA's cost accounting system must also be able to identify direct and indirect cost, where the indirect cost involves such items as administration and overhead. Furthermore, it is necessary to isolate indirect costs, such as the preparation of magnetic tapes for computer controlled photocomposition, so that these costs can be allocated to all of the derivative products that will come from these tapes. This cost will be discussed later with individual products and services.

Income is determined by the cost of producing the information products and services and the income derived from demand. The income derived from demand is found by multiplying demand by price per unit. However, since the information products and services provided by APA yield a direct value to society as a whole and not just to individual users, there is justification for society's partially funding these important operations through such means as the National Science Foundation. This kind of funding can best be accomplished through providing research and developmental capital in order to get a system operational, at which point the system can be self-sustaining. It is clear that a system such as the one envisioned at APA is not likely to be developed by a private organization since the capital outlay would extend over a long period and the return on investment would probably not accrue in a sufficient time to make the return worthwhile.

As indicated in Figure 2, demand is determined by the influence of the services themselves, promotion, and price. The relative importance of these factors depends largely on the characteristics of the market for the information products and services. There are two classes of market that APA will serve: individuals and institutions. Each of these two classes has substantially different resources available for purchasing APA's services For example, an individual subscriber may be able to spend only $50 to $100 per year, whereas an institution may spend anywhere from $1,000 to $20,000 per year. This means that the two markets may present substantially different demand curves. One would expect the institutional market to have a relatively inelastic (demand not highly sensitive to changes in price) demand curve, as shown in Figure 5.

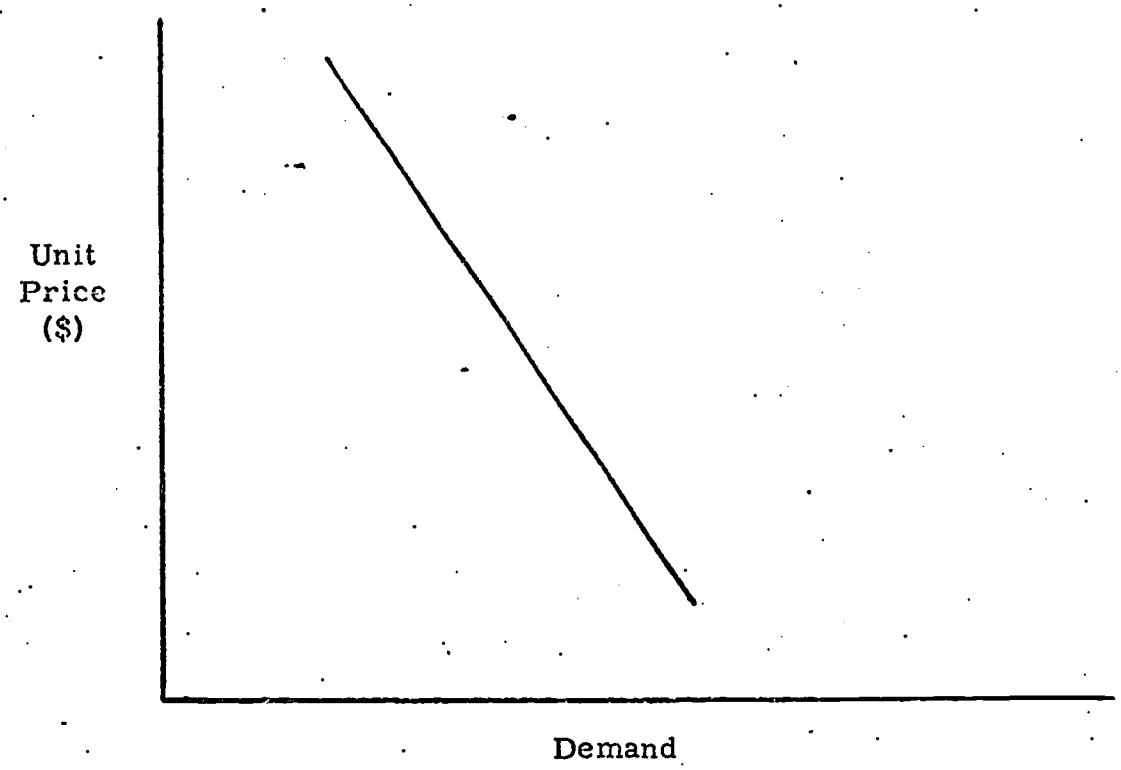
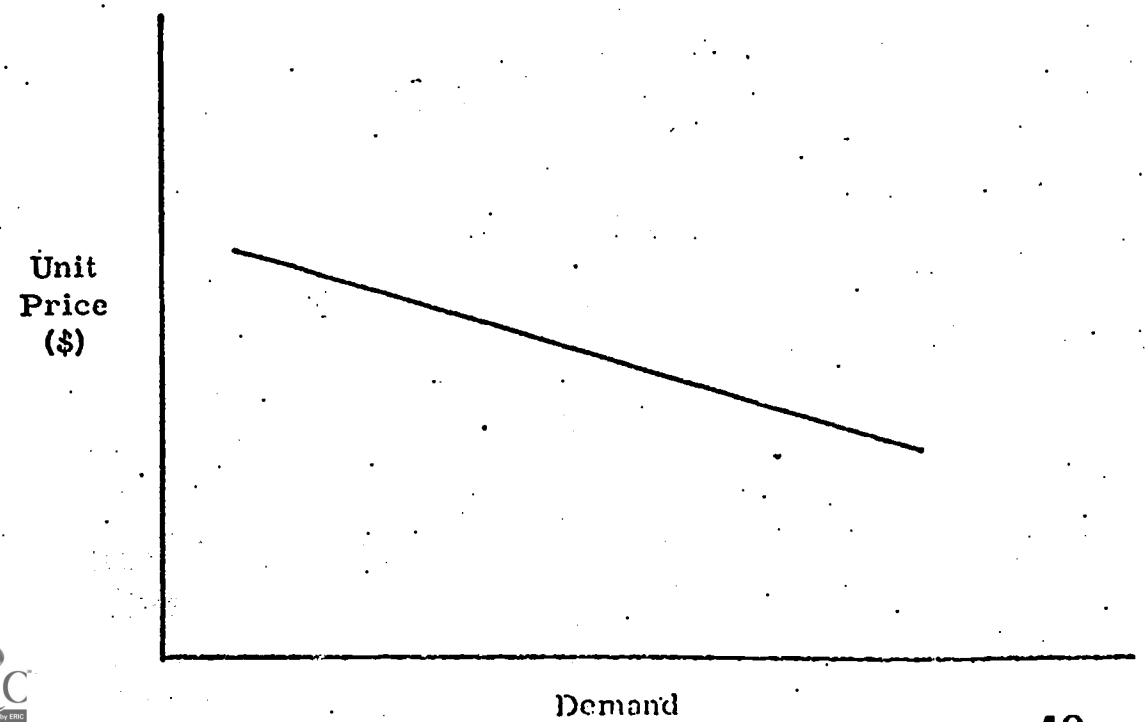


Figure 5. Typical inelastic price/demand curve

On the other hand, the market consisting of individual psychologists probably would have an elastic (highly sensitive to price changes) demand curve, as shown in Figure 6.



This means that APA will have to define carefully the market for each of its products and services and establish a corresponding pricing policy. Fixed and direct cost might be allocated as a component of the price in such a way that a major portion is allocated to those products and services that have an inelastic demand and the remainder to those that have an elastic demand.

BIBLIOGRAPHY

Reports and Articles

1. Bourne, Charles P. and Ford, Donald F. "Cost Analysis and Simulation Procedures for the Evaluation of Large Information Systems." American Documentation, April 1964.

2. Freeman, Monroe E. "Determining Costs of Information Systems." Journal of Chemical Documentation, Vol. 7, No. 2 (1967).

3. Helmkamp, John G. Managerial Cost Accounting for a Technical Information Center. Bloomington, Ind.: Indiana University, 1968.

4. Kidd, E.M. et al.. Study of the Data Central System for Information Retrieval Applied to NSA Data. Oak Ridge, Tenn.: Union Carbide Corporation, 1969.

5. Korfhage, R.R. and DeLutis, T.G. A Basis for Time and Cost Evaluation of Information Systems. Lafayette, Ind.: Purdue University, 1969.

6. Lancaster, F.W. Evaluation of the MEDLARS Demand Search Services. Bethesda, Md.: National Library of Medicine, 1968.

7. Landau, Herbert B. "The Cost Analysis of Document Surrogation: A Literature Review." American Documentation, October 1969.

8. Marron, Harvey and Synderman, Martin,Jr. "Cost Distribution and Analysis in Computer Storage and Retrieval. American Documentation, April 1966.

9. Penner, Rudolf J. "The Practice of Charging Users for Information Services: A State of the Art Report." Journal of the American Society for Information Science, January-February 1970.

10. Planning for An Experimental On-Line Bibliographic Access Service in 1970. Lister Hill National Center for Biomedical Communications, National Institutes of Health, Bethesda, Md.: March 1970.

11. Panel on Management of Information Activities, Committee on Scientific and Technical Information, Federal Council for Science and Technology. Report of the Sub-Panel on Unit Cost Analysis (draft). Washington, D.C.: October 1967.

12. Salton, Gerard. "A Comparison Between Manual and Automatic Indexing Methods." American Documentation, January 1969.

13. King, D. W. and Bryant, E. C.. Evaluation of Information Products and Services, Information Resources Press, Washington, D. C., due in October 1970.

14. King, D. W. and Bryant, E. C. "A Diagnostic Model for Evaluating Retrospective Search Systems". Information Storage and Retrieval June (1970).

15. King, D. W. and Brown, A. M. Some Comments on Marketing AIP Information Products and Services. Westat Research for American Institute of Physics, July 1970. Report #S-86.

Memoranda and Informal Reports

A. Bourne, Charles P. et al.. Abstracting and Indexing Rates and Costs: A Literature Review. Draft of report. May 1970.

B. Data Corporation. Rate Schedule. September 1969.

C. Marron, Harvey. "Cost Data for a Decentralized Information Network." Educational Resources Information Center, Washington, D. C. Internal memorandum.

D. Sessions, Mrs. V.S. Tables from "The Cost and Costing of Information Storage and Retrieval."