The Publication Speed of Information in Bibliographic Chemical Databases

I. Dióspatonyi,[†] G. Horvai,[†] and T. Braun*[,‡]

Division of Chemical Information Technology, Budapest University of Technology and Economics, 1111 Budapest, Gellért tér 4, Institute of Inorganic and Analytical Chemistry, L. Eötvös University, POB 123, 1443 Budapest, Hungary, and Information Science and Scientiometric Research Unit (ISSRU), Budapest, Hungary

Received April 15, 2001

The quality of bibliographic databases depends very much on the reliably fast follow-up of the pertinent literature. We have studied this quality feature of two important chemical bibliographic databases: *Chemical Abstracts* (*CA*) and *Analytical Abstracts* (*AA*). The follow-up rate (speed of information) of these databases has been determined with respect to 10 core journals in analytical chemistry. On the average the performance of *Chemical Abstracts* has been good over the past decade, while the formerly poor performance of *Analytical Abstracts* has improved considerably. Some quite unexpected nonuniformities can, however, still be detected in the publication time distributions of both databases.

1. INTRODUCTION

Scientific information sources such as journals and bibliographic databases need to maintain high quality standards. One important quality feature is how fast new information is being diffused. This speed may be characterized by the time elapsing between access to the information by the journal and access to the information by the readers in secondary sources. In a previous paper[1] we studied 10 core journals in analytical chemistry with respect to their publication speed. The present article is devoted to the study of the speed of two important chemical bibliographic databases: *Chemical Abstracts* and *Analytical Abstracts*. The follow-up of the performance of these databases for more than a decade gives insights which go in some respect against the expectations of database users. Although the average performance of both databases has been quite satisfactory in the past 5 years, the distribution of the time lapse over different journals, over different years, or over different papers in the same journal and the same year remains still too wide. This is an area where the quality of both databases needs to be improved.

The present work is an extension and follow-up of preliminary studies[2−4] which indicated the need of a more thorough examination.

2. METHODS

Ten analytical chemistry core journals have been selected (Table 1). The speed of *Chemical Abstracts* and *Analytical Abstracts*, respectively, in preparing abstracts of papers and making them available to users has been studied paper by paper for selected years (Tables 1 and 2) over more than a decade. The CD-ROM versions of both databases have been used as data sources. The time lapse between the earliest possible access to the printed paper and the earliest possible access to the abstract in the printed version of the database was selected as the measure of lapse time. We took as the first of these two time limits the date of publication of the printed issue of the analytical journal (as stated on the respective front page). We used the so-called "update" field of each database as the cross-reference between the abstract and the respective journal issue of the original paper. The second time limit was defined as the date of publication of the printed issue of the respective database containing the particular abstract. The date of publication was considered to agree with the date on the front page of the issue containing the abstract.

All abstracts of papers of the 10 selected analytical journals in the selected years were looked up in the respective database. The time lapse of access to the abstract has been determined in every case, using the definition of the previous paragraph. This time lapse was generally different for different articles, and therefore it was reasonable to study the distribution of the time lapses of individual abstracts. This has been done journal by journal and year by year. The results are shown as distribution curves (cumulative percentage of abstracts as a function of their lapse time) and as tables of statistical summary figures. The latter are as follows: (1) range, the shortest and the longest lapse time among the abstracts for all abstracted papers of a particular journal in a given year; (2) interquartile range, lapse time values on the distribution curve belonging to the first 25% and 75%, respectively, of the abstracts for one journal in the same year; and (3) median, the 50% point on the distribution curve.

Note that at the bottom of each column in the tables the statistical median of the values shown in the column is also presented.

For technical reasons the definition of "one year" is not exactly the same in all cases. For studying *Analytical Abstracts* we could use the CD 2000/12 which contained all abstracts since 1980. Thus the database certainly included

* Corresponding author phone: (36-1)-3115-433; e-mail: h1533bra@ella.hu.

[†] Budapest University of Technology and Economics.

[‡] L. Eötvös University and Information Science and Scientiometric Research Unit.

10.1021/ci010032l CCC: $20.00
Published on Web 10/24/2001

Table 1. Lapse Time of Abstracts (from Date of Publication in the Journal to Publication in the *AA* Database)

	publication lapse of abstracts [days]					
	1985			1990		
journal	median	interquartile range	range	median	interquartile range	range
Analyst	240	240−270	210−270	270	240−300	210−420
Anal. Chem.	240	240−270	90−300	346	316−406	166−466
Anal. Chim. Acta	300	271−300	150−330	330	300−338	240−479
Anal. Letters				390	360−450	120−480
Chromatographia	240	240−270	90−360	390	330−420	300−540
Fresenius J. Anal. Chem.	270	256−316	226−360	406	330−436	286−496
J. Chromatogr. A	253	244−282	131−482	331	295−354	209−405
J. Chromatogr. B	241	230−258	214−293	337	295−337	195−421
J. Chromatogr. Sci.	270	240−270	240−300	420	360−420	180−450
Talanta	210	210−240	210−270	330	300−360	240−420
median	241	240−270	210−300	342	308−383	210−458

	publication lapse of abstracts [days]					
	1995			1999		
journal	median	interquartile range	range	median	interquartile range	range
Analyst	150	120−150	120−180	150	120−150	101−150
Anal. Chem.	150	150−166	136−300	150	134−173	94−210
Anal. Chim. Acta	132	126−142	101−171	130	107−142	95−170
Anal. Letters	106	90−120	90−150	187	157−277	93−358
Chromatographia	90	90−120	90−120	150	127−150	115−180
Fresenius J. Anal. Chem	120	106−136	76−210	150	127−150	94−240
J. Chromatogr. A	147	129−160	45−178	131	121−144	86−179
J. Chromatogr. B	148	133−166	76−178	132	104−142	85−170
J. Chromatogr. Sci.	120	90−150	60−180	314	275−397	185−480
Talanta	120	120−150	120−180	150	128−150	96−180
median	126	120−150	90−179	150	127−150	95−180

Table 2. Lapse Time of Abstracts (from Date of Publication in the Journal to Publication in the *CA* Database)

	publication lapse of abstracts [days]								
	1990			1995			2000		
journal	median	interquartile range	range	median	interquartile range	range	median	interquartile range	range
Analyst	72	59−92	39−485	70	56−102	43−213	91	60−106	30−175
Anal. Chem.	8	1−17	−19−265	10	−4−35	−17−236	39	15−58	−11−148
Anal. Chim. Acta	99	83−117	59−464	51	38−72	20−184	68	34−83	−11−214
Anal. Letters	157	135−173	76−409	56	42−90	17−229	83	64−106	25−156
Chromatographia	134	118−157	84−256	116	93−151	58−248	120	97−132	54−272
Fresenius J. Anal. Chem	108	88−129	62−215	97	73−123	51−229	76	56−106	20−189
J. Chromatogr. A	86	67−106	44−314	64	51−95	30−280	86	59−99	16−168
J. Chromatogr. B	73	64−92	44−454	53	44−64	37−264	101	92−118	37−216
J. Chromatogr. Sci.	87	71−117	50−233	60	31−94	17−165	85	53−93	27−143
Talanta	89	76−112	53−188	115	88−144	42−252	56	34−83	7−199
median	88	76−117	53−314	62	51−95	37−229	84	59−106	25−189

all abstracts of papers that were published in 1985, 1990, or 1995. As for 1999, some very slow abstracts may still be missing.

In the case of *Chemical Abstracts* we used the 12th and 13th Collective index on CD ROM and the abstracts CD ROM 2000/133(17). The Collective Indexes ended with December 1991 and 1996, respectively. This could cause a slight truncation of the distributions. For the year 2000, we determined the lapse times for the abstracts from January to October 2000, instead of the papers from 2000. We estimate that none of these distortions is considerable.

3. RESULTS

3.1. Analytical Abstracts (AA). The performance of *AA* was studied in four years: 1985, 1990, 1995, and 1999, respectively. The 2000/03 CD-ROM version of *AA* was used as the data source. The results are shown in Table 1 and Figures 1−10.

What immediately strikes the eye when looking at these data is the very poor performance in 1990. The rate of following the publication of original papers was much slower in this year than either in 1985 or in 1995 and 1999. For example half of the abstracts relating to *Journal of Chromatographic Science* appeared more than 420 days after the publication of the original paper.

The time lapse of follow-up was not only extremely long in 1990 but its spread was also rather wide. Papers of *Analytical Chemistry* needed between 166 and 466 days for the abstract to be published. Thus the reader may have sometimes encountered abstracts relating to papers that were 6 months old and sometimes such relating to 16 month old papers. After 1990 changes in editorial practices have lead to great improvements in the *AA* database. By 1995 the

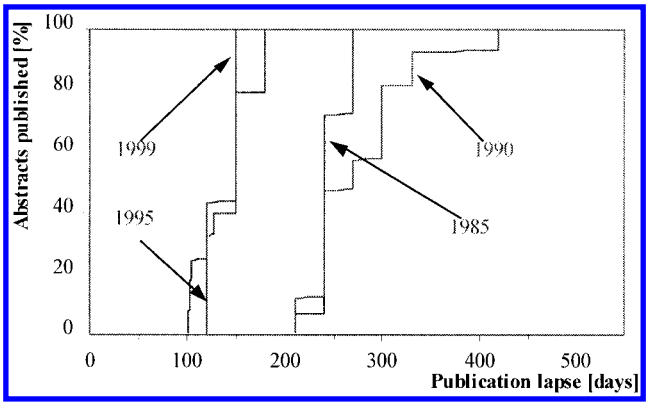

Figure 1. Publication lapse of abstracts from date of publication in *Analyst* to publication in the *AA* database.

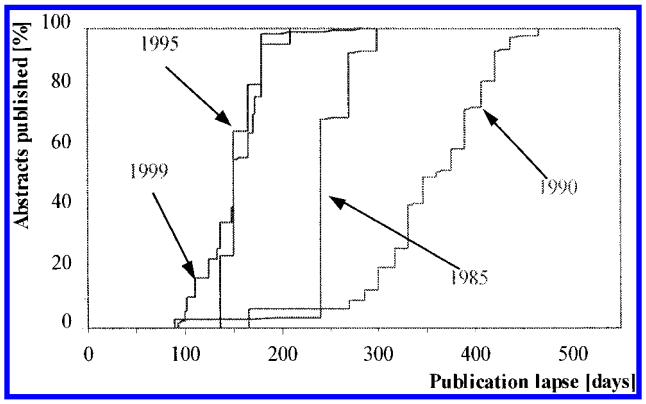

Figure 2. Publication lapse of abstracts from date of publication in *Analytical Chemistry* to publication in the *AA* database.

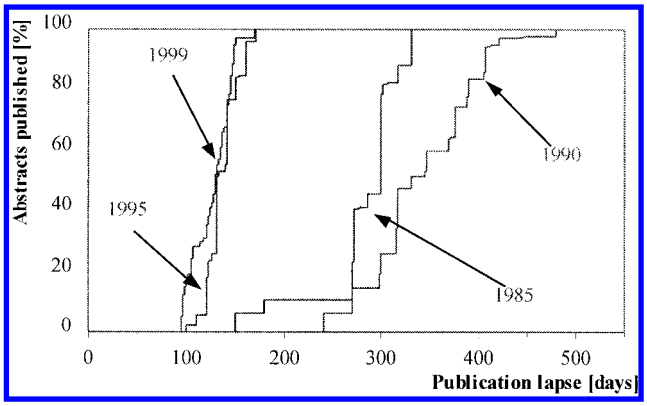

Figure 3. Publication lapse of abstracts from date of publication in *Analytica Chimica Acta* to publication in the *AA* database.

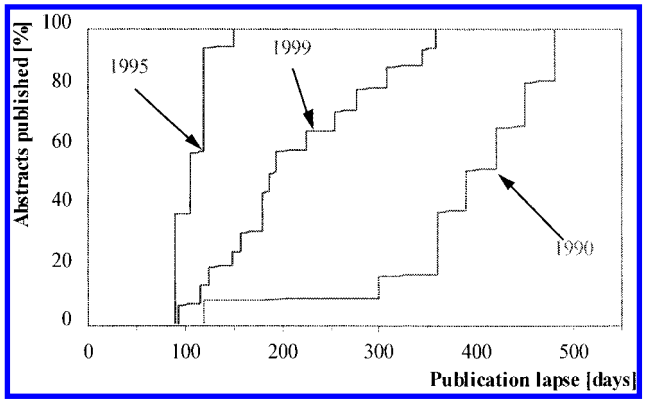

Figure 4. Publication lapse of abstracts from date of publication in *Analytical Letters* to publication in the *AA* database.

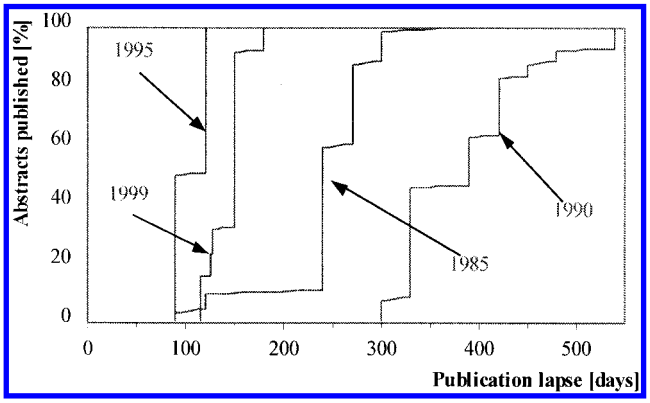

Figure 5. Publication lapse of abstracts from date of publication in *Chromatographia* to publication in the *AA* database.

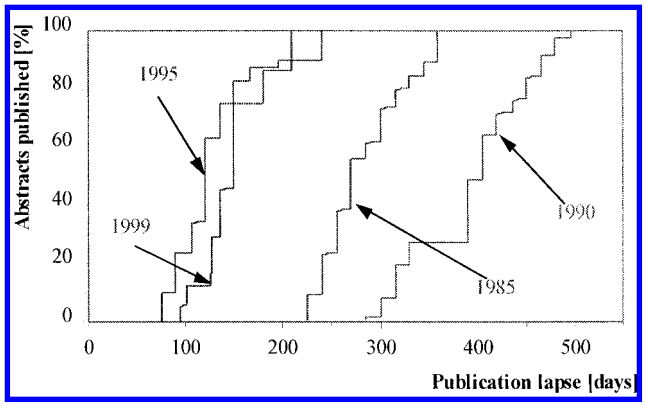

Figure 6. Publication lapse of abstracts from date of publication in *Fresenius Journal of Analytical Chemistry* to publication in the *AA* database.

median lapse for all 10 journals went down to 5 months or less. The spread of lapse times became also narrower and even the slowest abstracts appeared generally within 6 months.

Surprisingly, the 1999 performance of *AA* was significantly worse than that of 1995. For example the median lapse time for *Analytical Letters* and for *Chromatographia* increased by more than two-thirds. Ranges have also deteriorated. Looking again at *Analytical Letters* a 4-fold increase of the spread of lapse times (from 60 days to 265 days) is observed between 1995 and 1999. Stagnation or slight improvements with other journals hardly counterbalance the notable performance losses.

The most striking observation was to note the complete lack of abstracts for the 1999 papers of *Journal of Chromatographic Science* in the 2000/03 CD-ROM edition of *AA*. It has apparently escaped the quality assurance of *AA* that a full year of one of the core analytical journals that *AA* itself chose for abstracting "cover to cover",[6] was missing. This serious flaw was noted later by *AA* and the data shown in Table 1 and Figure 9 come from the 2000/09 CD-ROM.

A final note should be made about data for *Analytical Letters* in 1985. These are missing from our data set because the publication date was not indicated on the printed issues of this journal in 1985.

3.2. Chemical Abstracts (CA). The performance of *CA* was studied in three years: 1990, 1995, and 2000. This choice was dictated by the availability of CD-ROM versions of *CA* in the libraries accessible to us. We could use Collective Indexes 12 (1987−1991) and 13 (1992−1996)

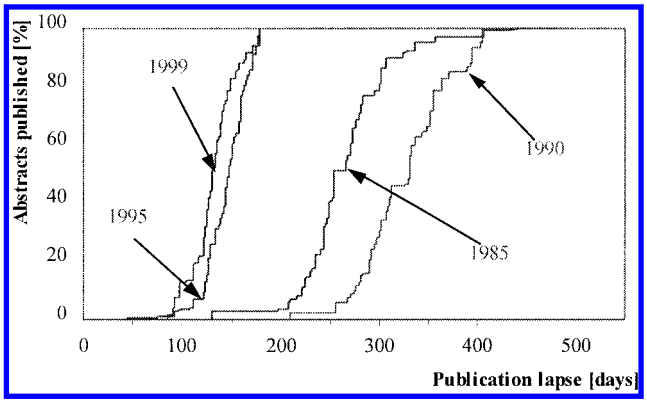

Figure 7. Publication lapse of abstracts from date of publication in *Journal of Chromatography A* to publication in the *AA* database.

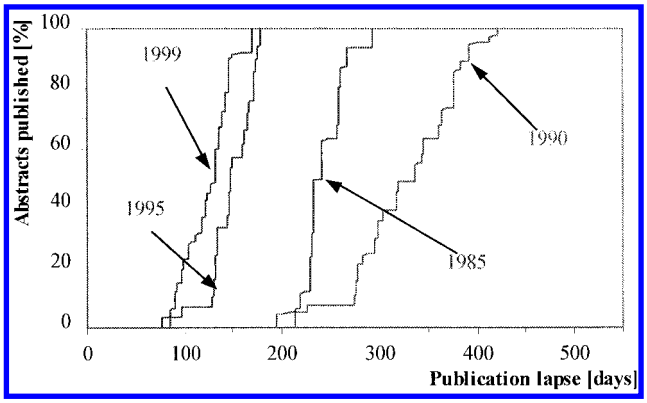

Figure 8. Publication lapse of abstracts from date of publication in *Journal of Chromatography B* to publication in the *AA* database.

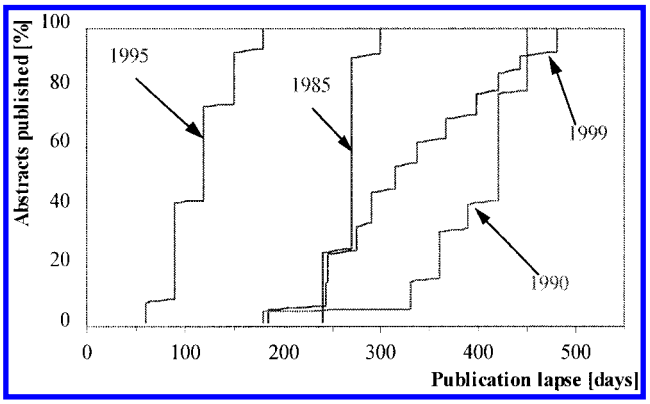

Figure 9. Publication lapse of abstracts from date of publication in *Journal of Chromatographic Science* to publication in the *AA* database.

and the *CA* CD-ROM 2000/133(17) (which was the latest CD-ROM when this study was made).

The results are shown in Table 2 and Figures 11−20. As the scales of Figures 11−20 are the same as those of Figures 1−10, one can immediately see that the abstracts of analytical papers appear in *CA*—on the average—faster than in *AA*. Lapse times are sometimes even negative. This means that *CA* had access to the paper well before the printed issue of the respective journal was published. This explains for example the extremely short median lapse times for *Analytical Chemistry*: 8 days in 1990 and 10 days in 1995.

It is surprising that the median lapse time for most journals increased from 1995 to 2000. On the other hand this has led to a more uniform quality of follow-up: the ratio of the largest to the smallest median decreased from 12 to 3.

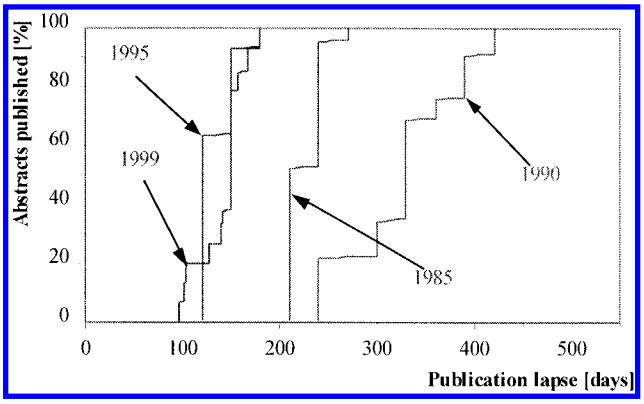

Figure 10. Publication lapse of abstracts from date of publication in *Talanta* to publication in the *AA* database.

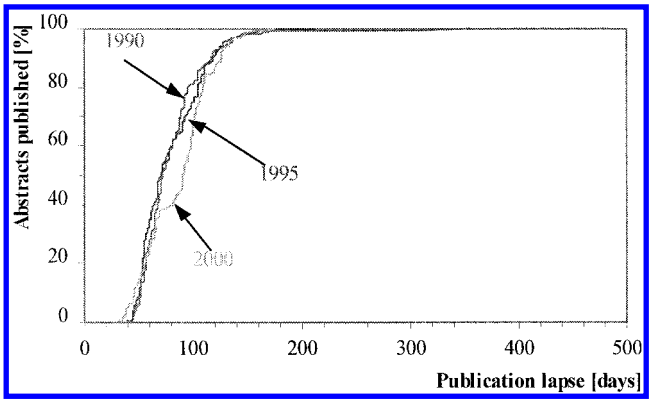

Figure 11. Publication lapse of abstracts from date of publication in *Analyst* to publication in the *CA* database.

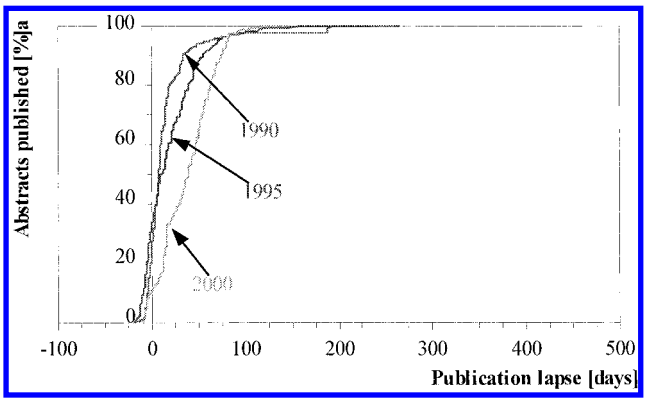

Figure 12. Publication lapse of abstracts from date of publication in *Analytical Chemistry* to publication in the *CA* database.

The quality of follow-up has been also considerably improved in the past decade by decreasing the upper limit of the range for nearly all journals. This means that the slowest abstract in a year was no more relating to well over 400 days old papers but only to papers about 200 days old.

4. COMPARISON OF AA AND CA

The quality of a bibliographic database depends on many factors. Here we compare *AA* and *CA* only with respect to their lapse time distributions observed on 10 core analytical journals in the past decade. During this time the average speed of *CA* in making abstracts available has been better than that of *AA*. Despite the observed great improvement of *AA* in the early nineties, *CA* remained faster. Both in 1995 and in 1999/2000 the median value of the median time lapses

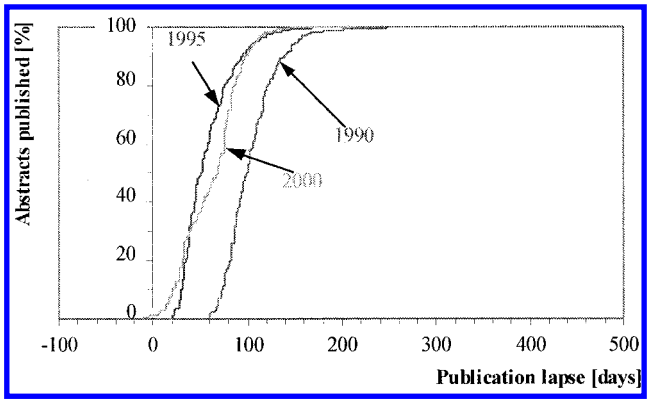

Figure 13. Publication lapse of abstracts from date of publication in *Analytica Chimica Acta* to publication in the *CA* database.

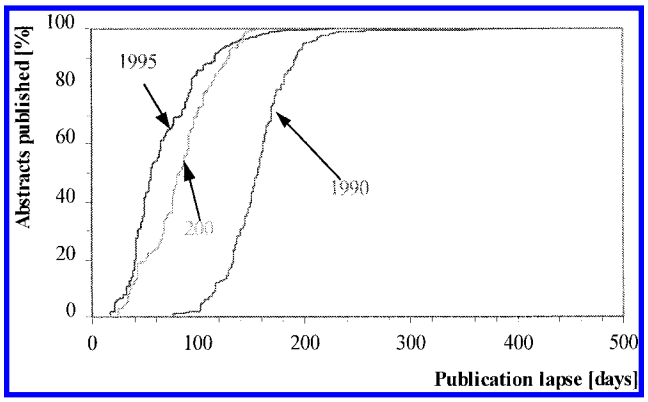

Figure 14. Publication lapse of abstracts from date of publication in *Analytical Letters* to publication in the *CA* database.

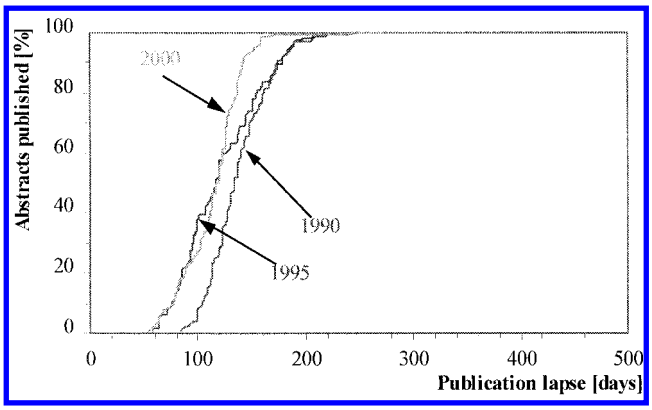

Figure 15. Publication lapse of abstracts from date of publication in *Chromatographia* to publication in the *CA* database.

of the 10 journals was about two times shorter with *CA* than with *AA*.

The spread of lapse times for abstracts of the same journal in the same year has not been very different in the two databases. In fact in 1990 and 1995 *CA* had a few abstracts linger much behind the others. Therefore the longest lapse times for some journals exceeded 400 days in 1990. Even in 1995 the slowest abstracts referred to papers 7−9 months old in the case of eight journals out of ten. This means that one could not be sure that all papers of a journal issue were covered unless one waited about 9 months after journal publication. In this respect the performance of *AA* was somewhat better in 1995: there was only one journal with maximum lapse time exceeding 7 months. In 1999/2000 the medians of the maximum lapse times were quite close for both databases: 180 days for *AA* and 182 days for *CA*.

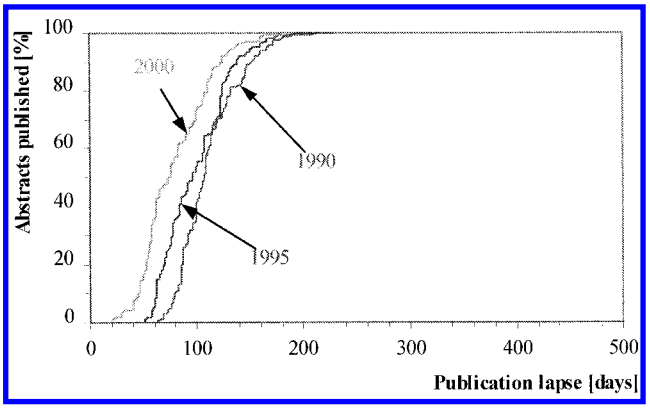

Figure 16. Publication lapse of abstracts from date of publication in *Fresenius Journal of Analytical Chemistry* to publication in the *CA* database.

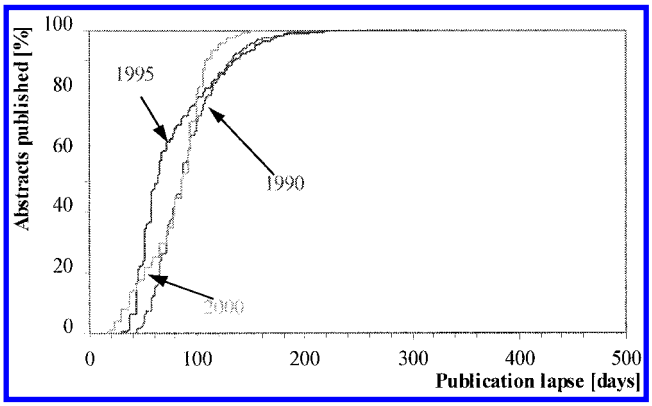

Figure 17. Publication lapse of abstracts from date of publication in *Journal of Chromatography A* to publication in the *CA* database.

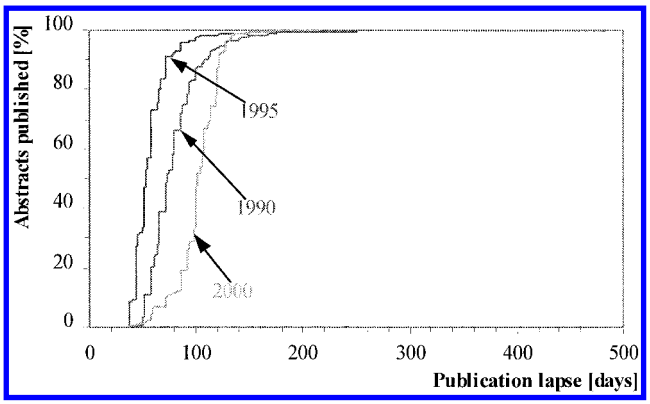

Figure 18. Publication lapse of abstracts from date of publication in *Journal of Chromatography B* to publication in the *CA* database.

5. CONCLUSION AND GENERAL REMARKS

We have studied the length of time needed for abstracts to become available to readers following the publication of an analytical chemistry paper. Since bibliographic databases are service providers, one has to judge the data by criteria of quality assurance. In this respect we consider most important the fulfilment of the expectations of the readers. Such expectations might be measured by user surveys. Here we deduce only basic expectations based on observations of the work of many colleagues. Bibliographic databases such as *AA* and *CA* are generally used for fast surveys of either the recent literature or literature retrieval of a longer time period. With respect to the first of these goals one hopes to be able to cover all recent papers of quality journals within

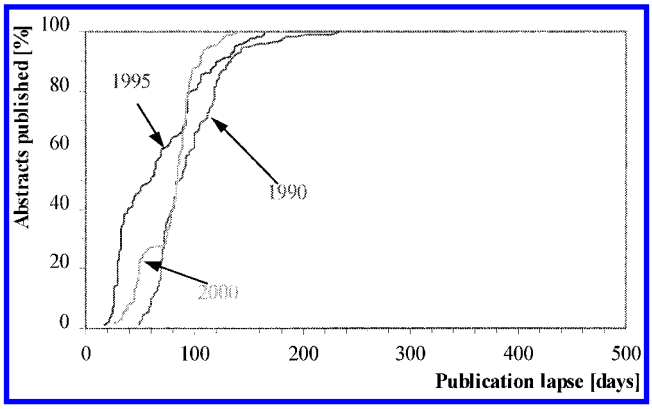

Figure 19. Publication lapse of abstracts from date of publication in *Journal of Chromatographic Science* to publication in the *CA* database.

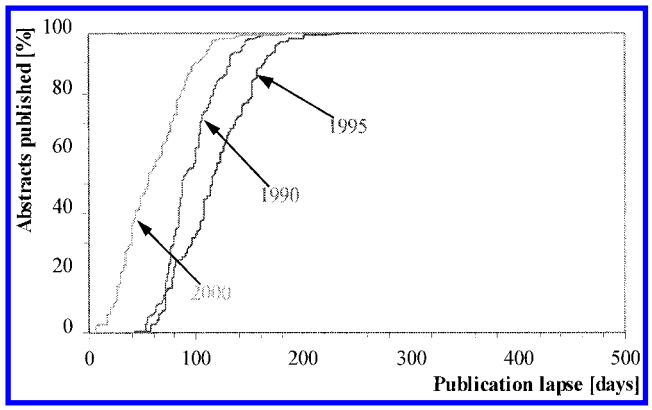

Figure 20. Publication lapse of abstracts from date of publication in *Talanta* to publication in the *CA* database.

a time frame when these papers are still relevant in supporting one's research. The word *all* should be stressed here: 80 or 90% may not be enough, as we would not like to miss 10 or 20% of the relevant papers.

When evaluated against this simple quality criterion, the performance of *CA* and particularly *AA* might still be improved. What we find the greatest problem is the reliability on uniform performance, particularly in the case of *AA*. Average users of the databases are not likely to make studies such as we did here, so they rely on their general experience with the database. Such a general experience would most likely reflect the median of medians, i.e., the median delay time of abstracts over many journals and over several years. Thus the average user is unlikely to be aware of the fact that during a browsing of the latest issue of the database he/she finds a wide mix of abstracts related to papers published in the last 7 (*CA*, 2000) or 16 (*AA*, 1999) months. In other words the user can be assured to have covered all important papers only 7 or 16 months after journal publication.

From our data we conclude that the editors of both databases have recognized that uniformity and reduction of the longest lapse time is important. They seem to have relaxed (between 1995 and 2000) the push for ever shorter median throughput time and increased their efforts to reduce the spread between journals and between papers from the same journal.

Based on this study we believe that the quality of bibliographic databases needs to be monitored regularly and the results should be made widely available to their users.

REFERENCES AND NOTES

(1) Dióspatonyi, I.; Horvai, G.; Braun, T. Publication lapse of analytical chemistry journals *J. Chem. Inf. Comput. Sci.*, pp 1452−1456.
(2) Braun, T.; Maczelka, H.; Zsindely, S. Lags in the publication of research findings in analytical chemistry. *Trends Anal. Chem.* **1993**, *12*, No. 8.
(3) Braun, T.; Glänzel, W.; Maczelka, H.; Zsindely, S. The image of analytical chemistry as reflected in the *Analytical Abstracts database. J. Chem. Inf. Comput. Sci.* **1993**, *33*, 164−173.
(4) Dióspatonyi, I.; Horvai, G.; Braun, T. A revisited auditing of the *Analytical Abstracts database J. Chem. Inf. Comput. Sci*. **2000**, *40*, 1085−1092.
(5) Braun, T. A new time-invariant definition of analytical chemistry aimed at to improve the coverage of *Analytical Abstracts. Anal. Proc*. **1993**, *30*, 216−217.
(6) http://www.rsc.org/is/database/aasour.htm.

CI010032L