

Accuracy of Cited References: The Role of Citation Databases

Robert A. Buchanan

The nature and extent of errors made by Science Citation Index Expanded™ (SCIE) and SciFinder® Scholar™ (SFS) during data entry have been characterized by analysis of more than 5,400 cited articles from 204 randomly selected cited-article lists published in three core chemistry journals. Failure to map cited articles to target-source articles was due to transcription errors, target-source article errors, omitted cited articles, and reason unknown. Mapping error rates ranged from 1.2 to 6.9 percent. SCIE and SFS also were found to correct errors made by authors in cited-article lists roughly one-half and one-sixth of the time, respectively.



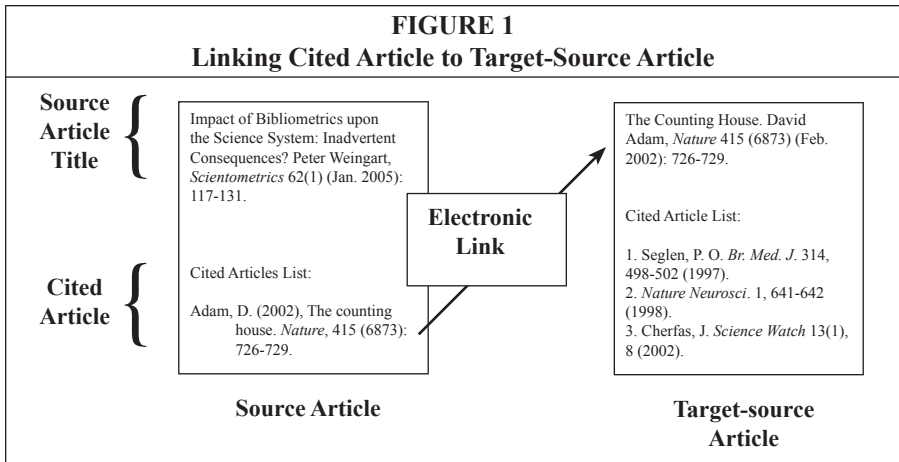
Citation databases serve two general purposes. First, they index the literature using cited articles as index terms. Second, they measure the number of times a publication has been cited in the literature, at least by that portion of the literature that the citation database indexes. Both purposes ultimately depend on the accuracy of the data on which a citation database is built. Authors make errors when creating the list of cited articles for their publications, and citation databases make errors during the data-entry process. Both types of errors can diminish the usefulness of the data in citation databases and the validity of conclusions based on those data.

This paper presents the results of a study on errors associated with cited-article lists in two citation databases that cover the chemical literature, Science Citation Index Expanded (SCIE) and SciFinder Scholar (SFS), and addresses two research questions.

- How prevalent are data-entry errors in the citation databases SCIE and SFS?
- How often do the citation databases SCIE and SFS correct errors by authors in cited-article lists?

Citation databases possess a specialized vocabulary and consist of two types of articles: those that cite and those that are cited. Those that cite are called source articles and are the basis of every record in a citation database. At a minimum, a record consists of the bibliographic information associated with the source article and its list of cited articles. The second type of article, those that are cited, is called the cited article. Many terms have been used to describe this concept, including citation, reference, and cited reference. In this article, the phrase *cited article* is used. Because the phrase *cited-reference search* has entered the popular lexicon, it is used in this article, even though *cited-article search* is more logically consistent with use of the term *cited article*.

Robert A. Buchanan is the Physical Sciences Librarian in RBD Library at Auburn University; e-mail: buchara@auburn.edu.



A target article is an article to which a cited article points.¹ At first glance, target article and cited article appear to be identical. Each can refer to the same article but performs a different role in the citation database. Electronic databases help illustrate the difference between target articles, cited articles, and source articles. When a target article is also a source article, SCIE and SFS provide an electronic link to connect the cited article to the citation database record for the corresponding source article. A new term *target-source article* has been coined because the concept is central to the current study. (See figure 1.) Target articles that are not source articles usually come from publications outside the mainstream journal literature, such as specialized journals, conference proceedings, patents, and book chapters.

Cited-reference searching in SFS and SCIE differs in two key aspects. The first difference is how the search is performed. Every cited article in the SCIE database can be the starting point for a cited-reference search. In contrast, cited-reference searching in SFS can begin only from a cited article that is also a source article (target-source article). The second difference appears in the search results and is related to the first. An SCIE-cited reference search locates the record for the source article and some records for incor-

rect forms of the source article created by author errors in cited article lists; SFS finds only the former.

Both differences arise from the history of SCIE and SFS. SCIE began its existence as the print resource Science Citation Index (SCI), which indexed scientific literature using cited articles as index terms.² Although SCI also provided title keyword access to scientific literature, it was primarily a tool to perform cited-reference searching. As a consequence, the problem of incorrect citation forms has been an important issue to SCI since its inception. Several electronic versions of SCI are available. In this article, the online database SCIE, which is available through the Web of Science (WoS) interface, has been used as the source for SCI data. Chemical Abstracts Service (CAS) began indexing chemical literature in 1907 and has only begun adding cited-article lists to CAplus database records for articles published since 1997.³ Data from CAplus are used for cited-reference searching in several CAS products, including SFS.

Literature Review

Author Errors in Cited-Article Lists

The literature on author errors in cited-article lists is extensive. James H. Sweetland evaluated the results of four large-scale studies published in the 1970s.⁴ Pooling the data from these studies, he estimated

that 7 percent of author errors were serious enough to make locating a publication difficult, such as when obtaining an article by interlibrary loan. These errors occurred in the publication title, publication year, volume number, and pagination. Errors in the name of the author, a critical element in cited-reference searches, were found in 15 percent of the cited articles. When the definition of author errors was expanded to include all types of errors, including those in the title of an article, author errors were found in 28 percent of cited articles.

A wide range of author errors has been reported in more than a dozen studies published since Sweetland's review. After organizing these studies according to their definition of author error, some general patterns emerge.

Studies that defined an author error as any error in the cited article, including those in the title of the article, reported the highest error rates.^{5,6} For example, 41 percent of cited articles in five social work journals contained at least one error⁷ and 44 to 56 percent error rates were reported in four anesthesia journals.⁸ Most of these studies are in the medical, social science, and library science literature, disciplines that include article titles as part of the cited article.

Several studies defined an author error as any error occurring in the publication title, publication year, volume number, or pagination, minimal elements of a cited article. Most reported author error rates between 2 and 4 percent,⁹⁻¹¹ but higher rates were reported in a surgery journal (10%)¹² and in a study of peer-reviewed nursing journals (15%).¹³

The definition of author error used in this study included the name of the author in addition to the publication title, publication year, volume number, and pagination. These are the five key data elements of an SCI record and are henceforth referred to as SCI data fields. Studies using this definition reported author error rates between 3 and 15 percent. Author error rates in SCI data fields have

been reported for nine analytical chemistry journals (0.7% to 6.6%),¹⁴ thirty-four biological and medical journals (15.0%),¹⁵ and an environmental medicine journal (3.4%).¹⁶

Helmut A. Abt evaluated the effect of author errors in cited-article lists on the retrieval of cited articles using the print version of SCI.¹⁷ All cited articles in one issue of *Astrophysical Journal*, which cited one of eight major astrophysics journals were examined for accuracy. Of the 12.2 percent cited articles with author errors in SCI data fields, only 3.6 percent were missing or displaced in the print version of SCI. This corresponded to SCI correcting 71 percent of author errors.

Based on matching twenty million cited articles to over eight million SCI source articles, Henk F. Moed estimated that 7 percent of cited articles contained an author error in SCI data fields.¹⁸ A thoroughly documented study compared 4,500 articles in five science journals to more than 25,000 citations to these articles using the SciSearch® database (analogous to SCIE).¹⁹ About one in ten (9.4%) of the cited articles was incorrectly cited in an SCI data field. According to Eugene Garfield, Moed's large-scale studies demonstrated the effect of author errors on establishing electronic links between cited articles and source articles in WoS.²⁰

Theoretical problems associated with citation analysis, including author errors in cited-article lists and data-entry errors, have been reviewed.²¹ The effect of clerical errors on citation analysis studies was not known, but because it was not expected to be systematic, the authors felt it should not invalidate citation analysis studies.

Errors by Citation Databases

The literature on errors by citation databases is scant. According to Garfield, no simple way exists to determine the rate of clerical errors introduced in the construction of the SCI database.²² To the best of my knowledge, no research study has focused on this topic. In particular, the accuracy of mapping cited articles to the

corresponding target-source articles in citation databases has not been reported. Only Moed and Vriens have attempted to quantify citation database errors in cited articles, albeit in a very small sample of 29 error-containing cited articles.²³ Although two of these errors might have been caused by SCI, they concluded that most errors in SCI data could be attributed to author errors.

The Institute for Scientific Information® (ISI) has employed several procedures to correct author errors in cited-article lists.²⁴ For example, a fourteen-character code based on the name of the first author, publication year, volume, and pagination matches newly entered cited articles to SCI source articles.²⁵ This not only saves data input time but also reduces the number of clerical data-entry errors. ISI has consolidated similar forms of a cited article under certain criteria, such as when one form is used more than twice as often as all other forms combined.²⁶ Garfield suggested that greater than 99 percent accuracy was achieved in the SCI Source Index but presented no data.²⁷

Jan Reedijk discussed the effect of SCI data entry errors on citation rates in the context of a few specific articles.²⁸ Slight variations in the form of cited articles led, in some cases, to severe undercitation. No systematic study on the source or effect of data-entry errors was presented. In a review comparing Web of Science and Scopus, David Goodman and Louise F. Deis observed that some articles are omitted in SCIE but provided no data.²⁹

Katherine M. Whitley compared cited reference searching in SFS and WoS by searching for source articles published between 1999 and 2001 that cited articles written by selected authors.³⁰ These databases shared 60 percent of source articles in common. Unique source articles were found in SFS and WoS: 23 and 17 percent, respectively. Most source articles from unique sources were attributed to greater coverage of the chemistry journal literature by SFS or to the greater coverage of science journals outside chemistry

by WoS. However, some unique source articles were from journals, such as the *Journal of Organic Chemistry*, which SCIE and SFS were expected to index cover-to-cover as source articles. Whitley speculated that the failure to find source articles published in major journals such as the *Journal of Organic Chemistry* appeared to be due to data-entry processing errors but provided no data.

Method

Author errors were defined as those occurring in the five SCI data fields: name and initials of the first author; publication title; publication year; volume number; and pagination. No more than one error was assigned to each cited article, in the order listed above. For example, if there was an error in the name of the first author and in the publication date, the cited article was assigned to the category for name of the first author. Although this introduced a slight bias toward name of first author errors, this was considered justifiable because incorrect names complicate cited reference searching in SCIE.

Database mapping error was defined as the failure to establish an electronic link between a cited article and the corresponding target-source article that can be attributed to a data-entry error. Database mapping errors were sorted into four categories, according to the type of error: transcription error; target-source article record error; cited article omitted from a cited-article list; and reason unknown.

If a cited article contained an author error and a database mapping error, it was assigned to the appropriate category for the former. This was done because it is unreasonable to expect citation databases to correct author errors that have not been caught by either the author or the editorial process. SCIE and SFS know that errors exist in their databases. ISI³¹ and CAS³² welcome having errors reported so that they can be corrected.

Cited-article lists from 204 articles published in three core chemistry journals were the source of the 6,313 cited articles

TABLE 1
Linkage of Cited Articles to Target-Source Articles

	SCIE (N = 5,460)	SFS (N = 5,648)
	Percent*	Percent
Not linked—author error	5.0	7.7
Not linked—mapping error	3.6	1.7
Not linked—internal page	0.3	0.2
Not linked—journal translation	0.2	0.9
Linked to target-source articles	90.8	89.5

*Total does not equal 100% due to rounding error.

used in this study. One article was randomly selected from each issue of *Inorganic Chemistry* and *Physical Chemistry Chemical Physics* (PCCP) published in 2003 and two articles from each issue of *Tetrahedron Letters* published in 2001. This generated roughly the same number of cited articles from each journal: *Inorganic Chemistry* (2,287), PCCP (2,022), and *Tetrahedron Letters* (2,004). A hard copy of the list of cited articles for each SCIE source article was printed. To this list were added the cited articles that SFS had included in its source article, but that SCIE had omitted. The resulting list was compared to a print copy of the cited-article list from the original journal article in order to add the cited articles omitted by both SCIE and SFS.

Target-source articles in SCIE that should have been linked to the corresponding cited articles were identified by

the following procedure. First, electronically linked cited articles were noted on the list described in the previous paragraph. Second, unlinked cited articles were compared to the original articles and to SCIE to determine why they were unlinked. This identified 953 cited articles that SCIE had not indexed as a target-source article; these were excluded from further

study. Third, the reason why any of the remaining 5,460 cited articles were unlinked was assigned: (1) author error; (2) citation database mapping error; (3) internal page of the article had been cited; and (4) SCIE indexed a different translation of the journal. (See table 1.) Fourth, cited articles that were unlinked due to an author error or a citation database mapping error were sorted into subcategories according to the type of error. (See tables 2 and 5.) Fifth, linked cited articles were compared to the original article to determine whether SCIE had corrected an author error during data entry (table 2). This procedure was repeated for cited article lists in SFS and generated 5,648 cited articles (excluding the 665 cited articles that SCIE had not indexed as a target-source article). Results were analyzed using an Excel® spreadsheet.

TABLE 2
Author Errors Corrected by SCIE and SFS

Error Type	SCIE		SFS	
	Number of Author Errors	Percent Corrected	Number of Author Errors	Percent Corrected
Name	201	50	203	11
Title	54	44	56	57
Year	30	73	31	10
Volume	134	57	133	11
Pagination	88	13	93	10
Total	507	46	516	16

TABLE 3
SCIE Mapping Errors

Mapping Error Type	<i>Inorganic Chemistry</i>		<i>PCCP</i>		<i>Tetrahedron Letters</i>		Combined	
	Number	% ¹	Number	% ²	Number	% ³	Number	% ⁴
Transcription	14	0.7	6	0.3	9	0.5	29	0.5
Target-source article	2	0.1	7	0.4	13	0.7	22	0.4
Omitted	118	6.1	7	0.4	21	1.2	146	2.7
Reason unknown	0	0.0	0	0.0	1	0.1	1	0.0
Total	134	6.9	20	1.2	44	2.5	198	3.6

¹Percent of 1,937 target-source articles. ²Percent of 1,732 target-source articles.
³Percent of 1,791 target-source articles. ⁴Percent of 5,460 target-source articles.

Results and Discussion

This study was based on 5,460 target-source articles in SCIE and 5,648 target-source articles in SFS identified from 204 cited article lists from three core chemistry journals. The larger number of target-source articles in SFS probably reflects the greater number of chemistry-related publications indexed by SFS. About 90 percent of the time, an electronic link connected cited articles to the corresponding target-source articles in SCIE and SFS. Careful examination of table 1 reveals that the reasons for not providing electronic links for the remaining 10 percent differed significantly.

Citation of an internal page instead of the first page of an article occurred in 0.3 percent or less of all cited articles. Almost without exception, when this happened,

neither SCIE nor SFS provided an electronic link to the target-source article. This occurred relatively infrequently because the standard practice in the chemical journal literature is to cite an article as an entity as opposed to citing specific pages of an article. Journals in disciplines in which citing specific pages is a common practice can be expected to have a higher percent of cited articles unlinked for this reason.

Journal translations pose problems for citation databases.³³ SFS and SCIE indexed source articles from either the original language journal or its English translation journal, but not both. This leads to a de facto undercount of the number of times an article published in a translation journal has been cited. It also can affect electronic linking of cited articles to

TABLE 4
SFS Mapping Errors

Mapping Error Type	<i>Inorganic Chemistry</i>		<i>PCCP</i>		<i>Tetrahedron Letters</i>		Combined	
	Number	% ¹	Number	% ²	Number	% ³	Number	% ⁴
Transcription	7	0.3	12	0.7	11	0.6	30	0.5
Target-source article	8	0.4	2	0.1	15	0.8	25	0.4
Omitted	21	1.0	9	0.5	0	0.0	30	0.5
Reason unknown	6	0.3	1	0.1	6	0.3	13	0.2
Total	42	2.1	24	1.4	32	1.7	98	1.7

¹Percent of 2,011 target-source articles. ²Percent of 1,775 target-source articles.
³Percent of 1,862 target-source articles. ⁴Percent of 5,648 target-source articles.

TABLE 5
Location of Database Mapping Errors

Error Location	SCIE		SFS	
	Transcription (Number)	Target-source Record (Number)	Transcription (Number)	Target-source Record (Number)
Name	9	5	29	15
Title	9	0	0	0
Year	3	0	0	2
Volume	3	1	0	2
Pagination	5	16	1	6
Total	29	22	30	25

target-source articles. SFS failed to link 0.9 percent of cited articles to the corresponding target-source articles because it had indexed a different version of a translated journal. In contrast, SCIE failed to link to only 0.2 percent. Indexing policies of SCIE and SFS accounted for this difference. SFS indexed source articles in the journal that first published the article, which was usually in the non-English-language journal. In most cases, SCIE indexed the English form of a translation journal. For example, in 1967, SCIE began indexing *Angewandte Chemie, International Edition in English*, the most commonly cited translation journal in this study. SFS did not begin indexing an English version of *Angewandte Chemie* until 1995.

Author errors and database mapping errors were the major reasons that SCIE and SFS did not establish a link between cited articles and target-source articles. Together, they accounted for approximately 90 percent of unlinked cited articles. These two topics are discussed in detail below.

Author Errors

Author errors in the cited-article lists of three chemistry journals have been summarized in table 2. Author errors occurred in 516 of 5,648 cited articles in SFS (9.1%) and in 507 of 5,460 cited articles in SCIE (9.3%). The total number of author errors in SCIE and SFS differed slightly because only those cited articles for which there

was a corresponding target-source article were included in this study.

The largest percent of author errors, roughly 40 percent, occurred in the name of the first author. Name errors included spelling errors in the name and initials, omission of diacritics, and the listing of a name other than that of the first author. Errors in the volume number were the second most common type of author error (26%). This was a higher percentage than other studies,³⁴ possibly because a number of cited articles were from journals, such as *Acta Crystallographica, Section C: Crystal Structure Communications*, that included the section letter with the volume number (i.e., C53). This is an example of a journal editorial policy that promotes formation of author errors. Moed observed that citation statistics for small data sets (i.e., a few individuals, a few research groups, or a few journals) can be overwhelmed by journal editorial practices.³⁵

Pagination errors accounted for about 18 percent of author errors. Citation of an internal page number was not included as a page error but was treated separately as discussed above. Most pagination errors were clerical errors, but a number arose from the practice of *Angewandte Chemie* and its English translations to begin review articles with a page-length diagram that was sometimes not recognized as the beginning of the article. Few author errors occurred in the publication title (11%), and most were due to the multiple title

changes by *Chemical Communications*. This journal began in 1965 as *Chemical Communications*, became *Journal of the Chemical Society D: Chemical Communications* in 1969, dropped the D in 1972, and returned to the title that chemists had always preferred, *Chemical Communications*, in 1996. This is another example of the effect of journal editorial policies on errors in cited articles. The least common error was in the publication year (6%).

Correction of Author Errors

Author errors were the major reason that an electronic link was not established between cited articles and target-source articles in SCIE and SFS (table 1). Although their coverage of target-source articles overlapped significantly, SCIE provided 159 more electronic links than SFS to cited articles in which the author had made an error. The reason for this large difference was because SCIE corrected many more author errors than SFS.

Both SCIE and SFS corrected a portion of author errors (table 2); however, SFS corrected significantly fewer. The only type of author error that SFS corrected at a higher rate than SCIE occurred in the publication title, most of which came from SFS recognizing situations when the title *Chemical Communications* should have been used instead of *Journal of the Chemical Society D: Chemical Communications* (with and without the "D") and vice versa.

In contrast to SFS, SCIE corrected almost half (46%) of all author errors. Although a high percentage, it was less than the 71 percent rate suggested in a study on cited articles that were missing or displaced in SCI print indexes.³⁶ SCIE was especially effective at correcting the names and initials of first authors, publication years, and volume numbers.

Unlike other types of author errors, SCIE corrected relatively few pagination errors. SCIE has used a number of error-catching procedures, which could account for the high percentage of corrections in the name of first author, publication year, and volume number errors.³⁷ However,

SCIE avoided introducing errors when applying its error-catching procedures.³⁸ SCIE treated page numbers carefully because authors sometimes publish more than one article in a single issue or a single volume. SCIE preferred to let putative author errors go uncorrected rather than introduce an error into the database by inaccurately making a correction. The percent of pagination errors corrected by SCIE (13%) was close to that corrected by SFS (10%). The scientific scholarly community is fortunate that SCIE and, to a lesser degree, SFS correct many of the mistakes made by scholars in cited-article lists.

The overall percent of cited articles that contained an author error was nearly the same in SFS (9.1%) and SCIE (9.3%). This result is deceptively similar to the rate (9.4%) reported in the largest study on author errors.³⁹ Like the present study, that study defined author error as any error occurring in SCI data fields. However, it included cited articles that referred to "in press" articles. Subtracting the author errors associated with "in press" cited articles (2.0%) resulted in a 7.4 percent error rate. Because the method used by Moed and Vriens did not identify author errors corrected by SCIE, those errors were excluded from their study, which underestimated the total percent of author errors. The low rate of author errors, less than 7 percent, disclosed in two studies that defined author errors as those in SCIE data fields also may have been due, in part, to exclusion of author errors that had been corrected by SCIE.^{40, 41}

Citation Database Mapping Errors

The most surprising result was the high number of mapping errors committed by SCIE and SFS (tables 3 and 4). Database mapping error has been defined as the failure to establish an electronic link between a cited article and the corresponding target-source article that can be attributed to a data-entry error. Four types of mapping errors were identified: transcription errors; errors in the target-

source article record; cited articles omitted from cited-article lists; and reason unknown.

Transcription errors occurred when cited articles were entered incorrectly during the indexing of a new source article. This type of error prevented the mapping of a cited article to the corresponding target-source article. Both SCIE and SFS committed transcription errors 0.5 percent of the time.

Most transcription errors committed by SCIE appeared to be clerical errors. Examples included listing the last name of an author as Kari instead of Kaji or indicating that a page number was 504 instead of 5204. SCIE made transcription errors in all five SCI data fields. SCIE distributed transcription errors relatively evenly among these categories, although errors in the name of the first author and in the publication title occurred somewhat more frequently (table 5).

Over 95 percent of transcription errors in SFS were associated with diacritics in names. For example, the name Müller can be transcribed as Muller or Mueller. A citation database can choose either transcription form but must be consistent. In most source articles, SFS included a second vowel to indicate diacritics. Unless a cited article was entered with the second vowel, SFS usually failed to map to the target-source article. Excluding diacritics, SFS made only one other transcription error. Unlike SFS, SCIE ignored diacritics when creating source article records, preferring Muller to Mueller, which explains why SCIE made no diacritics-related transcription errors.

Target-source article errors were caused by errors in the bibliographic data of a source record to which a cited article pointed. This type of mapping error had a more negative impact on a citation database than transcription errors, because it prevented any correctly entered cited article from mapping to the error-containing target-source article. Cited reference searches for an error-containing target-source record located few, if any, papers

that cited it, even when it had been cited. Target-source errors accounted for 0.4 percent of the mapping errors in SCIE and SFS.

Although 60 percent of target-source article errors in SFS occurred in the name of the first author, unlike transcription errors, none were due to diacritics. Pagination was the second most common type of target-source error in SFS (24%). SFS made few target-source article errors in the publication title, publication year, and volume number. The most common target-source error in SCIE occurred in pagination (73%). The practice of *Angewandte Chemie* and its English translations to publish reviews that began with a page-length diagram appeared to cause most pagination errors in SCIE and SFS. Like SFS, SCIE made few target-source article errors in the publication title, publication year, and volume number.

Occasionally, a target-source error has occurred in a high-profile article. For example, ISI initially treated the International Human Genome Sequencing Consortium as the author of a landmark paper on sequencing the human genome, instead of a long list of authors beginning with E.S. Lander.⁴² After ISI corrected this error in its source article, the number of source articles citing this paper increased dramatically.

The least common type of mapping error was caused by reason(s) unknown. In these cases, the data in the cited article appeared to be identical with the corresponding data in the source article, despite close examination of each and of the original paper. However, for an unknown reason, an electronic link between them did not exist. This occurred thirteen times in SFS. Although SCIE had only one unexplained mapping error, a related phenomenon was encountered during data collection. Cited-reference searches were used to locate randomly selected source articles in SCIE, but this failed to find 22 of 204 source articles used in this study. Instead, a keyword title search located them. Consistent with this behavior, database records for these source articles

incorrectly indicated that they had never been cited. The recentness of these source articles was not responsible for the lack of citing articles. Cited reference searches in SFS revealed that all twenty-two articles had been cited by source articles already indexed by SCIE.⁴³

The final type of mapping error, omission of cited articles, showed significant variation between SCIE and SFS and among the three chemistry journals (tables 3 and 4). Omission errors occurred when a cited article in a cited-article list was not included in the citation database record for the source article. Although omitted cited articles can be considered a form of transcription error, they have been treated separately because of their potential to skew analysis of mapping errors. Like transcription errors, errors of omission affect only the link between a single cited article and its target-source article. However, it is a type of mapping error that is not likely to be detected or corrected.

SFS omitted cited articles at an average rate of 0.5 percent (table 4). In contrast, SCIE omitted an average of 2.7 percent, nearly five times as many (table 3). SCIE omissions were unevenly distributed among the three journals; a single journal was responsible for 80 percent.

Several factors appeared to be associated with omitted cited articles. Placing more than one cited article under a single number (i.e., 6.a., 6.b.) had the greatest effect. Roughly 57 percent of cited articles omitted by SCIE happened when the first member of a list had been properly entered (i.e., 6.a.), but succeeding cited articles had been omitted (i.e., 6.b.). Use of footnotes instead of endnotes also appeared to correlate with higher numbers of omitted cited articles, perhaps because footnotes scatter cited articles throughout an article, making them more difficult to identify. The author can testify to the greater difficulties posed by these practices.

Extrapolation of the rate of omitted cited articles found in this study onto

the rest of the chemical literature or to the scientific literature as a whole would be premature. The method used in this study relied on how well the chemical literature was represented by the cited articles in 204 source articles from three journals. Nonetheless, that 6.1 percent of the cited articles from a core chemistry journal could be omitted from the SCIE database undermines confidence in the accuracy of its data.

Omitted cited articles strongly affected the overall rate of mapping errors. The average rate of mapping errors by SCIE for all three journals was 3.5 percent and varied significantly among journals. SFS had a lower average rate of mapping errors (1.7%) and showed less variation between journals.

What is an acceptable level of mapping errors in a citation database? Would 1 percent be acceptable as indicated by Garfield?⁴⁴ If so, the results of this study suggest that SCIE and SFS have room for improvement. Both databases made almost the same percent of transcription errors and target-source record errors: 0.5 and 0.4 percent, respectively. Combined, these two types of mapping errors almost reach 1 percent.

The biggest variable in mapping errors was the level of omitted cited articles. Can this type of mapping error be eliminated? After all, SFS omitted no cited references from *Tetrahedron Letters*. This study suggests otherwise. The level of omission errors was at least 0.4 percent in all other cases. On the other hand, it is likely that the 6.1 percent of cited articles omitted from *Inorganic Chemistry* is not representative of the SCIE database.

Comparing the number of author errors corrected by SFS and SCIE with the number of database mapping errors is instructive. These two database behaviors nearly cancelled out one another. Despite making roughly twice as many mapping errors as SFS (198 versus 98), SCIE corrected enough author errors (234 versus 82) to end up with a net reduction of errors (0.7%). SFS made many fewer map-

ping errors than SCIE but also corrected fewer author errors, resulting in a slight net increase in errors (0.2%).

Conclusions

This study examined two research questions: How prevalent are data-entry errors in the citation databases SCIE and SFS? How often do the citation databases SCIE and SFS correct errors by authors in cited-article lists?

The first question was addressed by determining why SCIE and SFS failed to map some cited articles to the corresponding target-source articles, excluding those caused by author errors. SCIE and SFS made transcription errors 0.5 percent of the time. Errors in the source articles of SCIE and SFS resulted in 0.4 percent more mapping errors. Combined, these two types of mapping errors indicate a data entry error rate of 0.9 percent. The largest, and most variable, data entry er-

ror came from omitting cited articles from cited-article lists. SCIE omitted an average of 2.7 percent; SFS omitted an average of 0.5 percent. Overall, the average percent of all database mapping errors in three chemistry journals was 3.5 percent in SCIE and 1.7 percent in SFS.

The second question was answered by examining more than 500 author errors in cited-article lists of three chemistry journals. SCIE corrected nearly one-half (46%) of these errors. In contrast, SFS corrected only about one-sixth (16%) of author errors.

This study suggests a couple of areas for further research: How representative of the chemical literature are the results of this study? How prevalent are mapping errors in the literature of other disciplines? How extensive is the phenomenon of omitted cited articles? How do other citation databases compare to the performance of SCIE and SFS?

Notes

1. Henk F. Moed and M. Vriens, "Possible Inaccuracies Occurring in Citation Analysis," *Journal of Information Science* 15 (1989): 95.
2. Eugene Garfield, *Citation Indexing: Its Theory and Application in Science, Technology, and Humanities* (New York: John Wiley, 1979): 6-18.
3. Chemical Abstracts Service, "Cited References in C_AplusSM and C_ASM," *STNNote* 24 Revised (Feb. 2005): 1-8. Available online from <http://www.cas.org/ONLINE/STN/STNOTES/stnote24.html>. [Cited 28 July 2005].
4. James H. Sweetland, "Errors in Bibliographic Citations: A Continuing Problem," *Library Quarterly* 59 (Oct. 1989): 291-304.
5. Candy K. W. Lok, Matthew T. V. Chan, and Ida M. Martinson, "Risk Factors for Citation Errors in Peer-reviewed Nursing Journals," *Journal of Advanced Nursing* 34 (Apr. 2001): 223-29.
6. Dana F. Wyles, "Citation Errors in Two Journals of Psychiatry: A Retrospective Analysis," *Behavioral & Social Sciences Librarian* 22, no. 2 (2004): 27-51.
7. Christina A. Spivey and Scott E. Wilks, "Reference List Accuracy in Social Work Journals," *Research on Social Work Practice* 14 (July 2004): 281-86.
8. M. Faith McLellan, L. Douglas Case, and Molly C. Barnett, "Trust, But Verify: The Accuracy of References in Four Anesthesia Journals," *Anesthesiology* 77 (July 1992): 185-88.
9. Margaret E. Hansen and Donald D. McIntire, "Reference Citations in Radiology: Accuracy and Appropriateness of Use in Two Major Journals," *American Journal of Roentgenology* 163 (Sept. 1994): 719-23.
10. Sung Yul Lee and Jong Suk Lee, "A Survey of Reference Accuracy in Two Asian Dermatologic Journals (the *Journal of Dermatology* and the *Korean Journal of Dermatology*)," *International Journal of Dermatology* 38 (May 1999): 357-60.
11. A. Vargas-Origel, G. Gómez-Martínez, and M. A. Vargas-Nieto, "The Accuracy of References in Paediatric Journals," *Archives of Disease in Childhood* 85 (Dec. 2001): 497-98.
12. James T. Evans, Howard I. Nadjari, and Sherry A. Burchell, "Quotational and Reference Accuracy in Surgical Journals: A Continuing Peer Review Problem," *Journal of the American Medical Association* 263 (Mar. 1990): 1353-54.
13. Lok, Chan, and Martinson, "Risk Factors for Citation Errors in Peer-reviewed Nursing Journals," 225.

14. T. Braun and Andrea Pálos, "The Accuracy and Completeness of References Cited in Selected Analytical Chemistry Journals," *Trends in Analytical Chemistry* 9, no. 3 (1990): 73–74.

15. Robert K. Poyer, "Inaccurate References in Significant Journals of Science," *Bulletin of the Medical Library Association* 67 (Oct. 1979): 396–98.

16. Jean-François Gehanno, Stefan J. Darmoni, and Jean-François Caillard, "Major Inaccuracies in Articles Citing Occupational or Environmental Medicine Papers and Their Implications," *Journal of the Medical Library Association* 93 (Jan. 2005): 118–21.

17. Helmut A. Abt, "What Fraction of Literature References Are Incorrect?" *Publications of the Astronomical Society of the Pacific* 104 (Mar. 1992): 235–36.

18. Henk F. Moed, "The Impact-Factors Debate: The ISI's Uses and Limits," *Nature* 415 (Feb. 14, 2002): 731–32.

19. Moed and Vriens, "Possible Inaccuracies Occurring in Citation Analysis," 95–107.

20. Péter Jacsó, "The Future of Citation Indexing: An Interview with Eugene Garfield," *Online* 28 (Jan./Feb. 2004): 38–40.

21. Michael H. MacRoberts and Barbara R. MacRoberts, "Problems of Citation Analysis: A Critical Review," *Journal of the American Society for Information Science* 40 (Sept. 1989): 342–49.

22. Eugene Garfield, "Errors: Theirs, Ours and Yours," in *Essays of an Information Scientist* (Philadelphia: ISI Pr., 1977), 2: 80–81. Originally published in *Current Contents* (June 19, 1974): 5–6.

23. Moed and Vriens, "Possible Inaccuracies Occurring in Citation Analysis," 102–106.

24. David Adam, "Citation Analysis: The Counting House," *Nature* 415 (Feb. 14, 2002): 726–29.

25. Eugene Garfield, "Project Keysave™: ISI's New On-Line System for Keying Citations Corrects Errors!" in *Essays of an Information Scientist* (Philadelphia: ISI Pr., 1980), 3: 42–44. Originally published in *Current Contents* (Feb. 14, 1977): 5–7.

26. ———, "Quality-Control at ISI: A Piece of Your Mind Can Help Us in Our Quest for Error-free Bibliographic Information," in *Essays of an Information Scientist*, (Philadelphia: ISI Pr., 1984), 6: 144–51. Originally published in *Current Contents* (May 9, 1983): 5–12.

27. ———, "Journal Editors Awaken to the Impact of Citation Errors. How We Control Them at ISI," in *Essays of an Information Scientist: Journalology, KeyWords Plus, and other Essays* (Philadelphia: ISI Pr., 1990), 367–75. Available online from <http://www.garfield.library.upenn.edu/essays/v13p367y1990.pdf>. [Cited 28 July 2005]. Originally published in *Current Contents* (Oct. 8, 1990): 5–13.

28. Jan Reedijk, "Sense and Nonsense of Science Citation Analyses: Comments on the Monopoly Position of ISI and Citation Inaccuracies. Risks of Possible Misuse and Biased Citation and Impact Data," *New Journal of Chemistry* 22 (1998): 767–70.

29. David Goodman and Louise F. Deis, "Web of Science (2004 version) and Scopus," *Charleston Advisor* 6(3) (Jan. 2005). Available online from <http://www.charlestonco.com/comp.cfm?id=43>. [Cited 28 July 2005].

30. Katherine M. Whitley, "Analysis of SciFinder Scholar and Web of Science Citation Searches," *Journal of the American Society for Information Science and Technology* 53 (Dec. 2002): 1210–15.

31. Adam, "Citation Analysis," 728.

32. Chemical Abstracts Service, "CAS User Comments and Questions Form." Available online from <http://www.cas.org/rform.html>. [Cited 28 July 2005].

33. Werner Marx, "Angewandte Chemie in Light of Science Citation Index," *Angewandte Chemie, International Edition in English* 40 (Jan. 2001): 139–43.

34. Sweetland, "Errors in Bibliographic Citations," 296.

35. Moed, "The Impact-Factors Debate," 731.

36. Abt, "What Fraction of Literature References are Incorrect?" 236.

37. Garfield, "Quality-Control at ISI," 144–51.

38. ———, "Journal Editors Awaken to the Impact of Citation Errors," 373.

39. Moed and Vriens, "Possible Inaccuracies Occurring in Citation Analysis," 99.

40. Braun and Pálos, "The Accuracy and Completeness of References Cited in Selected Analytical Chemistry Journals," 73–74.

41. Gehanno, Darmoni, and Caillard, "Major Inaccuracies in Articles Citing Occupational or Environmental Medicine Papers and Their Implications," 119.

42. "Errors in Citation Statistics," *Nature* 415 (Jan. 10, 2002): 101.

43. These statements were true as of March 1, 2005. On July 27, 2005, 11 of the 22 articles could not be located in SCIE by a cited-reference search.

44. Garfield, "Journal Editors Awaken to the Impact of Citation Errors," 367.