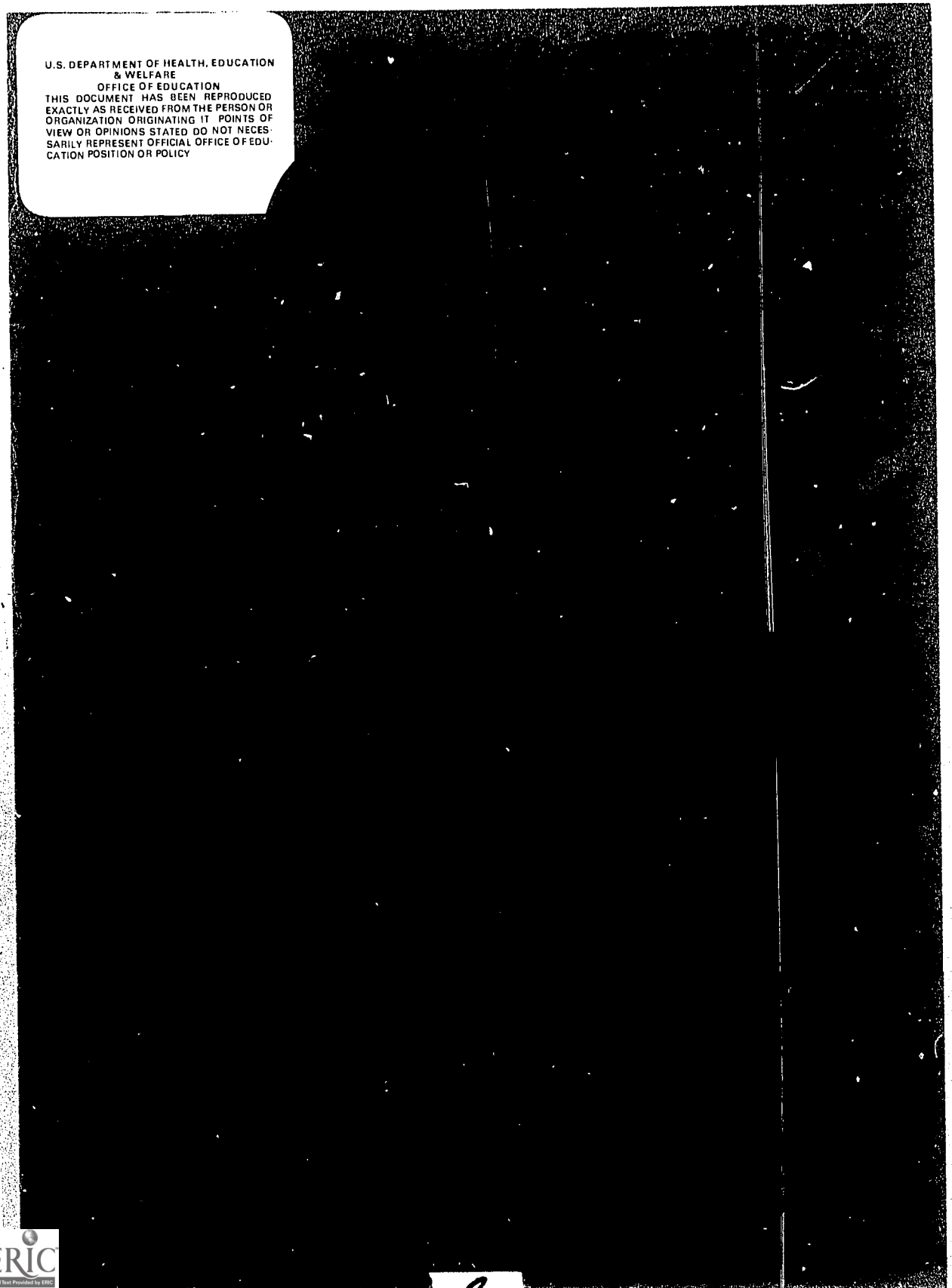
DOCUMENT RESUME

ED 051 837 LI 002 897

AUTHOR	King, Donald W.; Caldwell, Nancy W.
TITLE	Cost-Effectiveness of Retrospective Search Systems.
INSTITUTION	American Psychological Association, Washington, D.C.
REPORT NO	OC-3-71-TR-14
PUB DATE	71
NOTE	33p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	*Cost Effectiveness, *Information Retrieval, *Information Systems, Information Utilization, Models, Psychology
IDENTIFIERS	*National Information Systems

ABSTRACT

The purpose of the present study is to explore cost-effectiveness factors that affect the choice among alternative system designs for retrospective searching services. 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. (Author)



Preface

During the period from June 1968 to April 1970, the APA's Office of Communication Management and Development, working under the direction of the ad hoc Communications Committee, 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, Donald W. King, Executive Vice President of Westat Research, Inc., Rockville, Maryland, was selected to conduct a study of the cost-effectiveness of retrospective search systems. King is well known in the information science field for his work in evaluation and economics of information handling. This report, prepared by King and Nancy W. Caldwell under contract to APA, does not necessarily reflect the opinions or interpretations of the Association.

Harold P. Van Cott, Director
Office of Communication

TABLE OF CONTENTS

Highlights of Findings

The purpose of the present study was to explore cost-effectiveness factors that affect the choice among alternative system designs for retrospective searching services. 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.

As part of this study, a thorough literature search was conducted. To complement the sketchy information obtained from the search, a number of persons who manage operating information systems were contacted. Although much of their information was proprietary, many of them were able to provide information that is more realistic than that found in the literature, and estimates of "typical" costs are based heavily on these discussions. It should be emphasized that while these cost estimates are 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, "typical" costs of 36 system and subsystem alternatives are estimated.

Finally, if APA decides to charge for retrospective search services, it is necessary to have some knowledge of the 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 estimate reliably demand and a price/demand relationship, the model mentioned earlier, plus the examples given, yields an estimate of the cost/demand relationship. Some inferences are also drawn concerning frequency of use of the system.

COST-EFFECTIVENESS OF RETROSPECTIVE SEARCH SYSTEMS

One of the difficulties in 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. In the present study, a mathematical model that yields measures of cost-effectiveness for several combinations of system alternatives has been applied.[1]

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 earlier, 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.

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.

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 (see Lancaster, 1968), which used written request forms, indicated that a number of relevant documents were missed and nonrelevant documents retrieved due to an intermediary's not fully understanding the user's search requirements. The proportions of

---

[1] 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 King and Bryant (June, 1970; October, 1970).

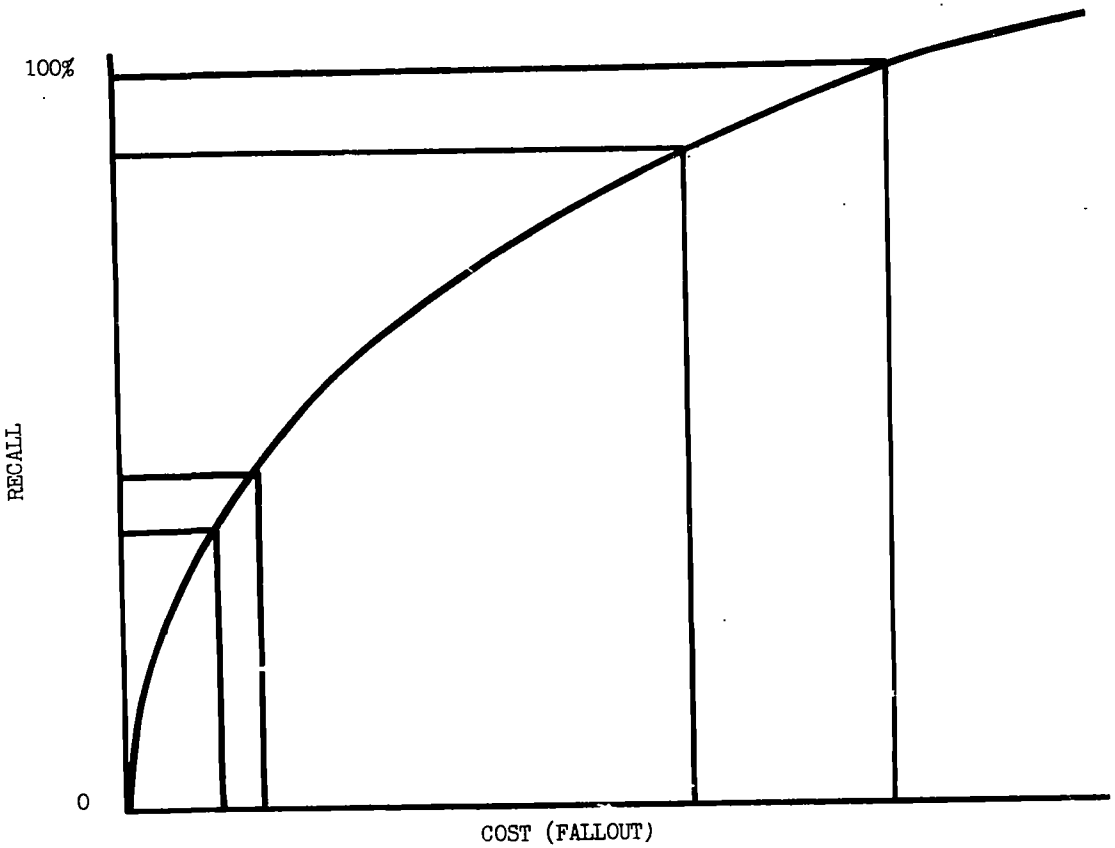


FIGURE 1. Cost versus recall for typical systems.

missed relevant and retrieved nonrelevant documents observed in this evaluation have been applied in the analysis given subsequently. As a second alternative, it was assumed that personal contact could be made with users, increasing costs but reducing failures by one-half. (One-half is conjectured, since no actual value for this number could be found 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, there are essentially three search systems in the 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, the present investigators established typical retrieval necessary to obtain recall levels of 40%, 60%, 80%, and 100% (see Salton, 1969). These values have been incorporated with the 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. The results of two experiments have been used to estimate recall and fallout. Results from the first experiment, involving tight screening, indicate 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 relevant documents missed.

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 systems, the analytical model was applied 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 4 years) and 50 total relevant documents in the file. The results of this analysis are given in **Table 1** for total retrieval per search and number of relevant documents retrieved per search. It is clear that the number of retrieved items that is 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 recall is increased from 80% to 100%. Thus, one must retrieve over 10 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.

TABLE 1

Total Retrieval and Number of Items Retrieved by
Alternative System Designs and Levels of Recall from Searching

Recall level	Written request			Telephone request		
	Batch processing	On-line index	On-line abstract	Batch processing	On-line index	On-line abstract
No screening						
.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
Tight screen on titles						
.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
Tight screen on abstracts						
.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

TABLE 1 (cont'd.)

Recall level	Written request			Telephone request		
	Batch processing	On-line index	On-line abstract	Batch processing	On-line index	On-line abstract
.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
Loose screen on titles						
.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
Loose screen on abstracts						
.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

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.

Cost Model

After many different retrospective search systems were carefully examined, a generalized cost model was developed, which will permit the 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 charges (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 used to display the full text of an abstract or document representatic screening search output. Fixed costs associated with input (C_3) ir: e 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), that is, 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

$$C'' = C_6 + X_5C_7.$$

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 (X4). 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, that is, the intermediary (C_9). For our purposes, two alternative methods of user/system communication are considered: written requests 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

$$C''' = C_8 + C_9 + X_3(C_{10} + C_{11} + C_{12}) + X_4C_{13}.$$

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, and
C = total annual cost.

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

Costs of Alternative Systems

Nearly all published literature on costs of information retrieval systems has emphasized the 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 the literature, "reasonable" cost estimates were derived that are used in the preceding cost model to estimate the general magnitude of total costs of various alternative retrospective search systems and subsystems.

Examples of typical values are given in Table 2 for cost items $C_1,\ldots,$ C_{13} for three general systems: (a) batch processing, index input, controlled vocabulary search; (b) on-line processing, index input, controlled vocabulary search; and (c) 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 the present 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 cost. 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.

TABLE 2A

Input and Search Costs for Three General Systems

Item	System		
	Batch index	On-line index	On-line abstract
Input			
C_3-Thesaurus[a]	$3,750	$3,750	
Staff[a]	500	500	$2,000
Update[b]	500	1,425	2,850
Tape conversion[b]	750	1,250	2,500
C_6-Index and abstract[c]	.30/item	.30/item	.80/item
Keyboarding[b]	.25/item	.25/item	.875/item
Other processing[b]	.025/item	.0875/item	
C_7-File loading	.000038/item/term	.000038/item/term	.00019/item/term
Search			
C_1-Computer rental	$65,000	$11,000	$15,000
Staff, other overhead	66,000	12,000	16,500
Terminal rent	--[d]	1,440	1,440
C_{10}-Computer processing	.05/item retrieved	--[e]	--[e]
C_{11}-Printout costs	.008/item retrieved	.008/item retrieved	.008/item retrieved

[a] One-half allocated to other processes and amortized over 4 years.
[b] Amortized over 4 years.
[c] Two-thirds allocated to other processes and amortized over 4 years.
[d] There is no terminal rent charge for remote batch processing.
[e] The search costs (C_{10}) for on-line index and on-line abstracts are not given in the table since they depend on additional variables. The costs used in this study came from prices quoted from firms that provide on-line search systems.

Other costs depend on the user/system interface (written requests versus telephone requests). Typical costs for these are given in Table 2B.

TABLE 2B
Search Costs Dependent on User/System Interface

Search	Written request	Telephone request
C_4 - staff, rent, other overhead	$500[a]	$500[a]
C_9 - staff, equipment	$11.25/search.	$15.00/search

[a]Amortized over 4 years.

Screening can be performed on titles or abstracts, which yield substantially different costs. Typical costs for these are given in Table 2C.

TABLE 2C
Screening Costs

Screen	Titles	Abstracts
C_2 - staff, rent, equipment, other overhead	$4,000[a]	$32,000[a]
C_{12} - staff	$.04/item retrieved	$.125/item retrieved

[a]Amortized over 4 years.

Finally, mailing costs also depend on whether titles or abstracts are sent to the user. Typical mailing costs are given in Table 2D.

TABLE 2D
Mailing Costs

Mailing	Titles	Abstracts
C_5 - staff, rent, equipment, other overhead	$250[a]	$250[a]
C_8 - preparation, staff materials	$.20/search	$.35/search
C_{13} - postage, materials	$.002/item sent	$.10/item sent

[a]Amortized over 4 years.

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 4-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% prior to screening -- see Table 1);
5. number of items mailed per search (X_4) is 22 (i.e., the recall level at 80% and after screening -- see Table 1); and
6. number of terms in authority list (X_5) is 1,000.

The model is:

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

$$= \$131,000 + \$32,000 + \$5,500 + \$500 + \$250 + 100,000[\$.575 + (1,000)(\$.0000375)] + 4,000[0 + \$15.00 + 126(\$.05 + \$.008 + \$.125) + (22)(\$.10)]$$

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

$$= \$391,532$$

The average cost per search is then \$97.90 (\$391,532/4,000); 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 preceding example 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.

Discussion

Table 3 gives the cost per search, cost per item retrieved, and cost per relevant item retrived for 1,000 requests for six alternative systems. Table 4 gives the same information for 4,000 requests. It appears that on-line index is consistently the best search system with regard to all three cost measures.[2] 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.

[2]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 3

Cost per Search, per Item Retrieved, and per Relevant Item Retrieved for Alternative Retrospective Search Systems, Subsystems, and Recall Levels (Demand = 1,000)

Recall level and cost ($) for each	Written request			Telephone request		
	Batch processing	On-line index	On-line abstract	Batch processing	On-line index	On-line abstract
No screening -- mail titles to users						
.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 item retrieved	11.20	6.20	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 item 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 item 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 item retrieved	4.80	3.20	5.80	4.80	3.20	5.80
No screening -- mail abstracts to users						
.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 item 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 item retrieved	7.60	4.20	8.80	7.70	4.30	9.00

TABLE 3 (cont'd.)

Recall level and cost ($) for each	Written request			Telephone request		
	Batch processing	On-line index	On-line abstract	Batch processing	On-line index	On-line abstract
.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 item 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 item retrieved	5.60	3.60	6.20	4.80	3.60	6.10
Tight screen on titles -- mail titles to users						
.40						
Search	218.00	123.00	259.00	221.00	127.00	262.00
Item retrieved	72.67	41.00	86.33	73.67	42.33	87.33
Relevant item retrieved	72.67	41.00	86.33	73.67	42.33	87.33
.60						
Search	220.00	126.00	261.00	224.00	130.00	264.00
Item retrieved	44.00	25.20	52.20	44.80	26.00	52.80
Relevant item retrieved	44.00	25.20	52.20	44.80	26.00	52.80
.80						
Search	226.00	149.00	265.00	230.00	153.00	269.00
Item retrieved	32.29	21.29	37.86	32.86	21.86	38.43
Relevant item retrieved	37.67	24.83	44.17	32.86	21.86	38.43
1.00						
Search	251.00	179.00	314.00	255.00	183.00	318.00
Item retrieved	25.10	19.89	34.89	28.33	20.33	35.33
Relevant item retrieved	31.38	22.38	39.25	31.88	22.88	39.75

TABLE 3 (cont'd.)

Recall level and cost ($) for each	Written request			Telephone request		
	Batch processing	On-line index	On-line abstract	Batch processing	On-line index	On-line abstract
Tight screen on abstracts -- mail abstracts to users						
.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 item 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 item 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 item 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 item retrieved	12.00	8.70	13.90	12.10	8.90	14.00
Loose screen on titles -- mail titles to users						
.40						
Search	217.00	123.00	259.00	221.00	127.00	262.00
Item retrieved	8.68	5.86	12.95	9.21	6.05	12.48
Relevant item retrieved	12.76	7.24	15.24	13.00	7.47	15.41
.60						
Search	220.00	126.00	261.00	224.00	130.00	264.00
Item retrieved	4.89	3.16	8.16	4.98	3.42	8.00
Relevant item retrieved	8.80	5.04	10.44	8.62	5.00	10.15

TABLE 3 (cont'd.)

Recall level and cost ($) for each	Written request			Telephone request		
	Batch processing	On-line index	On-line abstract	Batch processing	On-line index	On-line abstract
.80						
Search	226.00	149.00	265.00	230.00	153.00	269.00
Item retrieved	2.77	2.19	4.73	2.77	2.22	4.71
Relevant item retrieved	6.85	4.52	8.03	6.76	4.50	7.91
1.00						
Search	251.00	180.00	314.00	255.00	183.00	318.00
Item retrieved	1.12	1.34	2.78	1.13	1.36	2.78
Relevant item retrieved	5.98	4.29	7.47	5.93	4.26	7.40
Loose screen on abstracts -- mail abstracts to users						
.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 item 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 item 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 item 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 4

Cost per Search, per Item Retrieved, and per Relevant Item Retrieved for Alternative Retrospective Search Systems, Subsystems, and Recall Levels (Demand=4,000)

Recall level and cost ($) for each	Written request			Telephone request		
	Batch processing	On-line index	On-line abstract	Batch processing	On-line index	On-line abstract
No screening -- mail titles to users						
.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 item 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 item 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 item 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 item retrieved	1.70	1.60	2.50	1.70	1.70	2.50
No screening -- mail abstracts to users						
.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 item 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 item retrieved	2.50	1.80	2.90	2.60	1.90	3.00

TABLE 4 (cont'd.)

Recall level and cost ($) for each	Written request			Telephone request		
	Batch processing	On-line index	On-line abstract	Batch processing	On-line index	On-line abstract
.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 item 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 item retrieved	2.40	2.10	2.80	2.50	2.10	2.90
Tight screen on titles -- mail titles to users						
.40						
Search	65.00	46.00	82.00	69.00	50.00	86.00
Item retrieved	21.67	15.33	27.33	23.00	16.67	28.67
Relevant item retrieved	21.67	15.33	27.33	23.00	16.67	28.67
.60						
Search	68.00	53.00	84.00	72.00	57.00	88.00
Item retrieved	13.60	10.60	16.80	14.40	11.40	17.60
Relevant item retrieved	13.60	10.60	16.80	14.40	11.40	17.60
.80						
Search	74.00	67.00	101.00	78.00	71.00	105.00
Item retrieved	10.57	9.57	14.43	11.14	10.14	15.00
Relevant item retrieved	12.33	11.17	16.83	11.14	10.14	15.00
1.00						
Search	99.00	102.00	151.00	103.00	106.00	155.00
Item retrieved	9.90	11.33	16.78	11.44	11.78	17.22
Relevant item retrieved	12.38	12.75	18.88	12.88	13.25	19.38

TABLE 4 (cont'd.)

Recall level and cost ($) for each	Written request			Telephone request		
	Batch processing	On-line index	On-line abstract	Batch processing	On-line index	On-line abstract
Tight screen on abstracts -- mail abstracts to users						
.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 item 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 item 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 item 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 item retrieved	5.00	5.00	6.20	5.50	5.10	6.40
Loose screen on titles -- mail titles to users						
.40						
Search	65.00	46.00	82.00	69.00	50.00	86.00
Item retrieved	2.60	2.19	4.10	2.88	2.38	4.09
Relevant item retrieved	3.82	2.71	4.82	4.06	2.94	5.06
.60						
Search	68.00	53.00	84.00	72.00	57.00	88.00
Item retrieved	1.51	1.39	2.63	1.60	1.50	2.67
Relevant item retrieved	2.72	2.12	3.36	2.77	2.19	3.38

TABLE 4 (cont'd.)

Recall level and cost ($) for each	Written request			Telephone request		
	Batch processing	On-line index	On-line abstract	Batch processing	On-line index	On-line abstract
.80						
Search	75.00	67.00	101.00	78.00	71.00	105.00
Item retrieved	.91	.99	1.80	.94	1.03	1.84
Relevant item retrieved	2.27	2.03	3.06	2.30	2.09	3.09
1.00						
Search	99.00	101.00	151.00	103.00	106.00	155.00
Item retrieved	.44	.75	1.34	.46	.79	1.36
Relevant item retrieved	2.36	2.40	3.60	2.40	2.47	3.60
Loose screen on abstracts -- mail abstracts to users						
.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 item 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 item 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 item 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 item retrieved	3.80	3.40	4.10	3.80	3.40	4.10

When comparing written request against telephone request, we find that in some instances, written requests may cost less for the search but more for each relevant item retrieved (e.g., see the first section of **Table** 3 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 that of 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 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? Second, if the former 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 bette 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, similar systems were investigated. Two systems that do not charge for use of the retrospective search system (the Oefense Documentation Center, DDC; and MEDLARS) report demands of 24,000 and 12,000 (respectively). National Technical Information Service (formerly the Clearinghouse for Scientific and Technical Information), with a demand of 8,000, charges a nominal amount for its search service. The Smithsonian's **Science** Information Exchange 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 5**, cost is relatively insensitive to demand when demand is greater than 4,000. **Table** 5 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 that 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.

TABLE 5

Costs per Search with Two Different Allocations of the Indexing Costs 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'''		
1/3 of the indexing costs and 1/2 of the thesaurus development costs						
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
All of the indexing and thesaurus development costs						
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

As described in the preceding example, some of the input costs were allocated to various other services. For the costs presented in Table 3 and 4, one-third of the indexing costs and half of the thesaurus development costs were allocated to the retrospective search system. To investigate the sensitivity of the costs to different allocations of these fixed cost elements, the cost per search for this system[3] at various levels of demand was calculated for two different allocations of the fixed costs. The first allocation is that described earlier and used in Tables 3 and 4 for comparisons with other system alternatives. In the second, all of the indexing and thesaurus development costs were allocated to the retrospective search system. As shown in Table 5, the costs per search are relatively insensitive to this change in allocation.

It is clear that APA's 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. The 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 excerpted from King and Brown, 1970).

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 of his own time. For example, if a retrospective search results in 5,000 identified titles, he is not likely to be satisfied since he must pay such a high price of his own time to screen out those documents that 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 (1968). 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**. When one plots the marginal cost against

---

[3]Batch processing, telephone request, tight screening on abstracts, and abstracts mailed to users.

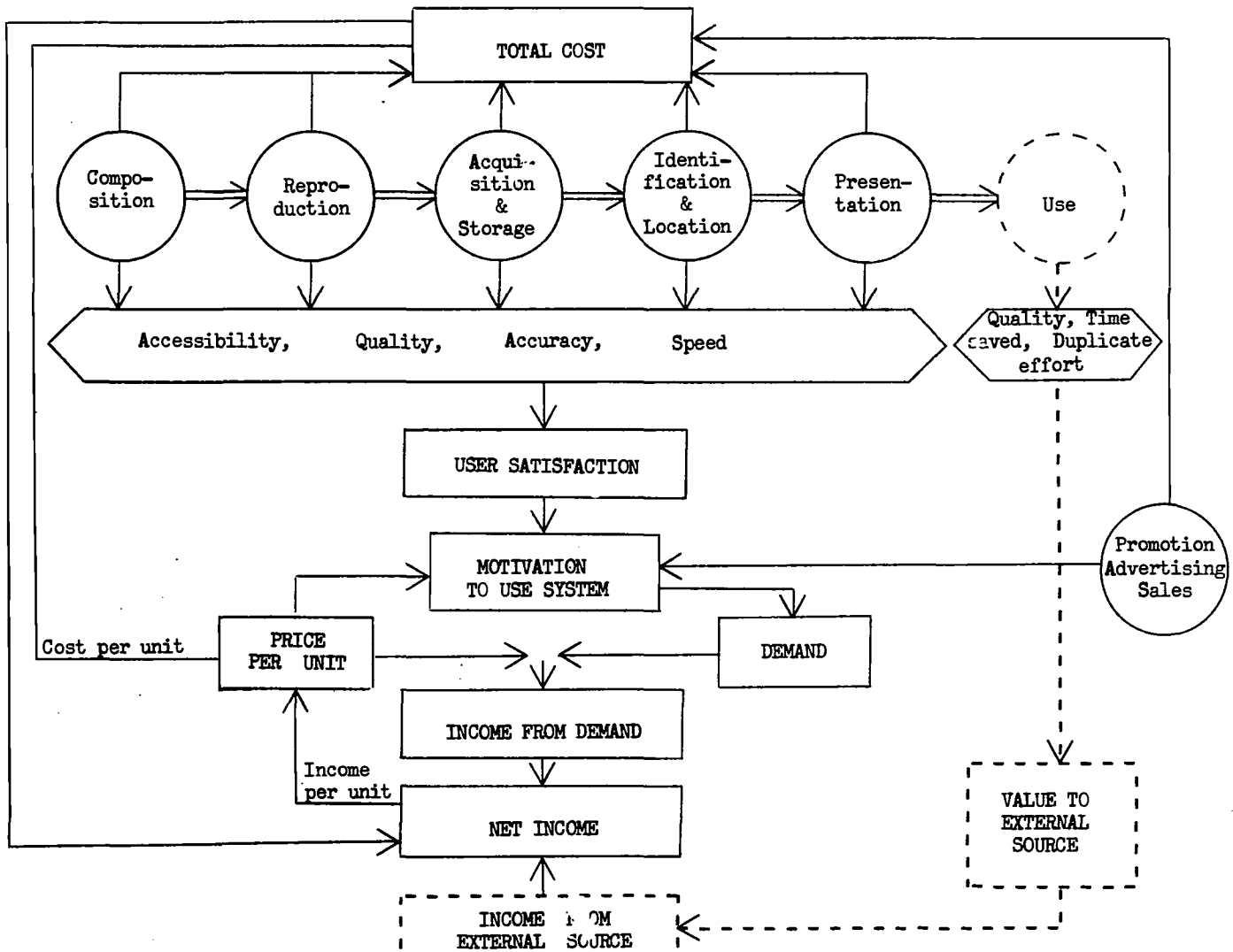


FIGURE 2. Relationships among functions and measures when users are not a part of the system being evaluated.

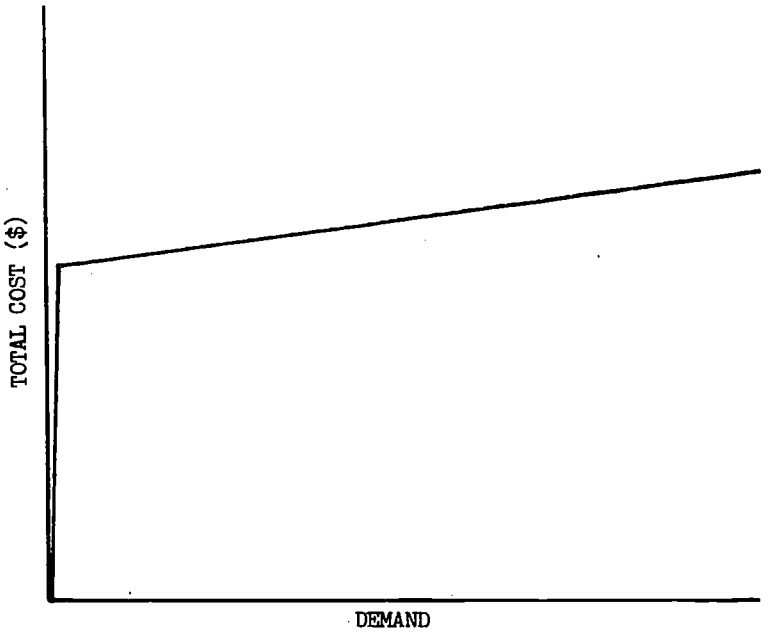


FIGURE 3. Typical cost curve for information products and services.

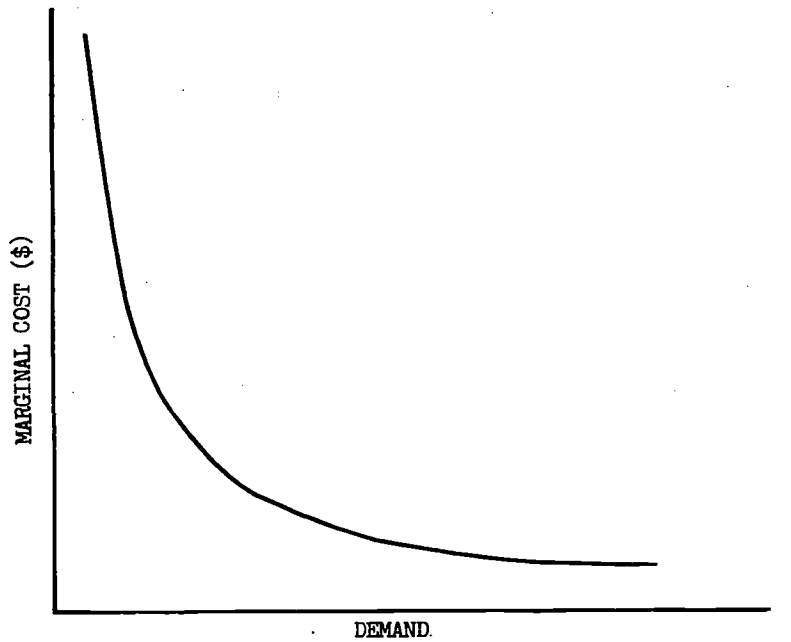


FIGURE 4. Typical marginal cost curve for information products and services.

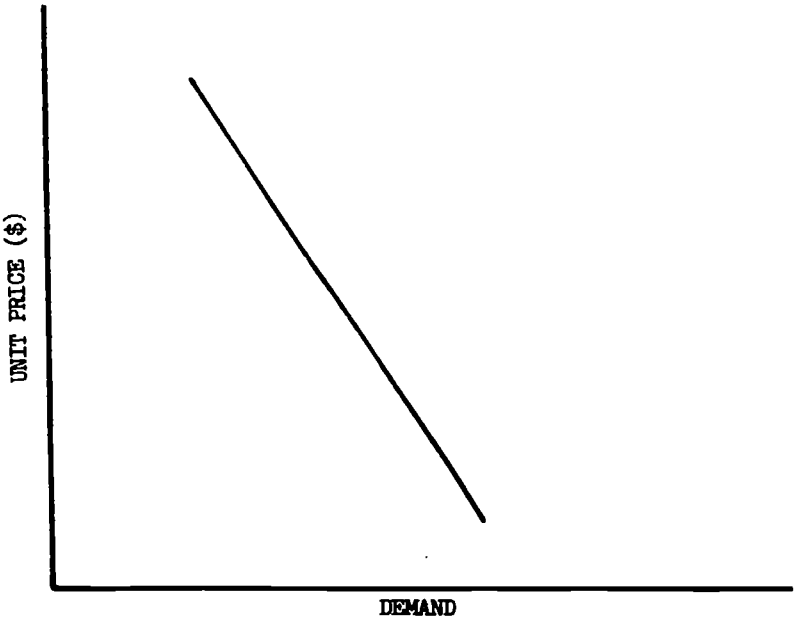


FIGURE 5. Typical inelastic price/demand curve.

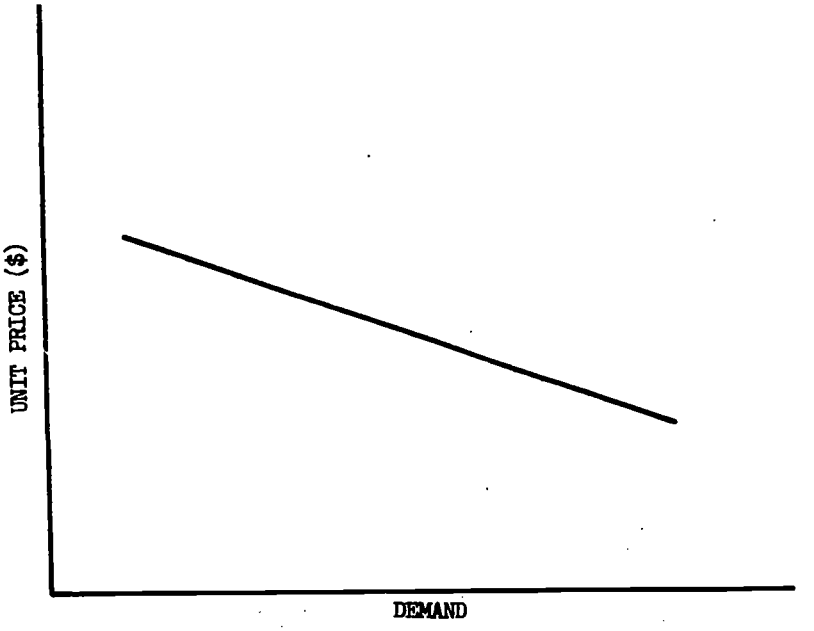


FIGURE 6. Typical elastic price/demand curve.

quantity or usage, the curve drops as shown in Figure 4. It is important to establish marginal cost over a likely range of demand for each of the products and services.

The 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.

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. 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

Bourne, C. P., North, J. B., & Kasson, M. S. Abstracting and indexing rates and costs: A literature review (draft). Palo Alto, Calif.: Information General Corporation, 1970.

Bourne, C. P., & Ford, D. F. Cost analysis and simulation procedures for evaluation of large information systems. American Documentation, 1964 15(2), 142-149.

Data Corporation. Rate schedule. Dayton, Ohio: Author, September 1969.

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

Freeman, M. E. Determining costs of information systems. Journal of Chemical Documentation, 1967, 7(2), 101-106.

Helmkamp, J. G. Managerial cost accounting for a technical information center. Bloomington, Ind.: Indiana University Press, 1968.

Kidd, E. M., Price, C. E., & Young, S. L. Study of the Data Central System for Information Retrieval applied to NSA data. Oak Ridge, Tenn.: Union Carbide Corporation, 1969.

King, D. W., & Brown, A. M. Some comments on marketing AIP information products and services (Tech. Rep. No. S-86) Washington, D.C.: American Institute of Physics, July 1970.

King, D. W., & Bryant, E. C. A diagnostic model for evaluating retrospective search systems. Information Storage and Retrieval, 1970, 6(3), 261-272.

King, D. W., & Bryant, E. C. Evaluation of information products and services. Washington, D.C.: Information Resources Press, May 1971, in press.

Korfhage, R. R., & DeLutis, T. G. A basis for time and cost evaluation of information systems. Lafayette, Ind.: Purdue University Press, 1969.

Lancaster, F. W. Evaluation of the MEDLARS demand search services. Bethesda, Md.: National Library of Medicine, 1968.

Landau, H. B. The cost analysis of document surrogation: A literature review. American Documentation, 1969, 20(4), 302-309.

Marron, H. Cost data for a decentralized information network (internal memo). Washington, D.C.: Educational Resources Information Center, undated.

Marron, H., & Synderman, M., Jr. Cost distribution and analysis in computer storage and retrieval. American Documentation, 1966, 16(2), 89-95.

National Institutes of Health, Lister Hill National Center for Biomedical Communications. Planning for an experimental on-line bibliographic access service in 1970. Bethesda, Md.: Author, 1970.

BIBLIOGRAPHY (cont'd.)

Penner, R. J. The practice of charging users for information services. Journal of the American Society for Information Science, 1970, 21(1), 67-74.

Salton, G. A comparison between manual and automatic indexing methods. American Documentation, 1969, 20(1), 61-71.