

## SECTION IV

### COST-PERFORMANCE ANALYSIS

Jeffrey Katzer\*

This section of the report is an extension of the performance study conducted with last year's version of SUPARS (Section V, 1971 Final Report). Our major goal was to obtain estimates of system performance -- specifically recall, paired with some measure of the cost required to achieve particular levels of performance. A cost-performance relationship for various sub-components of the system can be determined and compared, thereby providing valuable input to design considerations for future versions of SUPARS. Last year, for example, we found that searches using simple Boolean operators performed better and at less cost than searches using syntactical or positional operators. This finding affected our plans for user education, budget estimates and the design of this study.

#### 1. DEFINITIONS

SUPARS/DPS - The current version of soft-ware support for SUPARS is a highly modified and extended version of IBM's Document Processing System (DPS). (6) The input-output interactive user language was developed by the SUPARS programming staff for interface with the batch-mode DPS search modules. Since system performance is a function of input and output, it is misleading to separate the SUPARS components of the system from those of DPS. Thus, SUPARS/DPS correctly designates the total operating retrieval system.

Document - As used here, a document is the total representation of one item in Psychological Abstracts. Usually a document consists of a bibliographic citation and an abstract. Whenever Psychological Abstracts does not publish an abstract for an item, the SUPARS files represent the document as a citation only.

Document Data Base (DDB) - The DDB is that data base searched by SUPARS/DPS consisting of Psychological Abstracts documents (as defined above.) That portion of the DDB used in this study consisted of most documents published between February, 1970 and June, 1971.<sup>1</sup> (The size and growth of the total

---

\* The author wishes to acknowledge the contribution of several staff members to this report -- especially Miss Sandra Browning and Dr. Kenneth Cook for completing the onerous duties of search experts, and Miss Margaret Mucia for the lengthy task of data tabulation.

<sup>1</sup>Documents published in March and May, 1970, were discovered to be missing from the original machine-readable tapes. While we ultimately obtained these documents, they were received too late to be included in this study.

DDB available to users for searches is described in Section II of this report.) As used hereafter, DDB refers to that subset of the total DDB used in this study.

Vocabulary Data Base (VDB) - The VDB is a pseudo-data base searched by a simplified version of the SUPARS/DPS user language. The components of the VDB are the words in the system's vocabulary -- i.e. all non-common words found in the DDB derived from the free-text processing documents by DPS. The VDB is called a pseudo-data base because we did not have to create a new data base as defined by DPS. The components of VDB are stored and used during construction and searching of the DDB. Our problem was to access this vocabulary as a separately retrievable data base -- a complex, but not too difficult programming task.

Search Data Base (SDB) - The SDB is that data base searched by SUPARS/DPS consisting of previously made SUPARS search inquiries. Throughout last year's and this year's operating periods we attempted to store all interactions with the SUPARS system. Because of complex systems problems only a subset of the searches made in 1970 could be identified and stored. These, plus the great majority of the 1971 searches to the DDB make up the SDB. At the time this study was begun, the SDB consisted of 4,235 searches.

Information Requirement Statement (IRS) - In this study an IRS is a written statement of a person's information needs. We assume that the written statement actually represents that need.

Operators - The user language accepts as input one or more keywords (search words) which represent the IRS of the searcher. The keywords may be combined with Boolean operators (and, or, not); or grammatical operators (e.g. which specifies the desired proximity of keywords, or the truncation operator -- see below.)

Truncation Operator - The truncation operator is part of the user language. It is requested by typing the characters (?) at the end of a whole or partial word. Internally the truncation operator retrieves the union of documents containing any word in the vocabulary beginning with the letters to the left of the (?).

Search Inquiry - A SUPARS/DPS search inquiry is one complete input into the system. It consists of some system commands, an output specification statement, plus one or more user chosen keywords combined with the operators desired in order to represent concepts described by IRS. All correctly input search inquiries will not necessarily retrieve output. For one information requirement, a user may typically enter several search queries.

Logically Different Search Inquiry - One of the arbitrary parameters of the SUPARS system is the maximum number of retrieved documents which can be typed out in response to any one search inquiry. Using the fuller output format (LIST RECORD) each search will print a maximum of 10 documents. The citation output format (LIST BRIEF) will print up to 100 document identifications per search. If a searcher wishes to print more than this maximum,



the search inquiry must be repeated, (although he need not type the key words and operators again.) Each of these repeated search inquiries will be counted as one logical search inquiry. In order to have two logically different inquiries, the user would have to change one or more keywords or operators.

Estimated Recall - Recall, as a measure of performance, is defined as the retrieved proportion of relevant documents in the data base. For large data bases it is difficult to know how many documents in the total data base are relevant (which is the denominator of the recall ratio). To estimate a recall ratio, we have assumed that each of the ten documents identified as relevant by the writer of an IRS represents 10% of the total number of relevant documents in the entire data base.

Total Retrieval - This is simply the number of documents retrieved by one search inquiry. It can be used as the denominator of the precision ratio. By using total retrieval instead of precision, we were able to specify how many documents a user will have to scan through in order to find the relevant ones. In this sense, total retrieval is a measure of effort or cost of labor of the system. In this study, total retrieval is one of our major measures of system cost. It should be noted, however, that the effect of this cost, (effort) on users of a particular system, probably varies. It is not considered here.

Costs - Many factors contribute to the cost of a retrieval system. King (4) has identified the major ones. Many of these factors are installation dependent (e.g. salaries, whether or not an intermediary is used, whether retrievals are mailed to requester, etc.) For this reason we will limit our analysis to two measures of cost: (1) total retrieval -- which is reasonably independent of the particular installation, (2) computer costs in terms of CPU time and the number of I/O channel executions. Though these are dependent upon particular hardware configurations, comparisons among alternative computing systems can be made. CPU time and I/O executions are easily obtained and we chose them because they constitute a common standard for comparing different computer systems.

## 2. METHOD

The major objective of the study outlined below is to obtain an estimate of the cost-performance relationship for the current versions of the SUPARS/DPS.

At the end of our analysis of the 1970 SUPARS system, we concluded that performance could be improved by some form of vocabulary control. The VDB and the SDB were constructed as an initial attempt to "control" the vocabulary, but retain the free-text component of the system.

The VDB allows a searcher to determine if his keywords are in the vocabulary. Another option available with the VDB retrieves all vocabulary items with the same initial letters. We had hoped that this option could be

used to reduce the total retrieval of a search inquiry. If, for example, a searcher plans to use the truncation operator, prior use of the VDB might identify unrelated words which the truncated keyword would encompass. By eliminating the non-related words from his inquiry, the total retrieval will be reduced. This method of using the VDB could also improve the proportion of relevant documents retrieved by identifying synonym-like<sup>2</sup> keywords in the vocabulary.

The SDB was designed primarily to aid the searcher to obtain synonyms.<sup>2</sup> By typing as input into the SDB a keyword which is not general enough (i.e. does not retrieve enough documents), the searcher can retrieve all keywords and all stored searches which contain his keyword. On option, the searcher can have the entire retrieved inquiry printed -- including the operators which link the keywords. Another possible input into the SDB is the identification of a known relevant document; the output would consist of all searches which retrieved that document. All of these options were designed to make the vocabulary and logic of any searcher anonymously available to any other searcher -- hopefully increasing a users ability to search the DDB more effectively.

In order to test the effect of the DDB and the VDB on performance we conducted a laboratory experiment very much similar to the one conducted last year. In the experiment searching stimuli were obtained by hiring subjects to write a statement of their information needs (IRS). Upon completing his IRS each subject manually searched the published version of Psychological Abstracts to find documents relevant to his information needs. The ten most relevant of these documents were used in the study.

Expert searchers read each IRS and used SUPARS/DPS to retrieve 9 of the 10 documents previously identified as relevant. Each retrieved relevant document was taken to be an estimate of another 10% of all relevant documents in the DDB. Experts were restricted in their search inquiries, however, to one or more of the three data bases. (DDB, VDB, SDB.) Through proper experimental design we had hoped to obtain clear measures of the relevant effectiveness of the two new data bases.

#### a. Experimental Design

Our original plans called for four treatment conditions (controlling the data base permitted to be used by a search expert), four search experts and sixteen IRSs. These variables were arranged in a 4 x 4 Latin Square design replicated 4 times.

---

<sup>2</sup>The term "synonym" is used here as a general name for all words which may refer to the same class of objects. True synonyms fit in this class as do those items listed under a "See also" heading in an index. For example, synonyms of "human" which may help a user, are male, female, boy, girl, child, adolescent, student, subject, teenager, etc. -- plus all of their plurals.

For various reasons,<sup>3</sup> we had to delay the start of the experiment and consequently one of the experts had to resign from the study because of prior commitments. The study was redesigned with three treatment conditions, three experts and 15 IRSs arranged in a 3 x 3 Latin Square replicated 5 times. This design is presented below in Table XLI.

TABLE XLI  
EXPERIMENTAL DESIGN

|                 | IRS |   |    |    |    |    |   |   |   |   |    |    |    |   |   |
|-----------------|-----|---|----|----|----|----|---|---|---|---|----|----|----|---|---|
|                 | 8   | 7 | 10 | 15 | 12 | 16 | 9 | 2 | 4 | 3 | 13 | 14 | 11 | 5 | 1 |
| Search Expert A | S   | D | V  | S  | D  | V  | S | V | D | S | V  | D  | V  | S | D |
| B               | D   | V | S  | D  | V  | S  | V | D | S | V | D  | S  | D  | V | S |
| C               | V   | S | D  | V  | S  | D  | D | S | V | D | S  | V  | S  | D | V |

In Table XLI three treatment conditions identify the data bases each expert was permitted to use for a particular IRS: 'D' means the expert was restricted to the DDB, 'V' means the expert could use the VDB as well as the DDB, 'S' means that the SDB and the DDB could be used.<sup>4</sup>

#### b. IRSs

Subjects were needed to produce the required IRSs for the experts to search. They were contacted through advertisements asking for people with information needs in the social or behavioral sciences. A brief description of the fifteen subjects whose IRSs were used and of the IRSs themselves is presented in Table XLII.

Each subject was asked to write a detailed description of one of his current information needs. He was not told to construct his IRS in terms of the content of Psychological Abstracts, in fact, the data base was not mentioned at this stage of the experiment. The instructions given to the subjects are presented in the Appendix. In general, each subject was asked to state his information needs in such a way that we could go out and get the information for him.

<sup>3</sup>Notably, the delays caused by the computing center's conversion from the IBM 360/50 to an IBM 370/155 before this experiment was planned to be carried out.

<sup>4</sup>In the original 4 x 4 design, the fourth treatment would have allowed a search expert to search all 3 data bases. While it is unfortunate that this condition is missing from the study we can still estimate the effects of the 4 condition -- if it turns out to be additive.

TABLE XLII  
CHARACTERISTICS OF SUBJECTS AND IRSs

| IRS | Sex | Academic Status | Dept    | Topic  | Number of Relevant Documents Found |
|-----|-----|-----------------|---------|--|------------------------------------|
| 1   | M   | Fac             | Educ    | Vocational choice process of the disadvantaged                 | 18                                 |
| 2   | F   | Grad            | Educ    | Behavioral Theory, counseling, and education                   | 43                                 |
| 3   | M   | Fac             | Psy     | Activity level of animals in response to stimulation           | 73                                 |
| 4   | M   | Grad            | Law     | Change in suburbs and voting patterns                          | 16                                 |
| 5   | M   | Grad            | Psy     | Efficiency and discriminability of tests                       | 17                                 |
| 7   | F   | Undgr           | Nursing | Incompatability in marriage                                    | 33                                 |
| 8   | M   | Grad            | Psy     | Comprehensive services for unmarried pregnant teenagers        | 43                                 |
| 9   | M   | Grad            | Psy     | Development of human relationships in T group                  | 42                                 |
| 10  | M   | Grad            | Educ    | Education of Exoeptional Children and Instructional Technology | 18                                 |
| 11  | M   | Grad            | Educ    | Instructional t.v. in education                                | 65                                 |
| 12  | M   | Grad            | Psy     | Effects of intercultural contact and bilingualism              | 13                                 |
| 13  | F   | Grad            | Psy     | Errorless learning in adult problem solving                    | 19                                 |
| 14  | M   | Fac             | Lib Sci | Attitude of people confronted with technology                  | 17                                 |
| 15  | M   | Undgr           | Psy     | Police intervention in social/sexual deviancy                  | 54                                 |
| 16  | M   | Fac             | Psy     | Review of federally supported youth work programs              | 42                                 |

### c. Identification of Relevant Documents

Upon completing the writing of his IRS, each subject was given the February, 1970 -- June, 1971 issues of Psychological Abstracts. These issues were chosen for two reasons: (a) they constituted a data base of approximately the same size as that used in last year's study, and (b) they were the most recently received issues of Psychological Abstracts (the IRSs were collected in early Fall of 1971) which would also be included in the DDB.

Each subject was asked to find the entries in Psychological Abstracts which he judged to be relevant to his IRS. A subject could complete this task in any manner he chose. We suggested that he at least look at each month of the journal -- but there were no directions specifying how to find the relevant documents or how many to choose.<sup>5</sup>

The final task for each subject was to rate each of the documents chosen into terms of their relevance to his IRS. The 10 documents rated most relevant were used in this study.<sup>6</sup>

In last year's study we used the 5 documents rated most relevant and the 5 rated least relevant by the subjects. Upon analysis, we found a better cost performance relationship for the most relevant documents. Because of the plausability of this finding we did not think it necessary to replicate that portion of the study this year.

### d. Search Experts

Three SUPARS staff members served as search experts. All had considerable experience with SUPARS: they worked on last year's project, helped in the design of this year's system, made numerous searches with the system and prepared materials which taught others how to use the system.

Each expert was given the 15 IRSs and the identification (Volume and Abstract No.) of the 10 documents judged relevant by the subjects. The experts were instructed to use SUPARS as effectively as possible to retrieve any 9 of the 10 documents.

---

<sup>5</sup>None of the fifteen subjects completed this test without identifying at least 10 documents he considered relevant. If this had not happened, he would have been asked to continue the task.

<sup>6</sup>It was sometime after the subjects completed their job, when we discovered the two gaps in the DDB -- see footnote #1. At that time it was necessary to check each of the ten documents paired with each IRS. If the chosen document fell in one of the gaps of the DDB, the next most highly rated unused document not in the gap was substituted.

Their job was to use the system to maximize the retrieval of relevant documents with minimal total retrieval. To achieve this goal they were permitted to use any of the SUPARS/DPS operators available. The only restriction the experts were to observe was the treatment condition -- the data base combinations permitted for each IRS. (Experts were permitted, but not required, to use the SDB or the VDB when they were searching under the 'S' or 'V' conditions (see Table XLI.)\*)

#### e. Procedure

Each expert was given a packet of 15 IRS -- each paired with a list of 10 documents judged to be relevant to that IRS. The order of IRSs in each packet was randomized. Unfortunately, however, it became essential for the experts to begin their work before the 1971 searches were loaded into the SDB. The loading of searches took much longer than expected; consequently the three experts searched their 5 IRSs under the 'S' treatment condition last. That is, the effect of the 'S' treatment condition is confounded with a practice effect. This will be discussed, at more length, later.

All of the experts' terminal interactions with SUPARS were saved. For each IRS, the first 9 relevant documents retrieved were marked, and the cumulative number of non-relevant documents retrieved before each relevant document was counted. In this way, 9 measures were obtained for each of the 45 expert-IRS combinations. The mean number of non-relevant documents retrieved before the first relevant document was retrieved (across all forty-five expert-IRS combinations) provides the estimate of total fall-out, (or cost of screening) at the 10% recall level.

We also counted the number of search inquiries made because the computer cost of running SUPARS depends mainly upon the number of inquiries, not the number of documents retrieved. Our second measure of cost (CPU time and I/O executions) is obtained from data collected by STATPAC. For each search inquiry an expert enters for an IRS, STATPAC stores the CPU time and I/O executions (among other things). Summing these through each relevant document retrieved produces the second measure of cost.

#### f. Analysis

The Latin Square design was analyzed as a cross-over design. The partitioning of sources of variation is shown in Table XLIII. An analysis of variance procedure, followed by hypothesis testing or interval estimation will determine whether or not there is a significant difference among the 3 treatment conditions. There probably will be real differences among experts and IRSs.

---

\*Experts were given the option of using SDB under the 'S' condition or the VDB under the 'V' condition.



TABLE XLIII

## SOURCES OF VARIATION AND DEGREES OF FREEDOM IN CROSS-OVER DESIGN

| Sources of Variation  | df |
|-----------------------|----|
| Among Treatment       | 2  |
| Among Experts         | 2  |
| Among IRSs            | 14 |
| Residual <sup>1</sup> | 26 |
| Total                 | 44 |

1

The degrees of freedom for residual is not as large as desired. Our original design called for four 4 x 4 Latin Squares. The degrees of freedom in that design would have been more satisfactory.

### 3. RESULTS

The findings of this study are presented in 4 sections: First, a brief look at the treatment variable to see if search experts actually used the two optional data bases (VDB, SDB) as searching aids. Second, another look at the dependent variable, computer cost. Third, we will compare the cost-performance relationship of the different treatments. Finally, an attempt will be made to compare last year's and this year's systems in terms of cost-performance.

#### a. Use of Optional Data Bases

As noted earlier each search expert was assigned five IRSs he could search with the VDB as an aid. Five other IRSs could be searched with the SDB, and the remaining 5 IRSs could not be searched with any aid -- only the DDB was permitted to be used.

Since any differences between the use of different data bases is probably cumulative, we only need to examine the total number of search inquiries made to achieve the 90% recall level. The search frequencies and average use of the three data bases is presented in Table XLIV.

The DDB was necessarily used in all forty-five IRSs because it contained the documents to be retrieved. Of the 824 logically different inquiries

TABLE XLIV

USE OF THREE DATA BASES TO ACHIEVE THE 90% RECALL LEVEL

|                  | 1  | 2   | 3  | 4                              | 5   | 6   |
|------------------|--|---|--|--------------------------------|---|---|
| Data Base        | Total No. of IRSs Searched (By all Ex-perts) | No. of IRSs searched which did not include Any Use of the Data Base | No. of logically different search inquiries made | Standard for Computing Percent | Percent of Standard Number of Search Inquiries Made With Permitted Data Base (=Column 3÷Column 4) | Mean of Search Inquiries per IRS (=Column 3 ÷ Column 1) |
| DDB              | 45   | 0   | 691  | 824 <sup>1</sup>               | 84%   | 15.3  |
| VDB              | 15   | 2   | 88   | 318 <sup>2</sup>               | 27.6%   | 5.9   |
| SDB              | 15   | 5   | 45   | 216 <sup>2</sup>               | 20.8%   | 3   |
| All 3 Data Bases | 45   | --  | 824  | 824 <sup>1</sup>               | 100%  | 18.3  |

<sup>1</sup>Total number of logically different search inquiries made across 45 IRSs<sup>2</sup>Total number of logically different search inquiries made across those 15 IRSs in which this data base was permitted.

made (to achieve a 90% recall level for 3 experts each searching 15 IRSs) 691 (84%) were to the DDB. The remaining search inquiries were made to the VDB (11%) and the SDB (5%). Fifteen IRSs (5 for each expert) were searched under the VDB condition. Two experts each chose not to employ the VDB for one of their IRSs. Averaged over the 15 IRSs in which the VDB was permitted to be searched, experts chose to use it over 27% of the time.

The SDB had less usage. Search experts found at least one IRS in which they did not employ the SDB. Moreover, average use of the SDB dropped to approximately 20% of all inquiries made by the experts under this treatment condition. In each of the 5 times the SDB was not used, the experts achieved the 90% recall level with fewer total searches than the other experts. Across these 5 IRSs, the median difference between the fewest number of searches and the next fewest was 7 -- i.e. experts who chose not to use the SDB achieved a 90% recall level with an average of seven fewer searches than the next best performing expert.<sup>7</sup> Seen in this light, it appears that experts who are using a working strategy for retrieving relevant documents may not have the need to use the SDB for help. Perhaps they would turn to the SDB when and if their list of searching strategies became depleted. However, when asked, the search experts did not recall consciously working in the manner described.

There is an obvious need for further study of the use of the SDB; more will be discussed in this report. At this time, it is sufficient to point out that the SDB was infrequently used. The dilemma affecting an assessment of the performance of the SDB is apparent: (1) If the performance of the SDB does not differ from that of the DDB it may be due to its infrequent application rather than any intrinsic uselessness on its part; or (2) if the performance of the SDB does differ from that of the DDB it may be due to the fact that all 3 experts searched their SDB IRSs last -- so an increase in performance because of practice, or a decrease in performance because of fatigue or boredom is possible.

---

<sup>7</sup>A similar pattern is apparent, looking at the two IRSs which did not employ the VDB. One expert, not using the VDB achieved the 90% recall level with fewer logically different searches than either of the other experts. The other expert, achieved a 90% recall level with the second fewest number of searches -- but the total number of searches made (7) was small in absolute terms. The rationale postulated in the text supports the non-use of the VDB as well as the non-use of the SDB.

## b. Computer Costs

There are 4 different, but related, measures of computer cost collected by STATPAC. (1) The CPU time is the amount of processing required to execute one search. This includes input and output processing plus accessing the inverted file and using the search operators. CPU time will be presented in seconds. (2) The EXCP count is a measure of I/O activity plus any internal channel executions.<sup>8</sup> It will be given as a frequency count. (3) The number of completed searches is another indicator of computer costs. In general, only logically different, completed search inquiries were counted. (4) Finally, we have combined the CPU time and the EXCP count for a search into a dollar figure. How to combine these into one index of cost is somewhat arbitrary. We have chosen to use the rates currently being charged for the university's IBM 370/155: \$360 per hour of CPU time plus \$1.50 for every 1,000 EXCPs.<sup>9</sup>

Our plan was to retrieve from STATPAC these four cost measures for all of the completed inquiries made by the search experts. Unfortunately, several disc tracks were irretrievably lost, making it impossible to get accurate cost figures for searches in 23 IRSs. Eleven IRSs were missing cost figures for all searches, 12 IRSs had some cost figures, and the remaining 22 IRSs had cost figures for all searches. Two of the group of 12 IRSs with partial cost information had less than 20% of their searches missing. The missing costs in these two IRSs were estimated using the mean cost per search (See Table XLV). This raised the number of IRSs with complete cost data to 24. These are divided into seven complete IRSs searched under the 'D' condition, 12 searched under the 'S' condition, and 5 searched under the 'V' condition.<sup>10</sup> While each expert contributed at least one IRS

---

<sup>8</sup>The operating system's teleprocessing package executes one EXCP for every carriage return on input and output -- i.e. one EXCP for every line (regardless of size) of input and output. In addition, there are two EXCPs for every keyword in a search inquiry (one to the dictionary and one to the vocabulary.) Plus, there is approximately one EXCP for every use of a label as a keyword in a search inquiry. It is impossible to determine the exact number of EXCPs without using STATPAC or some computer monitoring system. Truncation operators, the length of internally generated temporary computation strings, and the size of buffers have an effect on the number of EXCPs.

<sup>9</sup>CPU time and EXCP count are only two of four components which determine the total computer charge for a search at Syracuse University. There is also a charge for (1) core residency (i.e. number of bytes of core required per hour); and (2) the length of time a terminal is connected to the computer. Neither of these are included as cost figures in this report as they contributed less of our understanding of the operating characteristics of SUPARS/DPS.

<sup>10</sup>These 3 experimental conditions refer to the design outlined in Table XLI. Each condition was a restriction upon which data base an expert could

TABLE XLV

## ESTIMATES OF THE COMPUTER COST OF SEARCHING THE THREE DATA BASES

|                  | DDB*   |                    | SDB**   |                    | VDB*** |                    |
|------------------|--------|--------------------|---|--------------------|--------|--------------------|
|                  | Mean   | Standard Deviation | Mean  | Standard Deviation | Mean   | Standard Deviation |
| CPU-Seconds      | 5.45   | 5.61               | 0.24  | 0.25               | 0.11   | 0.02               |
| EXCP Count       | 407.92 | 282.99             | 67.45   | 76.73              | 41.37  | 4.40               |
| Dollar Cost      | 1.16   | 0.97               | 0.13  | 0.14               | 0.08   | 0.04               |
| Searches per IRS | 15.85  | 3.93               | Estimates to these data bases are not based upon IRS's - see ** and *** |                    |        |                    |

\* Estimates are taken from the 111 logically different searches of the DDB in the seven complete IRS's in the 'D' condition.

\*\* Estimates are taken from the forty-two searches of the SDB retrieved by STATPAC whether or not they were part of complete IRS's.

\*\*\* Estimates are taken from the forty-one searches of the VDB retrieved by STATPAC whether or not they were part of complete IRS's.

under each of the three conditions, these 24 IRSs are not counterbalanced across IRSs, treatment conditions, and search experts. The cost findings presented represent less controlled estimates of cost than would be desired, and the reader is strongly cautioned to study the accompanying text material which will specify the basis for all estimates.

Table XLV presents cost estimates to the 3 data bases. All three measures of cost decrease as we go from the DDB to the SDB to the VDB. Since the DDB and the SDB are true data bases as defined by DPS, the differences in cost should be solely due to the differences in size of each data base. If the SDB contained as many searches as the DDB contains documents, we would expect the cost of searching each to be about equal.

The VDB, however, is not a true data base; it is merely part of the inverted file of the DDB. Searching it requires a much simpler set of computations. Cost for searching the VDB should remain quite low unless the number of words in the vocabulary increases by several magnitudes. The consistency of the cost of searching the VDB is shown by the very small standard deviation.

#### c. Cost-Performance

The relationships between 5 measures of cost and performance (in terms of recall ratios) for each data base are presented in Tables XLVI-XLVIII and Figures 10-11. Extreme care must be taken in interpreting these data. In Tables XLVI-XLVIII, the first column of data contains cumulative total retrievals -- that is, the number of irrelevant documents retrieved before the next one of the ten relevant ones were found. This data is based upon all 3 experts searching 15 IRSs -- 45 IRSs in all.

The other columns of data in Tables XLVI-XLVIII, though accurate, are less reliable, because they are not based upon a complete set of experimental data. The costs of the DDB were based upon 7 IRSs, those of the SDB were based upon 12 IRSs, and for the VDB, the data came from five IRSs.

The average cost per logically different search in the DDB is \$1.16 according to Table XLV. That figure can also be computed from Table XLVI by dividing the dollar cost by the number of searches (i.e.  $18.39 \div 15.85$ ). This correspondence between the two tables does not hold up with the other

---

search. Under 'D' only DDB could be searched, under 'S' both SDB and DDB could be searched, and under 'V' both VDB and DDB could be searched. It is important to distinguish between the cost of searching a data base and the cost of searching under a specific experimental condition. Only under the 'D' condition are they the same. Table XLV presents cost estimates for searching the three data bases. Tables XLVI-XLVIII present cost estimates for searching under the three treatment conditions.



TABLE XLVI

ESTIMATES OF THE MEAN COST OF SEARCHING UNDER THE 'D' EXPERIMENTAL  
CONDITION AT NINE LEVELS OF PERFORMANCE (RECALL)

(EACH FIGURE IS A MEAN; ALL COLUMNS ARE CUMULATIVE)

| Recall<br>Level | No. of Documents<br>Retrieved* | Based Upon 111 Searches in 7 IRSS |            |             |
|-----------------|--------------------------------|-----------------------------------|------------|-------------|
|                 |                                | CPU-Seconds                       | EXCP Count | Dollar Cost |
| 10%             | 297.27                         | 19.48                             | 1561       | 4.28        |
| 20%             | 534.60                         | 23.08                             | 1875       | 5.12        |
| 30%             | 614.93                         | 26.11                             | 2181       | 5.90        |
| 40%             | 745.20                         | 28.91                             | 2403       | 6.51        |
| 50%             | 1446.33                        | 39.68                             | 3334       | 8.96        |
| 60%             | 1887.40                        | 50.81                             | 4074       | 11.21       |
| 70%             | 3305.40                        | 68.85                             | 5146       | 14.57       |
| 80%             | 5200.67                        | 73.01                             | 5484       | 15.47       |
| 90%             | 7055.53                        | 86.42                             | 6468       | 18.39       |
|                 |                                |                                   |            | 4.00        |
|                 |                                |                                   |            | 4.57        |
|                 |                                |                                   |            | 6.14        |
|                 |                                |                                   |            | 6.57        |
|                 |                                |                                   |            | 8.14        |
|                 |                                |                                   |            | 10.28       |
|                 |                                |                                   |            | 12.14       |
|                 |                                |                                   |            | 13.57       |
|                 |                                |                                   |            | 15.85       |

\*This column is based upon the complete experiment -- Those 15 IRSS under the 'D' conditions

TABLE XLVII

ESTIMATES OF THE MEAN COST OF SEARCHING UNDER THE 'S' EXPERIMENTAL  
CONDITION AT NINE LEVELS OF PERFORMANCE (RECALL)

(EACH FIGURE IS A MEAN; ALL COLUMNS ARE CUMULATIVE)

| Recall<br>Level | No. of Documents<br>Retrieved* | Based Upon 163 Searches in 12 IRSs |            |             |  |
|-----------------|--------------------------------|------------------------------------|------------|-------------|--|
|                 |                                | CPU-Seconds                        | EXCP Count | Dollar Cost | No. of Logically Different<br>Different Searches |
|                 |                                |                                    |            |             |  |
| 10%             | 77.40                          | 5.96                               | 732        | 1.72        | To DDB and SDB 4.25<br>To DDB only 2.24          |
| 20%             | 155.53                         | 6.93                               | 892        | 2.06        | 5.00 2.66  |
| 30%             | 314.87                         | 9.71                               | 1119       | 2.68        | 5.75 3.24  |
| 40%             | 478.67                         | 11.07                              | 1241       | 2.99        | 6.00 3.50  |
| 50%             | 688.20                         | 13.37                              | 1442       | 3.53        | 6.67 4.16  |
| 60%             | 835.20                         | 15.99                              | 1675       | 4.14        | 7.42 4.91  |
| 70%             | 1045.53                        | 16.91                              | 1837       | 4.47        | 8.50 5.75  |
| 80%             | 1402.73                        | 24.59                              | 2554       | 6.31        | 10.83 8.15                                       |
| 90%             | 2233.80                        | 29.87                              | 3197       | 7.81        | 13.58 10.50                                      |

\*This column is based upon the complete experiment --- Those 15 IRSs under the 'S' condition.

TABLE XIVIII

ESTIMATES OF THE MEAN COST OF SEARCHING UNDER THE 'V' EXPERIMENTAL  
CONDITION AT NINE LEVELS OF PERFORMANCE (RECALL)

(EACH FIGURE IS A MEAN; ALL COLUMNS ARE CUMULATIVE)

| Recall Level | No. of Documents Retrieved* | Based Upon 104 Searches in 5 IRSS |            |             |                | No. of Logically Different Searches |  |
|--------------|-----------------------------|-----------------------------------|------------|-------------|----------------|-------------------------------------|--|
|              |                             | CPU-Seconds                       | EXCP Count | Dollar Cost |                |                                     |  |
|              |                             |                                   |            |             | To DDB and VDB | To DDB only                         |  |
| 10%          | 68.33                       | 4.09                              | 569        | 1.26        | 4.60           | 1.80                                |  |
| 20%          | 163.60                      | 4.95                              | 769        | 1.65        | 6.00           | 2.60                                |  |
| 30%          | 242.13                      | 5.32                              | 837        | 1.79        | 6.40           | 3.00                                |  |
| 40%          | 729.07                      | 19.69                             | 2005       | 4.98        | 10.00          | 5.80                                |  |
| 50%          | 916.49                      | 36.04                             | 3148       | 8.33        | 12.40          | 8.00                                |  |
| 60%          | 1219.60                     | 39.67                             | 3469       | 8.77        | 13.40          | 9.00                                |  |
| 70%          | 1517.47                     | 47.12                             | 3972       | 10.68       | 14.60          | 10.00                               |  |
| 80%          | 2014.27                     | 49.63                             | 4262       | 11.36       | 16.20          | 10.80                               |  |
| 90%          | 3193.80                     | 65.82                             | 5629       | 15.04       | 20.80          | 15.20                               |  |

\*This column is based upon the completed experiment -- those 15 IRSS under the 'V' condition.

data bases. For the SDB the average cost per search is \$0.13 (Table XLV), but according to Table XLVII it equals  $(7.81 \div 13.58) = \$0.57$ . The explanation for this difference lies in what is being counted in each table. Table XLV data is based upon searches to the SDB only, while in Table XLVII the data comes from searches completed under the 'S' condition. Under the 'S' condition, an expert was permitted to use the SDB as an aid in retrieving a document. To achieve this goal the expert was required to use the DDB to retrieve documents. The data in Table XLVII is taken from 12 IRSs under the 'S' condition. Four of these did not include any searches to the SDB. Of the 163 searches only 37 were to the SDB while 126 were to the more expensive DDB.

A similar situation occurs with the VDB. In Table XLV the average cost of a search to the VDB is \$0.08. But from Table XLVIII it is \$0.72  $(15.04 \div 20.80)$ . Again this difference occurs because Table XLVIII includes 73 searches to the more costly DDB as well as 31 searches to the VDB.

One tentative finding of this analysis of the average cost per search is that use of the VDB decreases the need to use the more expensive DDB and is, therefore, a valuable searching aid. This is evident when we look at the last column of Table XLVI and Table XLVIII, and when we compare the average cost. Since under both condition searches to the DDB had to be made, the only way the average cost per search to the VDB (\$0.72) can be less than that to the DDB (\$1.16) is if the use of the VDB reduces the need to search the DDB. At present this is a tentative finding because it is based upon incomplete unbalanced data. However, as we shall see, other evidence of a more reliable nature supports this finding.

A similar argument can be made for the SDB, but because the 'S' condition was confounded with a learning or practice effect, it is impossible to tell how valid it would be.

Another way of saying this is that the average number of logically different searches at all recall levels is about the same for the three data bases. The most they differ is approximately seven searches (at the 90% level.) While the average number of searches seems comparable, the other three estimates of computer costs differ widely across the 3 experimental conditions. In general, the 'D' is the most expensive, the 'S' the least expensive and the 'V' extends the full range between them. This general finding is true for all three measures and across the nine levels of recall. Figure 10 presents graphically the cost-performance differences between the 3 in computer dollar costs.

Upon inspection, we determined that the radical decrease in performance of the 'V' condition between the 30% and the 50% recall levels was due solely to one IRS. This may be artificial; if the remaining 10 IRSs in the 'V' condition were available, the performance decrease could conceivably be averaged down. A similar pattern emerges if we plot either of the other two computer cost measures.

It should be noted that all of the computer cost measures of the 3 conditions (with the possible exception of 'V') appeared to be log-normally

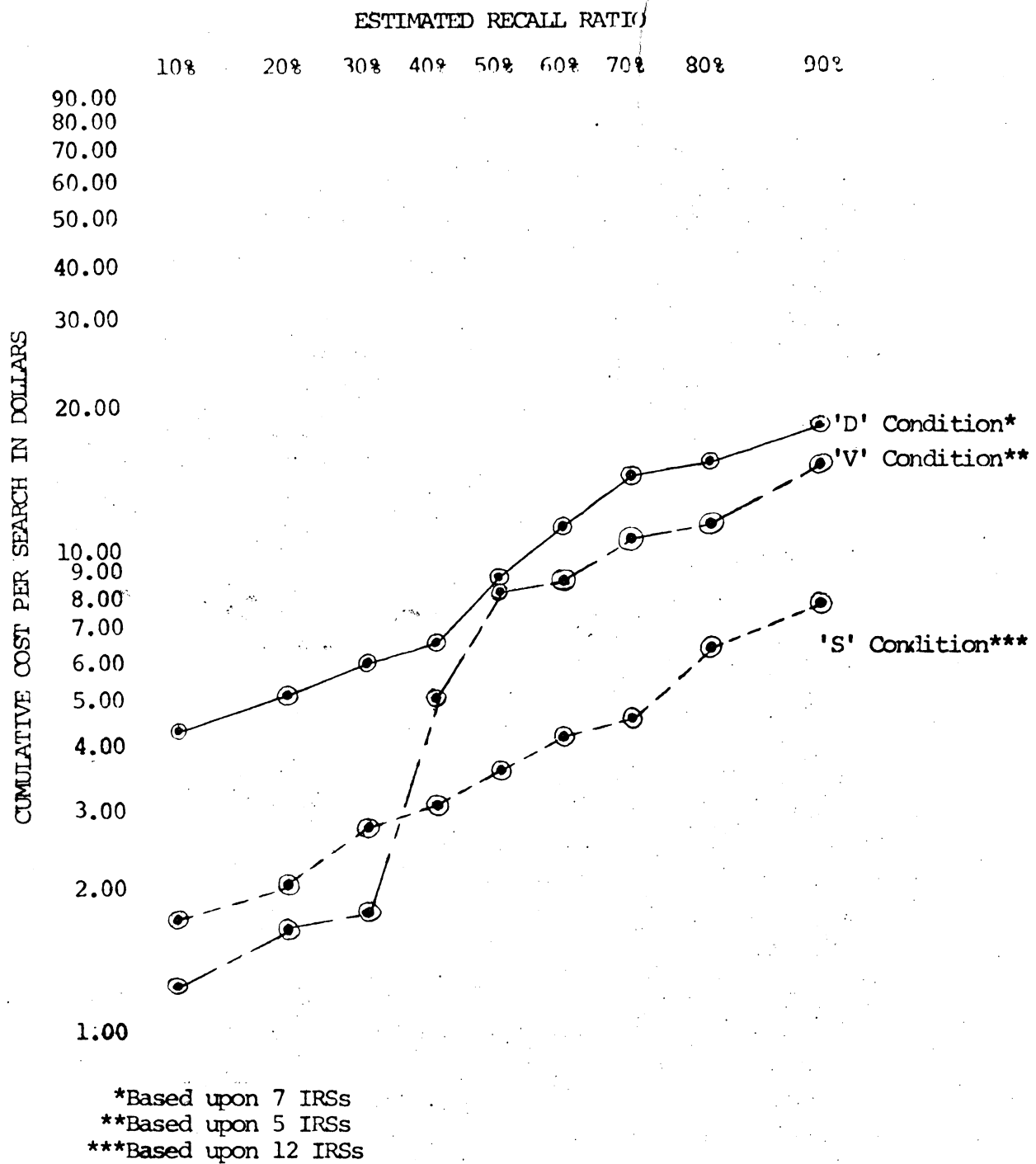
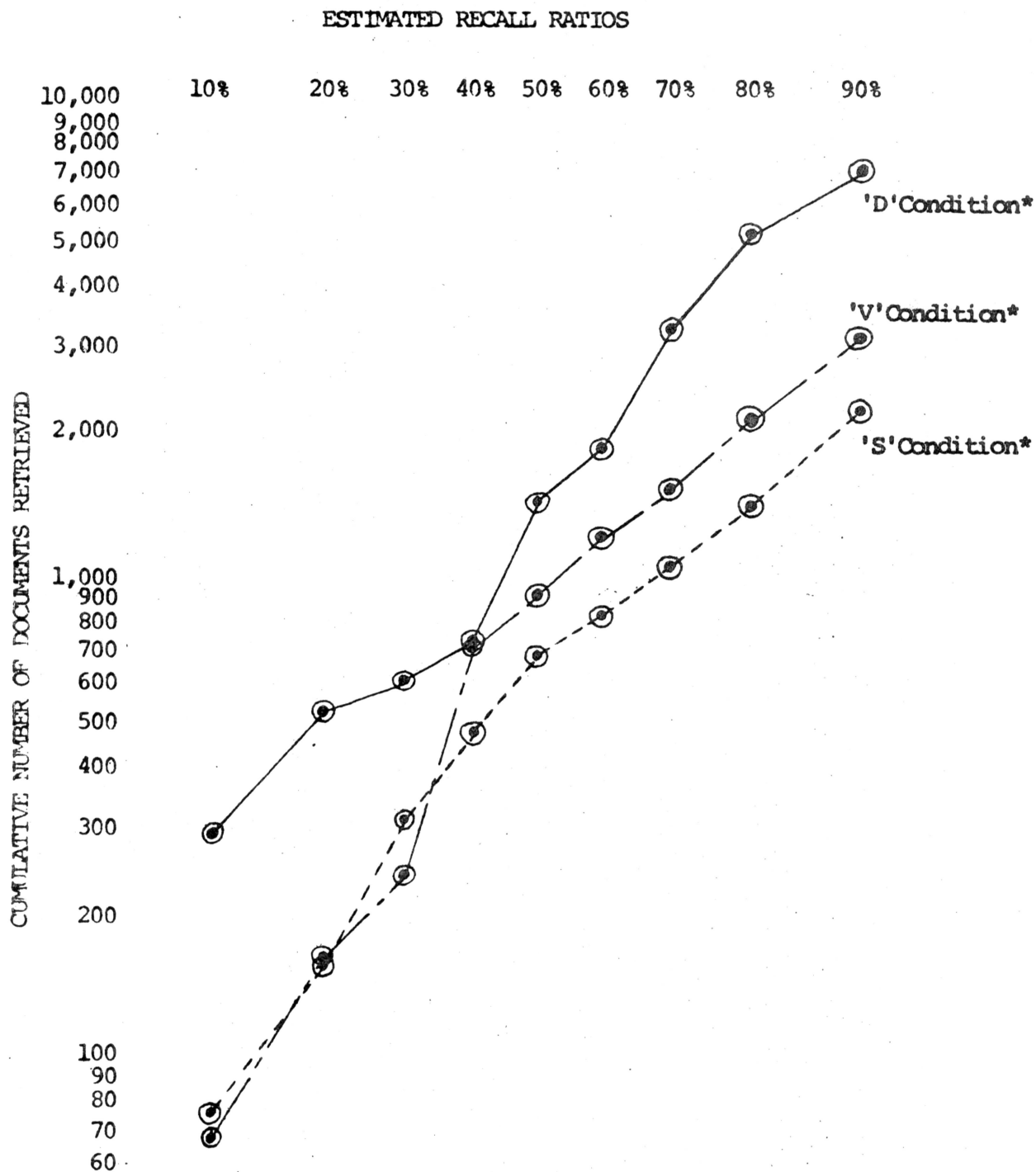


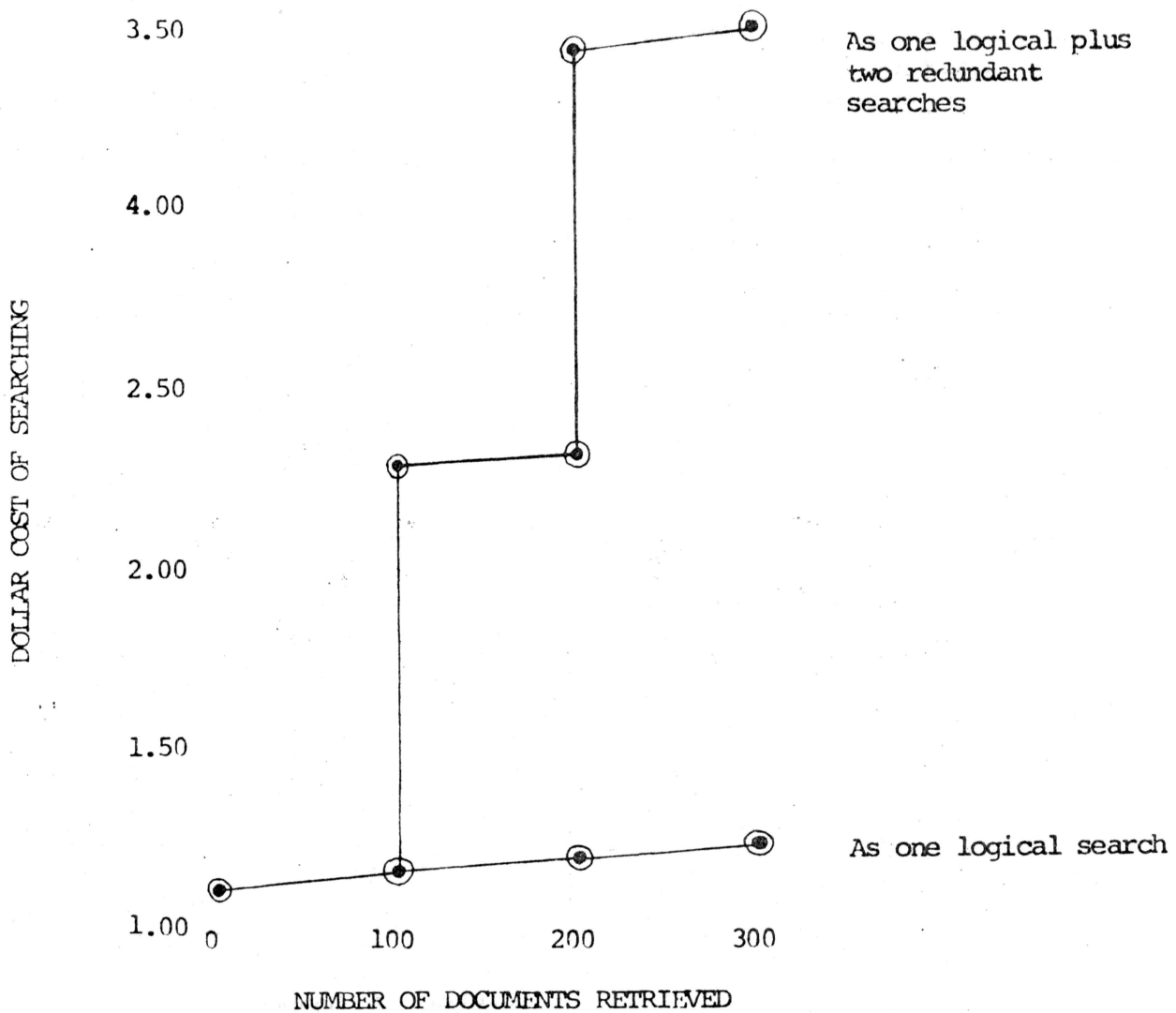
Figure No. 10. Computer Cost per Search To Achieve Nine Levels of Recall:  
A Comparison Among Three Treatment Conditions



\*Based upon 15 IRSs in Complete Study

Figure No. 11. Number of Documents Retrieved (Total Retrieval) to Achieve Nine Levels of Recall: A Comparison Among Three Treatment Conditions.





\*This splits the mean cost per search of \$1.16 (Table XLI) into \$1.12 for input processing and searching, and \$0.04 for printing 100 LIST BRIEF citations (25 lines of output).

Figure No. 12. Two Methods of Computing Dollar Cost of Searching on DDB\* Using LIST BRIEF Output Format

distributed. Since the computer cost estimates are derived from a portion of the total experimental design, it was impractical to test this finding statistically. For the same reason, we did not attempt to determine the "significance" of the differences between the data bases on any of the computer cost measures.

This is an appropriate place to explain the rationale underlying our reliance, this year, on logically different searches. As noted in the definitions, logically different searches are somewhat independent of the arbitrary system parameter which specifies the maximum amount of printed output. In this sense, logically different searches represent a good estimate of the minimum computer costs of operating SUPARS/DPS.<sup>11</sup> Figure 12 presents a plot of costs as a function of number of documents retrieved by a search. For each logically different search there is the initial cost of input and processing, plus a steady slight increase as a function of the number of documents printed. Under the current system's parameter with LIST BRIEF output, the user is required to pay for the input and search processing after every 100 documents are printed. Costs under the current system can be described as a step-function where the size of the step depends on CPU processing, and the number of steps for a given amount of output depends upon an arbitrary parameter of the system. By counting logically different searches, we attempted to simulate the minimal cost conditions in which the output parameter is set to the maximum.

Up to this point in the discussion of cost-performance characteristics of the 3 experimental conditions, we have been solely concerned with various estimates of computer costs as our dependent variable. The findings have been useful to get some idea of the cost of making different types of SUPARS/DPS searches. Comparisons among the conditions and data bases were considered tentative because of the incomplete nature of the data.

The major dependent variable, number of documents retrieved, has not as yet been discussed. This variable is based upon the complete experiment, and except for the interpretation of the performance of the SDB, is more reliably interpretable. Total retrieval for the 3 data bases is presented in the first column of Tables XLVI-XLVIII. It is the same measure used to evaluate the performance of last year's version of SUPARS.

Figure 11 graphs these data on a log-normal grid. A statistical analysis supports an inspection of the data; there is a difference between the treatment means at the higher recall levels (70%, 80%, 90%.) The analysis of variance summary tables are presented in the Appendix.

Figure 11 supports the notion that use of the VDB as a searching aid materially helps improve the cost-performance of the system. This is particularly true at the higher levels of recall. Figure 11 also suggests the value of the SDB, as it has the best performance curve. However, because

---

<sup>11</sup>This is a slight underestimation as we need to add \$0.0015 for every line of output above that permitted by the system's operator.

of the lack of control over the 'S' condition, a plausible alternative explanation is that search experts learn how to perform better over time. It is unfortunate that this study is incapable of adequately determining the value of the data base of searches.

d. Comparing the Cost-Performance of Two Years of SUPARS Operation

Two comparisons are of interest here. First, a comparison of cost-performance in terms of computer cost. This is useful because the two years of SUPARS operation included a major change in computer systems. Second, a comparison of cost-performance in terms of total retrieval. This will give us some idea of the stability of the DDB, IRSs, and experts' ability to search efficiently.

As would be expected, many changes have occurred between Fall, 1970 and February, 1972, which will affect this comparison. The major changes are outlined in Table XLIX. As we can see from the Table, it will be impossible to make straightforward comparisons. Three factors affect the comparison of computer cost; the third one affects the comparison of total retrieval; (1) the charges for computer service (from 360/50 to 370/155 increased 80% for CPU use and 36% for EXCP activity, (2) the performance capability of the two computer systems differ,<sup>12</sup> and (3) as noted in the last three columns of Table XLIX, there are differences in terms of what searches were counted and how EXCPs were treated.

In addition, there are several other differences between the two operations which may have an affect on any comparison. The size of the DDB grew from 35,874 in 1970 to 46,828 in 1972. Moreover the free-text words were coded differently: 16 bits per word in SUPARS I, and twice that in SUPARS II. This coding difference permitted an increase in the number of unique words in the inverted file from 64,534 to 106,702. The extent to which these changes affect any comparison between the two systems is not known.

To obtain a rough estimate of the average cost per search to the DDB from this year's data, which would be comparable to that of SUPARS I, we recommend the 3 computer cost estimates of searching the DDB reported in Table XLV by counting all 246 searches in the 7 IRSs. Of the 246 searches 111 are logically different (reported in Table XLV) and the remainder are the "redundant" ones.<sup>13</sup> Table L gives these cost figures.

---

<sup>12</sup>Depending upon the nature of the data and the type of operation being performed, the 370/155 is 3-4 times faster than the 360/50 in terms of computation. The 3330 disc devices have a transfer rate of 2.6 times faster than the 2314s. Seek time across a cylinder is 50% faster and time within a track is 33% faster in the newer discs. We cannot determine what this increase in computer performance means to SUPARS specifically. As a data retrieval system, the improvement in disc performance probably affects SUPARS more strongly than raw computation power.

<sup>13</sup>Since these make up 7 complete IRSs the relative proportion of logically different searches to "redundant" searches should be reasonably representative of the use of the DDB.

TABLE XLIX

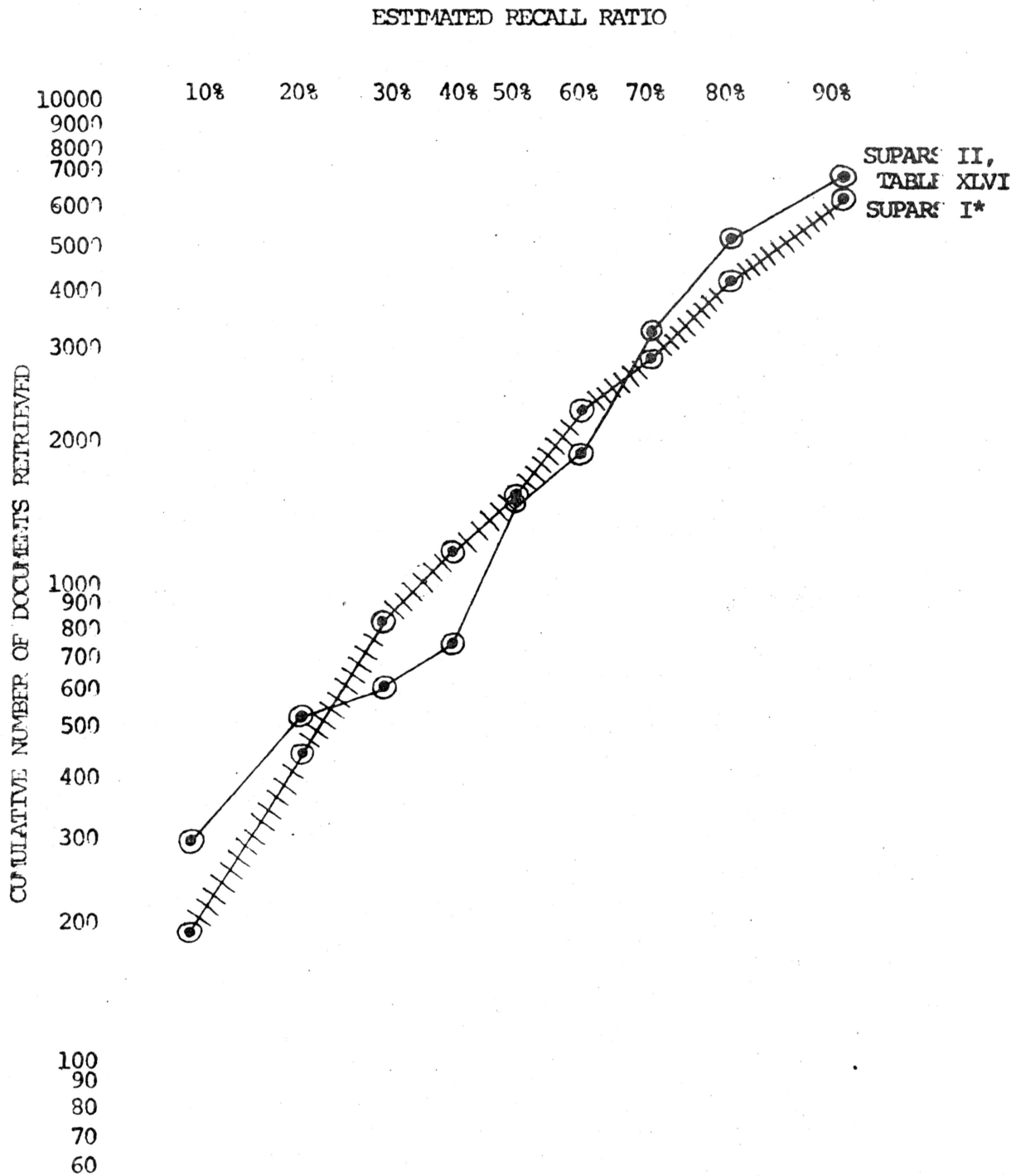
## DIFFERENCES BETWEEN SUPARS I AND SUPARS II

| Operating Period                 | Hardware Configuration and Charge for Use                                  | Searches to the DDB   | Experimental Conditions  | EXCP Count  |
|----------------------------------|--|---|--|---|
| SUPARS I<br>FALL 1970            | IBM 360/50 with 2314 Disc Drives at \$200/CPU hour plus \$1.10/1000 EXCPs  | All attempts at searching were counted including erroneous starts and "redundant" searches.   | Searches to the DDB were divided into two levels of search logic; simple and advanced. There are separate cost-performance estimates for each. | Only those EXCPs used in printing output were counted. Charge for these EXCPs was incorporated with CPU into a dollar cost. |
| SUPARS II<br>FALL 1971           | Same as above  | None of this data was analyzed. It consists of searches, performance characteristics, cost, etc. of SUPARS II operation by users. The only data studied is that produced by the search experts as they carried out the experiments. |  |   |
| Jan-Feb 1972<br>(Use by Experts) | IBM 370/155 with 3330 Disc Drives at \$360/CPU hour plus \$1.50/1000 EXCPs | Only logically different searches on the DDB, there is no completely entered separation into logic were counted.  | Within the data collected on the DDB, there is no separation into logic levels.  | All EXCPs were counted and reported as a separate cost.   |

TABLE L

ESTIMATES OF THE COMPUTER COST OF SEARCHING TO  
DDB COUNTING ALL ENTERED SEARCHES (N=246)

|                                       | Mean | Standard<br>Deviation |
|---------------------------------------|------|-----------------------|
| CPU-<br>Seconds                       | 6.45 | 5.79                  |
| EXCP<br>Count                         | 493  | 290                   |
| Dollar<br>Cost<br>Using 1972<br>Rates | 1.38 | 1.00                  |



\*Based upon SUPARS I data averaged at 9% simple and 91% advanced levels of search logic.

Figure No. 13. Number of Documents Retrieved (Total Retrieval) to Achieve Nine Levels of Recall: A Comparison Between SUPARS I and SUPARS II Searches to the DDB



The costs in Table L are somewhat higher than those in Table XLV. This is likely due to the additional internal processing required of many of the "redundant" searches. Search experts frequently used redundant searches to skip quickly down the list of retrieved documents. To do this they added to their search query an additional command specifying, in effect, how far to skip. This additional command was probably the cause of the cost increases.<sup>14</sup>

The estimated mean cost per search inquiry last year was \$0.589 for search inquiries employing simple logic, and \$1.097 for inquiries using advanced search logic.<sup>15</sup> A search inquiry employed the simple level of logic if it only used the AND or OR Boolean operators. As used by the search experts in last year's study, the advanced logic level included, for the most part, search inquiries employing the NOT and the Truncation Operators as well as the two simple level operators.

In order to determine the type of logic used by search experts this year, we counted all 293 inquiries completed under the 'D' condition. Simple logic operators made up only 9% of all the searches, and advanced logic operators appeared in the remaining 91% of the inquiries. Therefore, a hypothetical search inquiry to last year's DDB which would be comparable to the logic used by the experts this year, would be 9% simple and 91% advanced. The cost of such a hypothetical inquiry in terms of the rates charged on the IBM 360 would be \$1.05.

The \$1.05 represents the cost for an average search inquiry to last year's system which corresponds to the type of inquiries made this year. Our goal is to convert this figure to one based upon the IBM 370 charge rates (see Table XLIX). When this conversion is completed we can compare it with the actual charge for an average search to the DDB of SUPARS II, which is \$1.38 (Table L).

To convert the \$1.05 to one determined according to the IBM 370 rates,<sup>16</sup> we increased 40% of \$1.05 by 80% (the rate change for CPU time,) and 60% of \$1.05 by 36% (the rate change for EXCP.) The result of this arithmetic gives us an estimate of the cost per search inquiry last year, computed at this year's rates: it is \$1.63. This year's dollar cost of \$1.38 is about 14% lower than last year's.<sup>17</sup> Most of the increased efficiency this year is

---

<sup>14</sup>There is a minor reduction of cost of redundant searches because the entire search query does not have to be retyped. This decrease in the number of input EXCPs is too small to have a noticeable affect.

<sup>15</sup>These figures came from Table 7 of Section 5 of last year's Final Report.

<sup>16</sup>The dollar cost of an average search inquiry at last year's charges can be broken down as follows: 40% of it is due to CPU time and 60% is due to EXCP activity. To compute this corresponding rate for this year, we merely increased the appropriate portion of the dollar cost by the percentage rate increase for CPU and EXCP.

<sup>17</sup>The arithmetic did not take into consideration two factors which raise the converted cost for last year slightly above \$1.63; first, that last year's

probably due to the new hardware configuration (see Footnote 12.)

We can compare the two SUPARS systems in terms of total retrieval. Figure 13 graphically demonstrates the equivalence of the two systems when number of documents retrieved is used as a measure of cost. This suggests that on the average, experts, IRSs, and the DDB did not change between the two systems in any way to seriously affect total retrieval at any level of performance.

Finally, we can compare the best cost-performance relationship of SUPARS I, with SUPARS II. In this year's study experts did not choose to restrict search inquiries of the DDB to the simple level of logic -- which had the best cost-performance relationship last year. In fact, only 9% of their inquiries to the DDB used at most simple logic operators. Given these strong preferences of the experts, it would be worthwhile to see how use of the SDB and VDB improved the efficiency of searching SUPARS II, compared with the optimal simple logic strategy. Figure 14 plots these curves. Figure 14 duplicates Figure 11 with the additional cost-performance curve illustrating the most efficient version of SUPARS I. Without testing the difference among these curves, it appears that experts searching under the 'D' condition did not perform as efficiently as possible. This decreased efficiency, however, can be offset through judicious use of the VDB and possibly the SDB -- especially at the higher recall levels.

#### 4. DISCUSSION

##### a. The VDB

The remarkable finding of this study is the discovery of the power of the VDB. Though many others<sup>18</sup> have called for an on-line thesaurus with interactive retrieval systems, we were pleasantly surprised at its cost and its cost-performance characteristics.

Since our VDB is, in essence, a portion of the inverted file of the DDB, it does not take additional space. There are some user interaction commands to and from the VDB, plus searching routines which had to be written to access the inverted file, but these are relatively easy programming tasks, and the new instructions take little space. From a systems design point-of-view, the modifications of DPS to make the VDB accessible are not expensive.

The contribution of the VDB to the system seems quite clear. On two measures of cost (computer dollar cost and total retrieval) use of the VDB helps a searcher to use the DDB more efficiently. And, this help is most observable at the higher recall levels when the "going gets rough" for a

---

EXCP count was based solely on output printing charges, and second that last year's data base size and vocabulary size were less than this year's. In other words, this year's dollar cost for an average search inquiry is at least 14% lower than last year's.

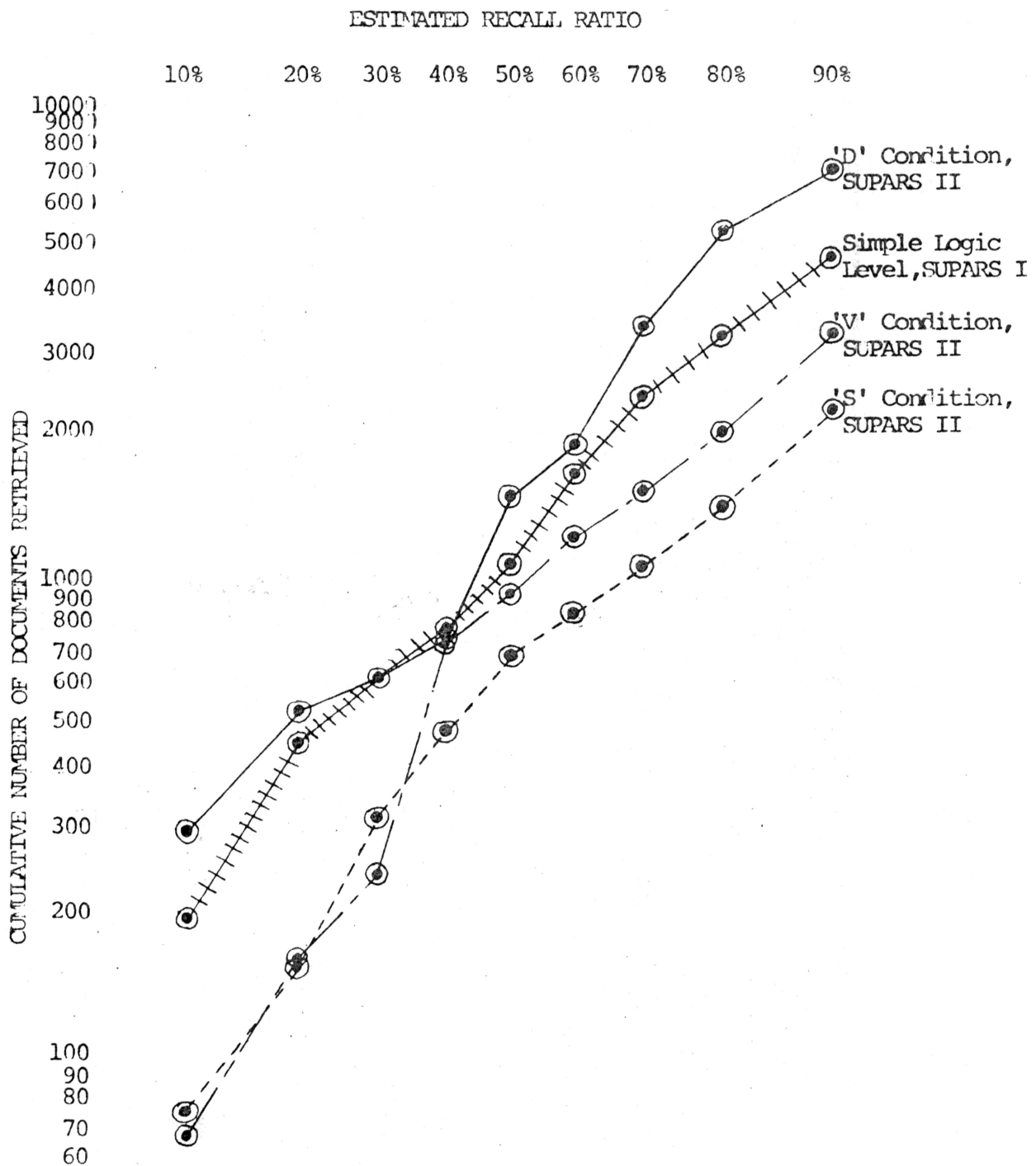


Figure No. 14. Number of Documents Retrieved (Total Retrieved) to Achieve Nine Levels of Recall: A Comparison Among the SUPARS II Conditions and the Most Efficient Use of SUPARS I

searcher. Apparently, use of the VDB decreases the need to search the DDB. This is supported by the data presented in the last columns of Tables XLVI and XLVIII.

The search experts used the VDB primarily as an adjunct to the truncation operator. With a free-text system, a truncation operator is a great help pulling together many words with the same root, but with different suffixes. To search a system such as SUPARS without a truncation operator might be more efficient (See Figure 13), but it would be more arduous for the searcher. This is suggested by the fact that approximately 90% of the search inquiries to the DDB included the truncation operator -- even though the experts were admonished to minimize total retrieval. Use of the truncation operator occasionally decreases the precision of a search because it may retrieve documents containing irrelevant keywords with the same root as the desired keyword. By entering the root into a search inquiry of the VDB, one can identify those irrelevant keywords and eliminate them from the inquiry to the DDB. This pairing of the truncation operator with the VDB allows an expert to keep the searching easy, but maintain a high precision ratio. An added benefit to the pairing of these two searching aids is that the cost of searching the VDB is quite low (See Table XLV); searchers can be encouraged to use the VDB liberally with marginal increases in cost.

Finally, it should be noted that the cost-performance of SUPARS II under the 'V' condition is better than that of the most efficient version of SUPARS I.<sup>19</sup> Thus, the relationship between the truncation operator and the VDB is not simply a trade-off -- what one loses in performance, the other gains. Rather, the VDB appears to be a worthwhile investment regardless of the use of the truncation operator.

#### b. The SDB

The SDB was developed as a means of synonym control working within the constraints of a free-text system.<sup>20</sup> The SDB is one way to develop a list of synonym-like words when such a list is not available from either the publisher of the data base or professional indexers.

As shown in Figures 10 and 12, the 'S' experimental condition proved to be most efficient. However, search experts are unanimous in their agreement that use of the SDB was not noticeably helpful. This, plus the fact

---

<sup>18</sup>For example, see J. H. Williams, Jr. "Functions of a Man-Machine Interactive Information Retrieval System." Journal of the American Society for Information Science, Volume 22. 1971. Pages 311-317.

<sup>19</sup>Though we have not tested this statistically. As shown in Figure 14, the curve for the 'V' condition is lower than the curve representing the most efficient version of SUPARS I, and this is true for all nine levels of recall.

<sup>20</sup>Another use of the SDB is to provide a data base for those who wish to study how searchers interact with an on-line retrieval system, but this is not of interest here.

that the 5 IRSs under the 'S' condition were searched last by the experts, leads to the conclusion that the cost-performance curves depict a learning effect. While not an unpalatable finding, it is useful to know that experts can improve their searching ability noticeably over time.

In terms of evaluating the SDB as a searching aid, the evidence suggests that in its present form it was not useful. The usefulness of the SDB may increase with its size. The current data base consists of 4,235 searches, which is not a large number considering the numerous topics covered in the 15 IRSs searched by the experts (See Table XLII). Another factor which lessened the use of the SDB was the knowledge of the search experts. As frequent users of SUPARS, with formal training in psychology, the 3 search experts many times found it easier to generate synonyms from memory than from a small SDB. Perhaps less frequent users of the system or less knowledgeable users in the subject area would have a greater need for the SDB.

Since the SDB is a true data base, a system designer has to allocate a large amount of storage for it. And, as its size increases he would expect the cost of a search inquiry to the SDB to rise to that of an inquiry to the DDB. Thus, a SDB is a relatively large investment for a retrieval system which might only pay off after several months or years when the total number of searches to the DDB is large enough to constitute a viable SDB. If a SDB is too costly, a system designer should consider alternative means to make synonym-like words available to the searcher of a free-text system.<sup>21</sup>

#### c. Differences Among IRSs and Experts

The 15 IRSs differ according to the ease with which experts can achieve a 90% recall level. The magnitude of the difference is staggering. Three experts achieved a 90% recall level for IRS 7 with an average total retrieval of 242 documents. The corresponding figure for IRS 2 was 11,321 -- 47 times as many documents. The difference has little to do with SUPARS per se and probably cannot be reduced (toward the lower number) by the system designer. Three interrelated factors contributed to the difference: (1) the generality and/or specificity of the IRS as written by the subject, (2) the stringency with which the relevance criterion was applied when the subject identified the relevant documents and (3) the breadth of topic areas published in Psychological Abstracts -- each topic differing in the specificity of its technical language.

It is likely that the cost-performance of SUPARS/DPS would improve if

---

<sup>21</sup>DPS has one such alternative. The system designer can load a list of synonyms or equivalent terms into a DPS file. The searcher can use this file, on option, to augment the keywords in his search inquiry. Since SUPARS does not have such a list available, we could not make use of this option.

its DDB covered fewer topics, and each topic had fewer terms referring to each important concept.

As one would expect, the 3 experts differ in their ability to use SUPARS efficiently. The size of this difference might be unexpected however. At the 90% recall level the total retrieval, averaged over 15 IRSs, ranged from 707 to 7,840.<sup>22</sup> Search experts, equally trained and equally knowledgeable have characteristically different ways of attacking an IRS. At this time we do not know what the differences are among the experts which account for the differences in total retrieval.

Once these differences are identified, one can choose search experts or intermediaries more selectively. Or, if the system is available to a general population, the system designers can use this information when they develop the user interactive language and when they prepare training materials for the potential user.

If differences between trained users of a retrieval system are of this magnitude generally, then intensive study into this matter is needed, because better training of users or a more adaptive interactive language might contribute more to improving the cost-performance characteristics of a system than costly software developments.

## 5. SUMMARY OF FINDINGS

Several of the major findings of this study of the cost-performance characteristics of SUPARS II are listed below. The listing is only meant as a partial summary and as such does not include explicit limits, exceptions, or explanations of the findings.

a. Search experts did not find the SDB to be a useful aid in retrieving documents. The 'S' experimental condition did have the best cost-performance characteristics, but this was attributed to learning effects, rather than to any intrinsic value of the SDB.

b. The VDB is a good investment in a free-text system such as SUPARS. It is relatively inexpensive to add to the system, and search inquiries into the VDB are the lowest costing of all 3 data bases and should remain so as the data bases grow. More importantly, use of the VDB decreases the need to search the more expensive DDB. Therefore, inquiries carried out under the 'V' condition had a better cost-performance relationship than searches under the 'D' condition.

c. Searches carried out under the 'D' condition had the poorest cost-performance curve.

---

<sup>22</sup>The differences among experts are statistically significant at all recall levels except the lowest (10%).

d. The cost-performance curves of searches to the DDB in SUPARS I and SUPARS II are remarkably similar insofar as total retrieval is concerned. DDB search inquiries in SUPARS II are about 14% less expensive in terms of dollar cost than similar inquiries carried out in SUPARS I (using the same charge for computer services.) The decreasing cost is attributed to the improvement of hardware performance in the new IBM 370/155.

e. Search experts found the truncation operator too valuable to discard, even though the evidence from SUPARS I shows it decreases cost-performance. The use of the VDB paired with the truncation operator seems to be a useful match: It meets the searching style of the experts and improves the cost-performance beyond that of the best sub-system tested in SUPARS I.

f. Some means of synonym control is needed. The SDB is one means of achieving it when other alternatives are not available. However, the value of the SDB may not become evident until the number of stored searches is quite large. For many systems this requires a large initial investment with no consistent pay-off for several months.

g. Some study into the differences between searching styles may conceivably contribute greatly to the cost-performance of a system.

APPENDIX VII  
INSTRUCTIONS TO SUBJECTS



## APPENDIX VII

### INSTRUCTIONS TO SUBJECTS

Choose a specific or general topic you need information for right now. If you are doing a paper or planning a talk you probably have a topic in mind. If you don't have any current topic you are working on, consider one you are familiar with.

In order to acquire this information for your topic we want you to write down your information requirements as if you were talking to a colleague who understood the field as well as you do.

(1) Start off by making a broad general statement. Give an overview of your topic.

(2) When that is completed give as much specific information on your topic as you would give to a colleague.

(3) After you have written the summary, make certain that you have given as much specific information as possible. Write down such things as the major author and/or people in your topic area, recent publications, related concepts to the topic, or any other clues that relate to your topic.

In writing this summary give us enough of your thoughts so that we can theoretically go out and get this information for you.

APPENDIX VIII  
ANALYSIS OF VARIANCE SUMMARY

| Source of Variation  | Sum of Squares | df | Mean Square |
|----------------------|----------------|----|-------------|
| Among                |                |    |             |
| Treatment Conditions | 504170         | 2  | 252085      |
| Experts              | 523088         | 2  | 261544      |
| IRSS                 | 2136598        | 14 | 152614      |
| Residual             | 3162212        | 26 | 121623      |
| Total                | 6326068        | 44 |             |

Standard Error of the Mean = 90.05

Standard Error of the Difference Between Means = 127.34

Figure No. 1. Differences Among Three Treatment Conditions  
at the 10% Recall Level

| Source of Variation  | Sum of Squares | df | Mean Square |
|----------------------|----------------|----|-------------|
| Among                |                |    |             |
| Treatment Conditions | 1406988        | 2  | 703494      |
| Experts              | 2223255        | 2  | 1111628     |
| IRSs                 | 5493483        | 14 | 392392      |
| Residual             | 5826493        | 26 | 224096      |
| Total                | 14950219       | 44 |             |

Standard Error of the Mean = 122.23

Standard Error of the Difference Between Means = 172.86

Figure No. 2. Differences Among Three Treatment Conditions  
at the 20% Recall Level

| Source of Variation  | Sum of Squares | df | Mean Square |
|----------------------|----------------|----|-------------|
| Among                |                |    |             |
| Treatment Conditions | 1171550        | 2  | 585775      |
| Experts              | 3558931        | 2  | 1779465     |
| IRSs                 | 10116553       | 14 | 722611      |
| Residual             | 6884394        | 26 | 264784      |
| Total                | 21731428       | 44 |             |

Standard Error of the Mean = 132.86

Standard Error of the Difference Between Means = 187.89

Figure No. 3. Differences Among Three Treatment Conditions  
at the 30% Recall Level

| Source of Variation  | Sum of Squares | df | Mean Square |
|----------------------|----------------|----|-------------|
| Among                |                |    |             |
| Treatment Conditions | 670002         | 2  | 335001      |
| Experts              | 10826397       | 2  | 5413199     |
| IRSs                 | 23116099       | 14 | 1651150     |
| Residual             | 19091616       | 26 | 734293      |
| Total                | 53704115       | 44 |             |

Standard Error of the Mean = 221.25

Standard Error of the Difference Between Means = 312.90

Figure No. 4. Differences Among Three Treatment Conditions  
at the 40% Recall Level

| Source of Variation  | Sum of Squares | df | Mean Square |
|----------------------|----------------|----|-------------|
| Among                |                |    |             |
| Treatment Conditions | 4538354        | 2  | 2269177     |
| Experts              | 23623521       | 2  | 11811761    |
| IRSSs                | 43113614       | 14 | 3079544     |
| Residual             | 32147044       | 26 | 1236425     |
| Total                | 103422533      | 44 |             |

Standard Error of the Mean = 287.10

Standard Error of the Difference Between Means = 406.02

Figure No. 5. Differences Among Three Treatment Conditions  
at the 50% Recall Level

| Source of Variation  | Sum of Squares | df | Mean Square |
|----------------------|----------------|----|-------------|
| Among                |                |    |             |
| Treatment Conditions | 8504225        | 2  | 4252113     |
| Experts              | 33061006       | 2  | 16530503    |
| IRSSs                | 71507851       | 14 | 5107704     |
| Residual             | 47738801       | 26 | 1836108     |
| Total                | 160811883      | 44 |             |

Standard Error of the Mean = 394.87

Standard Error of the Difference Between Means = 494.79

Figure No. 6. Differences Among Three Treatment Conditions  
at the 60% Recall Level

| Source of Variation  | Sum of Squares | df | Mean Square |
|----------------------|----------------|----|-------------|
| Among                |                |    |             |
| Treatment Conditions | 42632120       | 2  | 21316060    |
| Experts              | 77944731       | 2  | 38972365    |
| IRSs                 | 138288573      | 14 | 9877755     |
| Residual             | 83901321       | 26 | 3226974     |
| Total                | 342766745      | 44 |             |

Standard Error of the Mean = 463.82

Standard Error of the Difference Between Means = 655.94

Figure No. 7. Differences Among Three Treatment Conditions  
at the 70% Recall Level

| Source of Variation  | Sum of Squares | df | Mean Square |
|----------------------|----------------|----|-------------|
| Among                |                |    |             |
| Treatment Conditions | 124757078      | 2  | 62378539    |
| Experts              | 194712111      | 2  | 97356056    |
| IRSs                 | 327989517      | 14 | 23427823    |
| Residual             | 289814769      | 26 | 11146722    |
| Total                | 937273475      | 44 |             |

Standard Error of the Mean = 862.04

Standard Error of the Difference Between Means = 1219.11

Figure No. 8. Differences Among Three Treatment Conditions  
at the 80% Recall Level

| Source of Variation  | Sum of Squares | df | Mean Square |
|----------------------|----------------|----|-------------|
| Among                |                |    |             |
| Treatment Conditions | 195418483      | 2  | 97709242    |
| Experts              | 382735553      | 2  | 191367777   |
| IRSs                 | 485520111      | 14 | 34680008    |
| Residual             | 320914847      | 26 | 12342879    |
| Total                | 1384588994     | 44 |             |

Standard Error of the Mean = 907.12

Standard Error of the Difference Between Means = 1282.85

Figure No. 9. Differences Among Three Treatment Conditions  
at the 90% Recall Level