Table III. Number of Abstracts Prepared Annually

	CBAC	Ringdoc
1965	8,393	35,500
1966	14,187	38,500
1967	15,530	40,000
1968	14,423	44,000
1969	15,201	39,000
1970	17,026	43,000
1971	16,787	39,000

not be abstracted by CBAC unless biochemical aspects were also discussed.

It will be observed that Ringdoc too had its "misses"; in the Fenclonine case this amounted to a single reference which described the action of some drugs on the CNS after pretreatment with *p*-chlorophenylalanine. The aspect of "pretreatment" subordinates the significance of the *p*-chlorophenylalanine in the eyes of *Ringdoc* and was thus not coded. The single reference in the Herpes search and the two in the Penicillins search which *Ringdoc* "missed" were veterinary in nature and as such abstracted in the ancillary *Vetdoc* service instead.

The more extensive journal coverage of CBAC over *Ringdoc* is also apparent from the bar graphs in Figure 2 where the range of references from non-*Ringdoc* journals varied between 10–19%. An examination of the journal lists of the two data bases showed that 253 of the journals were common to both services (thus 75% of the *Ringdoc* journals were also covered by CBAC). The remainder of journals covered by each service was also scrutinized for any possible specialist trends but neither CBAC nor *Ringdoc* appeared to exhibit any. From the degree to which *Ringdoc* performed in these searches, it is very clear that their selection of 332 journals has been most successful. While more extensive journal coverage is in theory an asset, the increased rewards, as perhaps in the case of CBAC, are not an adequate substitute for careful article selection.

LITERATURE CITED

(1) Ish, C. J., and Terrant, S. W., Jr., "Chemical-Biological Activities: A specialized information service in Biochemistry," *Amer. J. Pharm. Educ.* **1968** (32) 201–10.
(2) Oatfield, H., "The ARCS System: *Ringdoc* as Used with a Computer," *J. Chem. Doc.* **1967,** (7) 37.
(3) Smith, D. R., Beauchamp, R. O., Jr., Barber, J. L., and Daugherty, M. A., "Computerized Drug Information Services," *J. Chem. Doc.* **1972,** (12) 9.

Comparative Searching of Computer Data Bases*

R. O. BEAUCHAMP, JR.,** MARVEL A. DAUGHERTY, JOSEPHINE L. GARBER, and JOHN D. MYERS
A. H. Robins Co. Research Laboratories, Richmond, Va.

Received June 1, 1972

Methods for retrieval of information on chemical compounds utilizing several computer data bases have been compared to determine scope of data base coverage. Queries for a single chemical, *N*-ethyl-α-methyl-*m*-trifluoromethylphenethylamine (fenfluramine) and generic chemicals (2-pyrrolidinones) were submitted to the services for searching through the recent literature. Data bases employed included MEDLARS, *Excerpta Medica, CA Condensates,* CBAC, *Ringdoc, Current Abstracts in Chemistry* and *Automatic New Structure Alert* (ANSA). Preparation of search questions is outlined and comparative results are reported indicating the yield from each data base.

In recent years several computer data bases have become available which encompass large areas of technical information. The spectrum of pharmaceutical information contained in these bases covers synthetic organic chemistry, biochemistry, pharmacology, toxicology and veterinary and clinical medicine. Other areas of interest include pharmacy, analytical chemistry, metabolism, radioisotopes, biopharmaceutics, microbiology, and physical chemistry. The user may obtain information from journals, symposia, reviews and patents, but many data are retrieved on a current awareness basis and retrospectively, from intermediary abstract sources. Such central sources are increasingly important in providing information necessary for the preparation of new chemicals and drugs and in keeping the scientist aware of developments in his chosen area.

The principal abstract sources available currently are *Chemical Abstracts, Biological Abstracts, Ringdoc, Vetdoc, Central Patents Index* (including *Farmdoc*), *Excerpta Medica,* and *Index Medicus.* Other services are *Patent Uniterm Index, Science Citation Index* (ASCA), and *Current Abstracts of Chemistry* and *Index Chemicus* (ANSA) and the experimental base, *Integrated Subject File* (ISF, the Chemical and Subject Files). Each of the aforementioned abstract compilations can be searched by means of computer. In the case of *Chemical Abstracts,* two computerized bases can be accessed: *Chemical Abstracts Condensates* (CAC) and *Chemical-Biological Activities* (CBAC). *Index Medicus* is searched via MEDLARS. Smaller sets of these data bases have been compiled for

*Presented at the Division of Chemical Literature, "Symposium on the Evaluation of Existing Chemical Information Services," 163rd Meeting, ACS, Boston, Mass., April 10, 1972.

**To whom correspondence should be addressed. Present address: Toxicology Information Response Center; Oak Ridge National Labs.; Oak Ridge, Tenn. 37830.

special searches—e.g., for limited coverage, an abridged MEDLARS is available for on-line bibliographic searching by MEDLine.

As part of a continuing program to improve present information retrieval at the Robins Research Laboratories and to evaluate developments of computerized data bases, our program has been extended from the previous current awareness on several products[5] to retrospective searching for either a single chemical entity or a generic chemical. Several reports have been published during the past year comparing computer data bases and overlap of journal coverage. Data bases evaluated include *CA Condensates*,[1,3,5] *Chemical-Biological Activities* (CBAC),[1,3,5] *Polymer Science and Technology*,[1,3] *Biological Abstracts*,[1,3,6] *Bio-Research Index*,[1,3] *Institute for Science Information Source Tape*,[1,3] *Excerpta Medica*,[4,5] MEDLARS,[2,4,5] *Ringdoc*,[4,5] and *Engineering Index*.[3,6]

In the present comparison, several computer data bases were employed in retrospective searching for a specific chemical or a generic chemical class. The chemical selected for study in evaluating retrospective searches was *N*-ethyl-α-methyl-*m*-trifluoromethylphenethylamine (fenfluramine—a nonstimulating anorexigenic). The generic chemical class was 2-pyrrolidinones and excluded references to polymeric applications such as PVP, fused ring systems, and spiro-containing pyrrolidinones. Dioxo derivatives such as succinimides were also excluded from computer queries when possible (Figure 1).

The computer data bases chosen for this comparative evaluation included: (1) *Chemical Abstracts Condensates* (CAC), (2) *Chemical-Biological Activities* (CBAC), (3) *Ringdoc*, (4) *Excerpta Medica*, (5) MEDLARS, and (6) *Automatic New Structure alert* (ANSA). Coverage was limited primarily to the two-year period 1970–1971. The following two sections present results from chemical searching of these computer data bases.

2-PYRROLIDINONES

The generic chemical search was confined to derivatives of 2-pyrrolidinone where groups may be substituted in positions 1, 3, 4 or 5 of the pyrrolidine ring. No polymeric pyrrolidinone references were included and disclosures to diketo structures such as succinimides were excluded. The computer data bases selected for comparison were MEDLARS, *CA Condensates*, CBAC, and *Ringdoc*. Two approaches were utilized for querying the *Ringdoc* file and consisted of (1) Codeless Scanning and (2) Ring Code. Codeless Scanning uses keywords, index terms, and free terms which may be truncated; Boolean logic may be employed. Ring Code allows entry to the file by means of a chemical fragment code developed by the Documentation Ring and modified by Derwent in 1964. The heterocyclic 2-pyrrolidinone ring was coded as either substituted or unsubstituted to obtain complete coverage from the Ring Code file of *Ringdoc*.

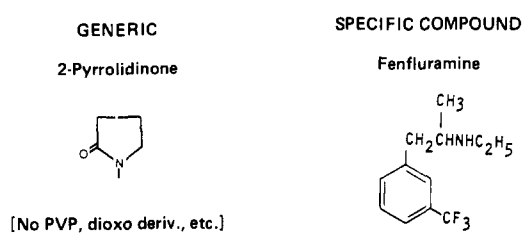

Figure 1. Computer search

The results obtained from the data bases for the two-year period 1970–71 are listed in Table I. The number of references cited and the number of hits from each data base are listed, along with the per cent of hits from each base relative to total number of significant references, the duplicate hits from each base, and, finally, the number of references which were listed by only one computer data base. Duplications were noted only when they occurred within either of the following two groups: (1) MEDLARS, *CA Condensates*, and CBAC and (2) *Ringdoc* Codeless Scanning and Ring Code (substituted and unsubstituted).

The principal results from Table I are the following:

1. Ringdoc provided the largest number of cited hits from a single base—i.e., 118 references.
2. Retrieval of generic chemicals by means of chemical fragments (i.e., Ring Code where 2-pyrrolidinone is either substituted or unsubstituted) was more effective than Codeless Scanning or the other data bases.
3. Ring Code provided the largest number of unique hits from a single data base (substituted plus unsubstituted)—i.e., 90 references.
4. Comparable numbers of unique hits were retrieved from each of three separate bases, MEDLARS, CAC, and CBAC (49, 53, and 32, respectively) with a small degree of overlap.

In Table II, the CAC and CBAC results were combined and compared with MEDLARS only. Also, the two Ring Code results (substituted and unsubstituted) were combined and compared only with the Codeless Scanning results. Data were also compiled from two other computer services, ANSA (new compounds) for 1970 and *Integrated Subject File* (Chemical Substance File only) for the last half of 1969. Results from manual searches of printed indexes to *Chemical Abstracts* for the last half of 1969 and *Chemical Substructure Index* (CSI, the index to *Current Abstracts in Chemistry* and *Index Chemicus*) for 1970 were also tabulated. The results from the aforementioned services are listed also in Table II, and the following principal conclusions were noted:

1. *Ringdoc* (Codeless Scanning and Ring Code) and Chemical Abstracts Services (CAC and CBAC) produced comparable numbers of hits and unique references.
2. MEDLARS contributed a significant number of references not listed in the other two services.

Table I. Comparison of Computer Bases—2-Pyrrolidinones

	Tot. No. Citations	No. Hits	Duplicates	Unique Hits
MEDLARS (1970-71)	217	66 (43%)	17	49
CAC (1970-71)	732	57 (38%)	4	53
CBAC (1970-71)	114	49 (32%)	17	32
(Total Diff. Refs.	*152)*			
Ringdoc (1970-71) – 118 separate references were retrieved				
Codeless Scanning (CS)	28	23 (19%)	16	7
Ringcode (unsubsti.)	90	12 (10%)	7	5
Ringcode (subst.)	287	100 (85%)	15	85

Table II. Summary 2-Pyrrolidinones

Data Base	1970 Only Total Refs.	1970-1971 Total Hits	1970-1971 Unique Hits.
MEDLARS		66	49
CAC - CBAC		103	99
Ringdoc – Codeless Scanning		23	7
Ringcode		104	90
ANSA (new cpds)	240		
ISF – Chemical Substance file only (no General Subject file; CA Vol. 71 only)	126 (only 6 months)		
Chemical Abstracts – Subject Indexes (Vol. 71)	109 (only 6 months)		
CSI (Curr. Abst. in Chem. - IC) Index	246		

3. ANSA and the CSI subject index gave comprehensive coverage of new compounds.

4. ISF (i.e., Chemical Substance File only) and CA subject index gave comparable coverage—i.e., 126 hits *vs.* 109.

In summary, generic chemical searching by means of computer data bases potentially offers a rapid method of obtaining reference material. Each data base covers a limited to fairly broad portion of the information spectrum; for example, MEDLARS is principally medically-pharmacologically oriented while *CA Condensates* is chemically and biologically oriented with practically no clinical data. Computerized indexing procedures vary from a restricted number of keywords to chemical fragments, to accessibility of entire abstracts (as with CBAC). The figures reported in these tables should not be considered absolute, but as indicative of certain advantages these bases offer when a generic search is required. From these results, it appears that no single computer base can be expected to give a yield which may be considered complete for the chemical industry and particularly for the pharmaceutical industry.

Results referred to in Table II concerning ANSA (Automatic New Structure Alert) for 2-pyrrolidinones are summarized in Table III. This computer data base is accessed by means of WLN (Wiswesser Line Notation). The two notations for pyrrolidinones utilized structures which contained substituents other than hydrogen on the nitrogen atom of the pyrrolidinone ring and those having a hydrogen in the 1-position. The WLN codes are T5NVTJ and T5MVTJ and both of these codes were used in searching the 1970 file. A total of 240 different derivatives containing the WLN fragments were retrieved from the computer file as compared to 246 from a manual search of CSI (Chemical Substructure Index).

FENFLURAMINE

In the following tables are listed the results from the retrospective searching of five computer data bases for literature citations to fenfluramine for the two-year period 1970–71. The five computer bases were (1) MEDLARS, (2) *Excerpta Medica*, (3) *CA Condensates* (CAC), (4) CBAC and, (5) *Ringdoc* (using both Ring Code and Codeless Scanning searching procedures). The total number of hits reported for each data base as well as unique references (found in only one source) are listed. In addition, the duplication of references between each data base is indicated.

The largest number of references was obtained from MEDLARS—i.e., 82 of 148 from all sources, or 55% of the different references reported by these data bases in 1970–71 (see Table IV).

Duplicates are shown in the following grid arrangement (Table V) which illustrates the degree of overlap between each two data bases. The largest overlap or duplication occurred between Codeless Scanning and Ring Code, followed by MEDLARS and Ring Code. There was also a fair degree of overlap between the following pairs: (1) Ring Code-*Excerpta Medica;* (2) *Excerpta Medica*-MEDLARS; (3) CAC-CBAC and (4) Codeless Scanning-MEDLARS.

The number of citations from Chemical Abstracts Services (CAC and CBAC) have been combined, as have those from Ringdoc (Ring Code and Codeless Scanning). The results have been tabulated in Tables VI and VII to allow for ease of comparison. The yields from each data base were compared, and the total number of individual references to fenfluramine from all these sources was 148. (Subsequent to the presentation of this paper, a rerun of the *Excerpta Medica* base (1970–1971) search for fenfluramine references yielded approximately 90 references instead of the 60 as reported. The number of distinct ref-

Table III. ANSA—Automatic New Structure Alert

[13,000-17,000 abstracts/month;
12,000-17,000 new compounds/month;
Chemical Substructure Index (CSI) of Current
Abstracts in Chemistry (CAC)
and Index Chemicus (IC)]

2-Pyrrolidinones

T5NVTJ or T5MVTJ

T = 1st Symbol Het. ring
5 = Size of ring
N = Nitrogen atom (no H)
V = Carbonyl group next to nitrogen
T = Saturated ring
J = End of ring notation

M = imino –NH gp.

RESULTS – 240 hits

Manual Search

Pyrrolidinones	T5NVTJ	147
	T5MVTJ	99
		246

Table IV. Fenfluramine 1970–1971

Data Bases	Total No. Refs.	% Refs. per Base	No. Unique Refs./Base
MEDLARS (MED)	82	55	31
Excerpta Medica (EM)	60	40	22
CA Condensates (CAC)	45	30	13
CBAC	31	22	3
Ringdoc			
Codeless Scanning (CS)	43	29	2
Ringcode (RC)	57	38	3
(Total Diff. Refs.	*148)*		

Table V. Fenfluramine 1970–1971

Data Base	Total No. Refs.	Duplication				
MEDLARS (MED)	82					
Excerpta Medica (EM)	60	28				
CAC	45	20	12			
CBAC	31	17	17	28		
Ringdoc Codeless Scanning (CS)	43	25	23	15	14	
Ringdoc Ringcode (RC)	57	33	30	24	20	40
(Total Diff. Refs.	*148)*	MED	EM	CAC	CBAC	CS

Data Bases

Table VI. Fenfluramine 1970–1971

Data Base	Total No. Refs.	% Refs. per Base	No. Unique Refs./Base
MEDLARS	82	55	31
Excerpta Medica	60	40	22
CAC - CBAC	49	33	16
Ringdoc	60	40	10
(Total Diff. Refs.	*148)*		

Table VII. Fenfluramine 1970–1971

Data Base	Total No. Refs.	Duplication*		
MEDLARS	82			
Excerpta Medica	60	28		
CAC – CBAC	49	19	22	
Ringdoc	60	33	39	25
		MED	EM	CAC - CBAC
		Data Bases		

*Combined totals for CA Services and Ringdoc searching.

erences to be added to the 148 total and duplication of citations between the computer bases is available on request from the author.)

From the results listed in Tables VI and VII, the following general conclusions are apparent:

1. MEDLARS, citing 55% of the different references, provided the most complete coverage. *Excerpta Medica*, Chemical Abstracts Services and *Ringdoc* produced 40, 33, and 40%, respectively.

2. MEDLARS and *Excerpta Medica* gave the highest number of unique references.

3. A significant number of references was retrieved by only one data base indicating a need for searching multiple bases to assure greater coverage.

4. Highest duplication occurred between *Ringdoc* and *Excerpta Medica*, followed by *Ringdoc* and MEDLARS, *Excerpta Medica* and MEDLARS, and the combination CAC–CBAC with either *Ringdoc* or *Excerpta Medica*.

The WLN code for fenfluramine was utilized in an ANSA search. As expected, no citations were listed from the 1970 file because fenfluramine is not a new compound.

SUMMARY AND CONCLUSIONS

The present study summarizes the availability and potential use of several computer data bases for the purpose of retrospective searching for either specific chemical structures or generic chemical structures. Conclusions are tabulated below:

1. Computer bases such as *CA Condensates*, CBAC, ANSA, and *Excerpta Medica* permit retrospective searching for a single compound as well as generic searching; ISF searching will be most valuable as the files expand and may well become a preferred source.

2. Specific compound searching in MEDLARS is facilitated when a Medical Subject Heading (MeSH) term has been designated; generic searching presents some difficulty owing to broad coverage of generic headings.

3. Use of WLN code provides access to data bases for either specific or generic chemical.

4. Chemical fragment and Codeless Scanning procedures employed by *Ringdoc* allow for both specific and generic coverage, but a single program combining both methods will prove to yield greater coverage in a single search. Such a program is being sponsored jointly by Searle and Pfizer (SPRING).

ACKNOWLEDGMENT

The authors wish to express their sincere appreciation for the cooperation of: Computer Center, University of Georgia for *CA Condensates*, CBAC, and ISF searches (Margaret Park and Margaret Caughman), 3i Computer Company for *Excerpta Medica* searches (Arthur Haines), MEDLARS (PMA Science Information staff under the direction of Charles Cleveland) and Institute for Scientific Information for ANSA searches (Charles Granito and Murray Rosenberg). In addition, the able assistance of Virginia Kelley was most valuable in compiling data and preparation of the manuscript.

LITERATURE CITED

(1) Garfield, E., Granito, C., and Petrarca, A. E., pp. 510–535, in "Kirk-Othmer Encyclopedia of Chemical Technology," 2nd ed., Suppl. Vol., Wiley, New York, 1971.

(2) Goode, D. J., Penry, J. K., and Caponio, J. F., "Comparative Analysis of Epilepsy Abstracts and a MEDLARS Bibliography," *Bull. Med. Libr. Assoc.* **58**, 44–50 (1970).

(3) Schipma, P. B., Williams, M. E., and Shafton, A. L., "Comparison of Document Data Bases," *J. Amer. Soc. Inform. Sci.* **22**, 326–32 (1971).

(4) Scott, E. J., Townley, H. M., and Stern, B. T., "A Technique for the Evaluation of a Commercial Information Service and Some Preliminary Results from the Drugdoc Service of the Excerpta Medica Foundation," *Inform. Stor. Retr.* **7**, 149–65 (1971).

(5) Smith, D. R., Beauchamp, R. O., Jr., Garber, J. L., and Daugherty, M. A., "Computerized Drug Information Services," *J. Chem. Doc.* **12**, 9–13 (1972).

(6) Wood, J. L., Flanagan, C., and Kennedy, H. E., "Overlap in the Lists of Journals Monitored by Biosis, CAS and Ei," *J. Amer. Soc. Inform. Sci.* **23**, 36–8 (1972).