

An Indexing Coverage Study of Toxicological Literature*

RUTH REINKE MONTGOMERY

E. I. du Pont de Nemours and Co., Haskell Laboratory for Toxicology and Industrial Medicine,
Newark, Del. 19711

Received September 26, 1972

A data base of 1873 citations dated 1960–1969 was obtained through an author survey of the members of the Society of Toxicology. Coverage was determined by checking the author index through a maximum of three years after publication or through the end of the decade. *Chemical Abstracts* consistently provided coverage of more than two-thirds of the 1960–1967 citations in the data base. *Chemical Abstracts*, *Biological Abstracts*, *Index Medicus*, and *Science Citation Index* each provided coverage of 79% to 85% of the 215 citations dated 1966 and their combined coverage exceeded 98%. *Excerpta Medica* (Section IIC) and *Chemical-Biological Activities* provided 43% and 58% coverage, respectively, of this same group of 215 citations.

PROCEDURE

The goal of this study was an appraisal of indexing coverage for the interdisciplinary field of toxicology. Toxicology was approached as that branch of science dealing with the harmful effects of chemicals on biologic mechanisms.¹ Information of research significance was limited to primary literature, defined as published documents containing original data from experimental work or case studies.

The study of coverage was restricted to that offered by six major secondary sources: *Chemical Abstracts* (CA); *Biological Abstracts* (BA); *Index Medicus* (IM); *Excerpta Medica* (EM), Section IIC in the listing then current; *Chemical-Biological Activities* (CBAC), printed version; and *Science Citation Index* (SCI). For an abstracting and/or indexing service, the most fundamental performance criterion is coverage of available information. The study of coverage provides a basis for comparative evaluation which is entirely dependent on the service or system, as contrasted to other major performance criteria such as recall, precision, and user effort, which can be significantly affected by user training and experience.

The technique used was a modification of the author survey approach of Orr² and Orr and Crouse³ and also of Garvey and Griffith.⁴ For this study, the survey was based on professional society membership rather than presentation of a paper at a single meeting, in order to obtain a larger and more representative sample suitable for studying the still diffuse field of toxicology. In checking the six services selected for study, the author index provided the only common approach. Additionally, the author index is generally regarded as a relatively certain means of checking for coverage. Only six-month, yearly, or collective author indices were used for checking coverage to expedite searching.

The Society of Toxicology, founded in 1961, is a professional society oriented to toxicology. Contact was made with the Secretary and the President of the Society to explain the proposed canvass of membership. Each member listed in the 1970 directory was then sent a letter requesting a personal bibliography, 1960 through 1969, marked to show original data of toxicologic importance. The letter was enclosed with a reply form and a postage-paid return envelope. A month after the individual letters were sent out, a letter intended for duplication in the next issue of the Society newsletter was sent to the Secretary of the Society. This letter expressed gratitude to the members who had already replied, answered several questions, and served as a reminder to those who had not yet answered.

The raw data of heterogeneous citations from the bibliographies were passed through three filters: first, time as indicated by a date of 1960 through 1969; second, pertinence to toxicological research as judged by an author listed in each respective citation; third, suitability for checking in secondary sources. Citations were not considered suitable for checking in secondary sources if the complete document did not appear to be available through a library or other routine means. For example, papers presented at most meetings, government-classified or government-limited reports, and privately-printed documents were not considered available. Although abstracts published separately were generally excluded from the data base, citations found in a service as references from *Dissertation Abstracts* were included as a copy of the complete dissertation can be obtained through a routine procedure. In one case, the number of citations was reduced by sampling, from 164 citations for the 1960–1969 span to 46 citations, to equal the largest number of publications contributed by any of the other members to the data base.

Citations were checked for correspondence by page reference. Occasional obvious errors, either by the member or by the service, such as incorrect spelling, slight variance in title, omission of an author, or transposition or truncation of page numbers, were not considered reasons for rejection. Government documents were considered to correspond on the basis of title-author match. Book

*Presented at the 7th Middle Atlantic Regional Meeting, ACS, Philadelphia, Pa., Feb. 16, 1972. Revised and excerpted from the author's thesis, Drexel University, Graduate School of Library Science, June 1971, Master of Science (Information Science). Thesis M-2951, University Microfilms, Ann Arbor, Mich.

chapters were matched on the basis of an entry for either the monograph or the chapter cited under an author listed in the book chapter.

All citations in the data base were checked in the collective or volume indices of CA which were issued for the 1960–1969 decade, either until a citation was found, or through three years after publication up to the 1969 time limit. Citations dated 1966 through 1969 were similarly checked in the annual or semi-annual author indices which were issued for these four years by BA, IM, EM, CBAC, and SCI. Citations dated 1966 which were not found in any of the six services were additionally checked in the publication listed as the primary source. Yearly totals were then prepared and analyzed. The 5% level of significance was used.

RESULTS

The data base consisted of 1873 citations submitted by 221 members (41% of the membership). Twenty-four of these 221 members were resident in countries other than the United States: eight in Canada; two each in Belgium, England, France, Germany, and Japan; one each in Czechoslovakia, Denmark, India, The Netherlands, New Zealand, and Switzerland.

Coverage within three years of the publication date was determined on 1233 citations dated 1960 through 1966 as shown in Table I. Differences in the percent of citations indexed by CA from 1960 through 1966 were significant ($0.05 > P > 0.02$), but the data indicate that the differences were significant because of the marked increase in 1966 to 84.2%. Differences among the percentages prior to 1966 were small and not significant at the 0.05 level.

Differences in the rates of indexing among the six services for the 215 citations of 1966 as illustrated in Figure 1 were highly significant ($P < 0.001$). Excluding EM, differences among the remaining five services were significant ($0.005 > P > 0.001$). The services with the two lowest rates of indexing, EM and CBAC, were selective by policy. CBAC limited coverage to organic compounds and EM operates as a classified service—Section IIC was one of many subject sections for the broad area of medicine. Also, both EM and CBAC screened a smaller number of journals than the other four services. In relation to size and scope, the amount of coverage offered by EM and CBAC compared favorably to that offered by the other four services. (The size and scope of CBAC changed in January 1972 when the printed CBAC was merged with the biochemical section of CA.)

Table I. Proportions of Coverage Within Three Years

Year of Publ.	Index	Number of Citations			% Indexed
		Total	Indexed	Not Indexed	
1960	CA	137	105	32	76.6
1961	CA	141	106	35	75.2
1962	CA	140	97	43	69.3
1963	CA	199	152	47	76.4
1964	CA	199	148	51	74.4
1965	CA	202	146	56	72.3
1966	CA	215	181	34	84.2
1960–1966	CA	1233	935	298	75.8
1966	BA	215	169	46	78.6
1966	IM	215	182	33	84.7
1966	EM	215	92	123	42.8
1966	CBAC	215	124	91	57.7
1966	SCI	215	182	33	84.7

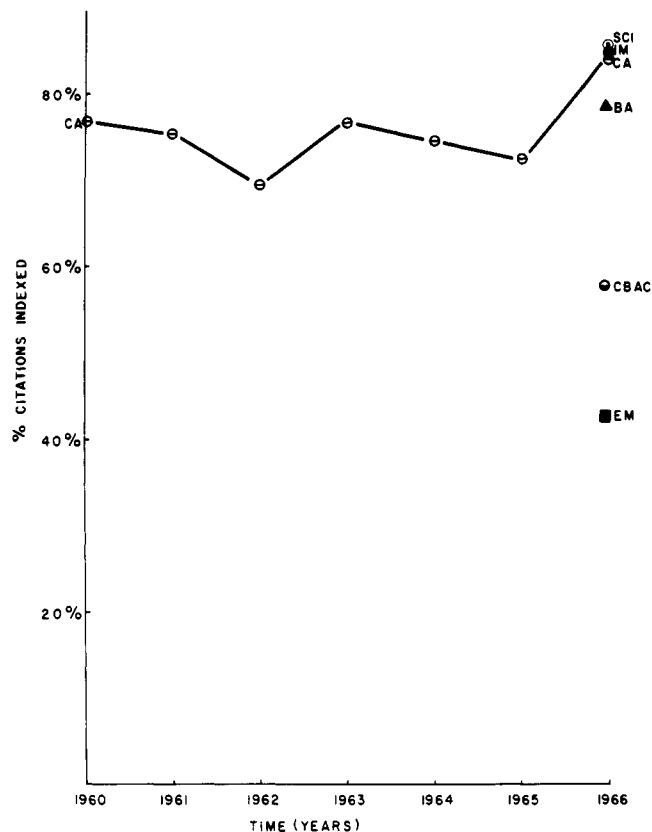


Figure 1. Proportions of coverage for citations dated 1960–1966

Differences in the rates of indexing among the four services CA, BA, IM, and SCI for these 215 citations were not significant. These four larger services all offer comprehensive coverage in their respective fields.

The results of checking the 855 citations dated 1966–1969 at yearly intervals after publication are shown in Table II. These data are also illustrated in Figure 2. In general, differences in the percent of citations indexed by CA, BA, IM, and SCI were not substantial at two or more years after publication but were substantial at less than two years. Specifically, at the end of 1969, differences between these four services for citations dated 1967 were not significant but the differences for citations dated 1968 were significant ($0.005 > P > 0.001$). A major and consistent difference in the four-year period 1966 through 1969 was the high percentage of early indexing by SCI. SCI releases its annual cumulation only after indexing all issues in the SCI coverage which were issued during the year being cumulated. In retrospective searching, the initial delay in release time results in the appearance of unusually early coverage.

In this study, delayed indexing tended to be more frequent with BA. In partial defense, one 1966 citation abstracted by BA at three years was from a Brazilian journal not indexed in any of the other five services. Delayed indexing in this case represented a contribution not available elsewhere.

Coverage is directly related to overlap—how much of the material is covered in more than one source (or only one source, or not at all). The data on overlap for the 215 citations dated 1966 which were checked through three years in each of the six services are presented in Table III. More than 30% of these citations were found in all six services. Only four citations—less than 2%—were not found in any service. As might be expected, no citations were

AN INDEXING COVERAGE STUDY OF TOXICOLOGICAL LITERATURE

Table II. Proportions of Coverage at Yearly Intervals after Publication

Year of Publication	Index	Number of Citations Indexed Each Year				Cumulative Percentage of Citations Indexed			
		Same	1 yr	2 yr	3 yr	Same	1 yr	2 yr	3 yr
1966 (N = 215)	CA	134	45	2	0	62.3	83.3	84.2	84.2
	BA	53	99	15	2	24.7	70.7	77.7	78.6
	IM	56	124	2	0	26.0	83.7	84.7	84.7
	EM	33	59	0	0	15.3	42.8	42.8	42.8
	CBAC	95	29	0	0	44.2	57.7	57.7	57.7
	SCI	177	5	0	0	82.3	84.7	84.7	84.7
1967 (N = 212)	CA	106	42	1		50.0	69.8	70.3	
	BA	54	81	15		25.5	63.7	70.8	
	IM	89	74	1		42.0	76.9	77.4	
	EM	31	40	1		14.6	33.5	34.0	
	CBAC	68	27	0		32.1	44.8	44.8	
	SCI	141	15	0		66.5	73.6	73.6	
1968 (N = 218)	CA	112	32			51.4	66.1		
	BA	53	84			24.3	62.8		
	IM	116	67			53.2	83.9		
	EM	14	48			6.4	28.4		
	CBAC	81	21			37.2	46.8		
	SCI	177	12			81.2	86.7		
1969 (N = 210)	CA	103				49.0			
	BA	52				24.8			
	IM	100				47.6			
	EM	6				2.9			
	CBAC	72				34.3			
	SCI	165				78.6			

N = Number of citations checked.

Same, 1 yr, 2 yr, 3 yr = Same as year of publication and one, two, and three years after the year of publication.

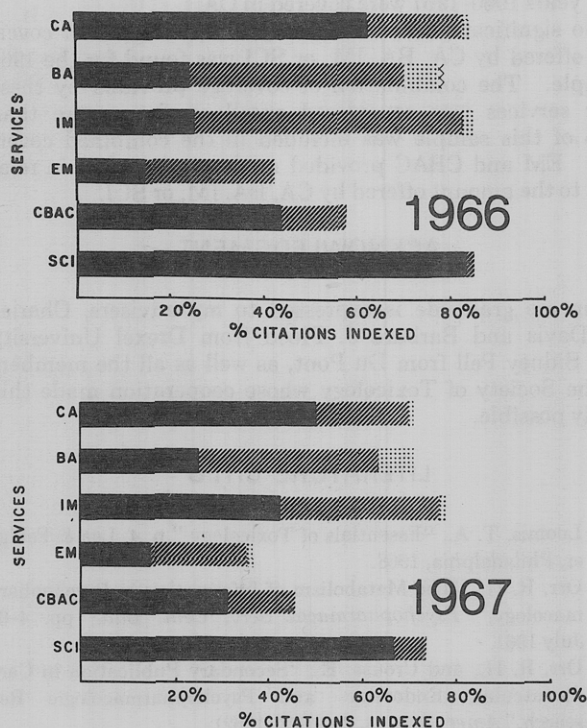


Figure 2. Coverage for citations dated as indicated in 1966-1969

Time of Indexing
 ■ Year of publication
 ▨ 1 year after publication
 ▩ 2 years after publication
 ▤ 3 years after publication

Table III. Overlap in Indexing for 215 Citations Dated 1966

Number of Citations	Number of Services Providing Coverage	Indexes Providing Coverage					
		CA	BA	IM	EM	CBAC	SCI
66	6	x	x	x	x	x	x
39	5	x	x	x		x	x
9	5	x		x	x	x	x
6	5	x	x	x	x		x
3	5	x	x		x	x	x
1	5	x	x	x	x	x	
15	4	x	x	x			x
3	4	x		x		x	x
3	4	x		x	x		x
3	4		x	x	x		x
2	4	x	x			x	x
2	4	x	x	x		x	
1	4	x	x		x		x
1	4	x		x	x	x	
13	3		x	x			x
9	3	x		x			x
6	3	x	x				x
3	3	x	x	x			
2	3	x				x	x
3	2	x	x				
3	2	x		x			
2	2		x	x			
1	2	x					x
1	2		x				x
1	2			x	x		
1	2			x			x
7	1	x					
2	1		x				
2	1			x			
1	1						x
4 ^a	0						

^a Checked in the original publication.

found only in EM or only in CBAC, and coverage in CBAC was always accompanied by coverage in CA. Coverage in EM was found with different overlap combinations but was most common with coverage in all six services.

DISCUSSION

The value of the data base as a sample for toxicology literature depends on the boundaries set for the field of toxicology. One major publication using a broad definition of toxicology, *Toxicity Bibliography*, listed over 27,000 citations for its initial two years of operation during 1968-1969. Comparison with this source suggests that the data base provided a yearly sample in the range of 1%, emphasized research testing rather than case reports, and concentrated on English-language publications, particularly those of United States origin.

Over-all, published toxicity reports appear to be generally well covered in secondary sources. CA alone consistently provided coverage of more than two-thirds of the data base. The combination of the four services CA, BA, IM, and SCI offered coverage of more than 98% of the 1966 data base. By way of comparison, Martyn⁵ found that "in general, about a fifth of such literature as a bibliographer deems worthy of mention is not covered at all

by abstracting and indexing services ... although there are considerable variations between disciplines ...". A major factor in the particularly high coverage found in this study lies in the restriction of the data base to complete published documents. This restriction was established to sift the raw data from personal bibliographies. Basically, however, the restriction is realistic: services, especially subject services, can be expected to provide comprehensive coverage only of complete accessible material. Users find indexes and abstracts of limited value if the original documents are unavailable.

The general factors governing the effectiveness of a published index have been identified.⁶ Faced with specific practical questions in toxicology, users may ponder the probable benefits against the anticipated cost (time, effort, money) associated with search of secondary sources. The six services checked for coverage in this study differ appreciably in their approach: chemical (CA, CBAC), biological (BA), medical (IM, EM), or science in general (SCI). The type and depth of abstracting and/or indexing reflect the primary subject approach as well as size, scope, and individual characteristics. At present, three services (CA, BA, EM) provide a publication with abstracts while two (IM, SCI) offer the convenience of direct reference to citations. Analysis of the specific query in relation to the distinctive features of available services will promote optimum cost-effectiveness. Coverage is a basic criterion in this analysis.

CONCLUSIONS

The coverage of toxicological literature provided by CA alone was evaluated as consistently good. More than two-thirds of the literature samples for the eight consecutive years 1960-1967 were covered in CA.

No significant difference between the amount of coverage offered by CA, BA, IM, or SCI was found for the 1966 sample. The combination of coverage provided by these four services was considered excellent since more than 98% of this sample was included in the combined coverage. EM and CBAC provided moderate coverage in relation to the amount offered by CA, BA, IM, or SCI.

ACKNOWLEDGMENT

Sincere gratitude is expressed to my advisers, Charles H. Davis and Barbara J. Flood from Drexel University and Sidney Pell from Du Pont, as well as all the members of the Society of Toxicology whose cooperation made this study possible.

LITERATURE CITED

- (1) Loomis, T. A., "Essentials of Toxicology," p. 4, Lea & Febiger, Philadelphia, 1968.
- (2) Orr, R. H., "The Metabolism of Information in Psychopharmacology," *Psychopharmacol. Serv. Cent. Bull.*, pp. 4-6, July 1961.
- (3) Orr, R. H., and Crouse, E., "Secondary Publication in Cardiovascular, Endocrine, and Psychopharmacologic Research," *Amer. Doc.* 13, 297-303 (1962).
- (4) Garvey, W. D., and Griffith, B. C., "Studies of Social Innovations in Scientific Communication in Psychology," *Amer. Psychol.* 21, 1019-36 (1969).
- (5) Martyn, J., "Tests on Abstracts Journals: Coverage Overlap and Indexing," *J. Doc.* 23, 45-70 (1967).
- (6) Lancaster, F. W., "The Cost-Effectiveness Analysis of Information Retrieval and Dissemination Systems," *J. Amer. Soc. Inform. Sci.* 22, 12-27 (1971).