E. Ashley Rogers Brown. Building a Better Mousetrap: Capturing New Data in ISI Journal Citation Reports and Local Journal Utilization Reports to Support Academic Collection Managers. A Master's Paper for the M.S. in L.S degree. April, 2006. 62 pages. Advisor: Catherine Blake

The ISI provides librarians with tools such as the Journal Citation Reports (JCR) and the institution specific Local Journal Utilization Report (LJUR) to aid in the management of serials collections. These tools enable librarians to respond quickly to changes in publishing practices and purchasing options. While available literature often criticizes ISI data, few studies provide concrete recommendations for improvement. This study explores two extensions to LJUR: (1) adding citation date and (2) creating institution specific impact factors. In addition, I explore the degree to which self-citations influence the ISI impact factor. Publication and citation calculations are made for three prominent southern universities' research chemists using a corpus of full text articles drawn from 27 American Chemical Society (ACS) journals and stored in an Oracle database. The ACS research corpus impact factor simulation and ACS research corpus self-citation omission impact factor are also created and compared with current JCR data.

Headings:

ISI, Impact Factor, Citation Analysis, Publication Analysis, Collection Management, Journal Selection and Deselection

BUILDING A BETTER MOUSETRAP: CAPTURING NEW DATA IN ISI JOURNAL CITATION REPORTS AND LOCAL JOURNAL UTILIZATION REPORTS TO SUPPORT ACADEMIC COLLECTION MANAGERS

by
E. Ashley Rogers Brown

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Library Science.

Chapel Hill, North Carolina

April 2006

Approved by

____

Catherine Blake

Table of Tables

Table of Figures

1 Introduction

Academic collection managers use existing data in conjunction with faculty feedback, library mission and local collection development guidelines to analyze journal collections to best serve institutional faculty, students and researchers while maintaining collection and budget control. Updating current data tools available in Thomson Scientific's ISI Journal Citation Reports (JCR) and Local Journal Utilization Report (LJUR) through the addition of citation date to the LJUR database citation reports, creating a university specific Impact Factor as part of the LJUR package and providing a Self-Citation Omission Impact Factor to the current JCR will supply tools addressing the more detailed citation and impact factor data necessary to build and maintain a journal collection that effectively supports the institutional faculty research. I will utilize an American Chemical Society corpus (ACS research corpus) of 27 electronic chemistry journals and the aggregated publication and citation data available from the articles in this corpus.

This research considers 103,262 published articles (publications) and their 3.2 million citations in 27 ACS journals. To explore the possible value of institution specific information I have included chemists at Duke University (Duke), North Carolina State University (NCSU), the University of North Carolina at Chapel Hill (UNC). In comparison, ISI holds "over 5,900 leading international science journals" (Thomson Scientific, Journal Citation Reports, Features), from which to draw publication and citation data for their JCR and LJUR data.

The JCR includes journal and subject specific publication and citation data garnered from 5,900 journals. Publication data counts all articles published in a journal in a given year while citation data counts citations in the respective journals. The citation counts include self-citations, defined as articles published in a journal that cite that same journal. The ISI Impact Factor incorporates the citation and publication data as a ratio of citations from a two year period to publications in that same two year period. (See ISI Impact Factor Definition and Use in the Background Chapter of this work for the Impact Factor formula) My research simulates these categories (publication, citation and impact factor), to answer the following questions:

(1) Does the JCR and LJUR data allow collection managers to identify gaps in their collection, including backfile and archival needs?

(2) Is the JCR too general for collection managers to effectively support the needs of institutional faculty, students and researchers?

(3) Is the current LJUR data sufficient for collection managers to make difficult purchasing decisions considering current publishing trends?

These questions warrant an overview of current journal publishing trends.

The number of available journals is rising, as are costs and available formats and package structures for existing materials in print and electronic form. Many publishers offer "current" subscriptions to electronic journals covering a set number of years (e.g.1995-present) for an ongoing cost, while also offering electronic archives with coverage, in some cases, from the journal's inception to the first year of the "current subscription (e.g. 1901-1995) for a one-time fee and nominal ongoing fee.

Publishers now offer "bundled" title packages with publisher defined content inclusion. These packages offer a lower per title cost, but often pair the most reputable titles with lesser known or lower quality titles. Title-by-title purchase options are still available from most publishers, at a higher per title cost. Consortia agreements with publishers may allow bundled title access to all consortia members with all institutions contributing to a higher package cost.

Chemistry journal collection evaluation pervades library literature. In 1927, Gross and Gross were utilizing citation analysis for the identification of Chemistry journals that would support graduate education. While a more recent study, at the University of Illinois at Urbana-Champaign Chemistry Library determined that "…84 percent of journal use was generated by 20 percent of the collection…" (Davis, 2002, p. 156) The distinct temporal contexts support an ongoing theme of academic library journal collection evaluation.

The assertion of Line and Pan that "no measure of journal use other than one derived from a local use study is of any significant practical value to librarians," (Line & Pan, 1978, p. 313) still holds today. While the literature often criticizes ISI data, I believe affecting change warrants analytical investigation and proposal of action. Journal collection management necessitates specific data that is both comprehensive and institution specific. The research described herein is an attempt to identify this data.

It is my supposition that LJUR publication data is effective in its current form. The corresponding citation data; however is insufficient. The current format of year of journal publication-citation count lacks the information necessary for archival collection decisions. Without consideration for the citation year, differentiating between citations from current research or research from twenty years ago is impossible. A simple change in the query structure gathers this information. Armed with this addition the librarian could alter rank lists as well as craft an informed faculty survey to determine if gaps exist in the current collection. Identification of such gaps could determine whether the collection team should make a one-time archival purchase or maintain an ongoing cost subscription.

The creation of an ACS research corpus impact factor simulation (ACS-IF-Sim) and an ACS Self-Citation Omission Impact Factor (SCO-IF), I will address literature suggestions that the ISI Impact Factor self-citation inclusion establishes an opportunity for publisher manipulability of the impact measure.

As an introduction to ISI, the Background Chapter defines the Impact Factor and identifies limitations. The Methodology Chapter steps through the pre-processing of the ACS research corpus, description of publication, citation, ACS-IF-Sim and SCO-IF collection process, as well as the collection of ISI data for comparison. The Results and Analysis Chapter steps through the findings of the research and is followed by the Recommendations and Conclusions Chapters.

2 Background

2.1 ISI Impact Factor Definition and Use

In 1927, P.L.K. Gross and E.M. Gross attempted to address the question, “What files of scientific periodicals are needed in a college library successfully to prepare the student for advanced work, taking into consideration also those materials necessary for the stimulation and intellectual development of the faculty?” (Gross & Gross, 1927, p. 386) More specifically, Gross and Gross were interested in these issues in relation to Chemistry as an area of institutional study. Their methods consist of a citation analysis of “a single volume of *The Journal of the American Chemical Society*.”(p. 386) Since this time, citation analysis has been a frequent tool for librarians to analyze journal collections for selection and deselection.

Though Gross and Gross were one of the first to utilize citation analysis for collection assessment, Eugene Garfield is probably the most famous. Garfield proposes in his 1955 article, the creation of “new bibliographic tools that can help to span the gap between the subject approach of those who create documents…and the subject approach of the scientist who seeks information.” (Garfield, 1955, p. 108) Garfield points to the Gross and Gross methods and suggests the creation of an “impact factor,” but increases the complexity of the citation analysis metric. Ultimately, Garfield’s ideas came to fruition and are currently known as *Thomson Scientific’s (ISI)Impact Factor* (previously known as *The Institute of Scientific Information*) and "is the ratio of articles published to articles cited during a rolling two-year window." (Cameron, 2005, p. 108) Peter Vinkler defines the ISI Impact Factor as follows:

> The impact factor of journals calculated by Equation 1 has been termed as the *Garfield Factor* [sic] (GF):
>
> $$GF_y = \frac{C_y}{P_{y-1} + P_{y-2}}$$
>
> [where GFy is the Garfield Factor of a journal in year y, Cy is the total number of citations (self-citations included) received in year y by the papers (articles, reviews, letters, notes) published in the respective journal in years (y-2) and (y-1), and Py-1 and Py-2 stand for the number of papers published in the respective years. [Note: impact factors published in SCI Journal Citation Reports by the Institute for Scientific Information (Philadelphia, PA) refer to the number of citations (numerator) received by the journals to any papers but in the denominator only articles, notes, and reviews are counted as source items.] (Vinkler, 2004, p.431)

Today *Thomson Scientific* offers a variety of products to libraries and research institutions around the world. Some of these products include citation data on the institutional and national levels, custom analysis and publication data. (Thomson Scientific,Research Services) Thomson also provides a "Local Journal Utilization Report," or LJUR. According to the Thomson Scientific website:

> The Local Journal Utilization Report is a statistical database listing the frequency with which an institution's researchers publish in ISI indexed journals, and the frequency with which they cite ISI journals and other works (theses, government reports, etc.) in their publications. Frequencies are calculated both annually and cumulately from 1981 through the current year. (Thomson Scientific, Local Journal Utilization Report)

These citation and publication reports "[cover] more than 7,500 of the world's most highly cited, peer-reviewed journals in approximately 200 disciplines." (Thomson Scientific, Journal Citation Reports, Features)

2.2 Limitations of ISI Data

Throughout the Thomson Website, you will also find data on the use and abuse of ISI Impact Factors and citation data. In a 2005 article, Brian Cameron suggests, “There is now clear evidence that impact factors can be (and are being) manipulated by publishers.” (p. 105) He further asserts,

> Editors can take several steps to artificially improve the journal’s impact factor. Increasing the number of review articles and simply increasing the number of articles published in the journal can positively affect impact factor. In addition, serial publication of segments of research, multiple publication of the same research either in identical or modified form, publishing more informal items like letters, editorials, and discussion papers will increase the impact factor. An editor may require the inclusion of more references in every article published, assuming that their journal would benefit by receiving some of these citations. Requiring that a certain number of citations be made from articles previously published in their journal will increase the journal’s impact factor. In addition, a journal may choose to unveil new paradigms, host controversies, or solicit papers from authors with a good citation history, something that Garfield feels many editors do instinctively. (p. 117)

Yet another consideration is the addition of electronic resources. Cameron again suggests that, “…it has been proven that electronic access to a journal increases its impact factor, partly because the journals are available 24 hours per day.” (p. 120) The exclusion of conference proceedings, which some disciplines rely more heavily upon for their research, indicates further Impact Factor flaws.

In an essay entitled “*The ISI Impact Factor,*” found on the Thomson website, Garfield promotes the responsible use of available data: “Informed and careful use of these impact data is essential. Users may be tempted to jump to ill-formed conclusions based on impact factor statistics unless several caveats are considered.” (Garfield, 1994) He goes on to state, “These data must also be combined with cost and circulation data to make rational decisions about purchases of journals,” and later:

> The Institute for Scientific Information® (ISI®) does not depend on the impact factor alone in assessing the usefulness of a journal, and neither should anyone else. The impact factor should not be used without careful attention to the many phenomena that influence citation rates, as for example the average number of references cited in the average article. The impact factor should be used with informed peer review. In the case of academic evaluation for tenure it is sometimes inappropriate to use the impact of the source journal to estimate the expected frequency of a recently published article. Again, the impact factor should be used with informed peer review. Citation frequencies for individual articles are quite varied.
> (Garfield, 1994)

Important to note here are the components of the ISI Impact Factor. Impact factor calculations include “articles…reviews…editorials, letters, news items, and meeting abstracts.” (Journal Citation Reports, Help, Source Data) These citations are then utilized to calculate the impact factor as described below:

> A journal’s impact factor is based on two elements: the numerator, which is the number of cites in the current year to any items published in the journal in the previous 2 years; and the denominator, the number of substantive articles (source items) published in the same 2 years. (Garfield, 2005, p. 5)

The rolling two year scope of the impact factor, the inclusion of self-citation to manipulate impact factors and the use of impact factors in faculty review for tenure comprise a large portion of the debate over the validity of impact factors. In this paper, the primary focus will be self-citation factors.

According to Hoeffel (1998), “Impact Factor is not a perfect tool to measure the quality of articles but there is nothing better and it has the advantage of already being in existence and is, therefore, a good technique for scientific evaluation.” (p. 1225) While many research institutions follow this tenet, the voices of descent also need careful consideration. Over the years, ISI’s Impact Factor has been both widely used and highly controversial. (Cameron, 2005) The overuse of impact is the subject of myriad articles, one of which is Joseph Wible’s 1990 article, comparing citation analysis, swept use and the ISI Impact Factor to determine the most comprehensive method for deselection evaluation. (Wible, 1990) His results indicate that "citation and swept use data complement each other, and together provide a more complete picture of an individual title's value to the collection." (1990, p. 110)) His statements about the ISI Impact Factor are less flattering, "relying solely on ISI's data would result in serious deselection errors." (1990, p. 116) The return to a more straightforward use of citation analysis is reflected in anumber of articles written in the recent past on citation analysis as a collection development tool. (Black, 2001; Brown, 2002; Calhoun, 1995; Davis, 2002; Fuseler-McDowell, 1990; Walcott, 1994; Wible, 1990)

In some instances, citations are still hand collected from a random sample of representative units (i.e. theses or print article copies provided by an individual department). A review of the literature, however, reveals in more recent years, the use of manual citation analysis is limited to the analysis of theses and dissertations. (Edwards, 1999; Gooden, 2001) In Edwards'1999 study 25% of available theses and dissertations in the "area of polymer science and polymer engineering from 1990-1996" were analyzed to identify and categorize the citations into "periodical, monograph, conference proceeding/paper, patent, standard, thesis/dissertation, or other." (p. 14) Of these categories, periodicals were then pulled into "Top Twenty Five Journals Cited" (p. 16) ranked list tables. Edwards believes "it is appropriate to consider citation analysis in combination with other quantitative measures, such as in-house use data and interlibrary loan statistics, in order to develop a more comprehensive picture of users' research needs." (1999, p. 19) Hand coding is time-intensive and illustrates some of the preference given to automated methods such as ISI data, as suggested by Davis (2002), "While not supplanting other rigorous study methods…[electronic] bibliographic analysis…provides an easy, fast, and low-cost study that can be adapted to various bibliographic databases and scaled as needed…" (p. 156) Davis warns "while impact factors are an excellent guide for identifying key journals in particular fields, no campus is typical and as such, impact factors (if used at all) should be supplemented by local data." (p. 157)

In more recent citation analysis studies, librarians and researchers alike, pull on their available electronic resources to identify citations by their own faculty. (Bauer & Bakkalbasi, 2005; Davis, 2002; LaBonte, 2005; Vaughan & Shaw, 2005) The studies that use electronic resources, result in a ranked list of journals, applied to the current collection. (Vaughan & Shaw, 2005) The resulting lists are used to make decisions in journal selection and deselection, the purchase of print and electronic subscriptions as well as identification of the library's core collection. (Davis, 2002) Unfortunately, unlike Wible, the researchers often use these citation analysis statistics alone. Karen LaBonte, for instance simplistically describes the citation analysis research in her conclusion, "Citation analysis is a practical tool to evaluate how a library is meeting the needs of local users…If the need arises to make cuts to serials budgets…this data can be used to find the least cited material." (2005) While her statements are true, citation analysis is but one tool for selection and deselection.

While the progression of citation analysis as a tool is well documented, a further step is indicated. As suggested by Line, "no measure of journal use other than one derived from a local use study is of any significant practical value to librarians."(Line & Pan, 1978, p. 313) Bollen, et al. employ these recommendations through a study which “applies…metrics to journal networks…derived from the Journal Citation Records and from sequential journal download patterns registered in the log files of a large Digital Library (DL).” (2005, p. 1420) In this way they create their own impact formula as a local alternative to the more global impact factor data. The uses of citation analysis, criticisms of ISI data and creation of more institution specific tools, drives the research herein.

3 Methodology

3.1 Pre-Processing

The American Chemical Society (ACS) provided Dr. Catherine Blake, Assistant Professor at the University of North Carolina at Chapel Hill, with more than 100,000 HTML articles from 27 journals for research purposes. Under the direction of Dr. Blake the HTML documents were pre-processed to identify headings, figures, tables, text, and citations. This initial work was complete before this study began.

The HTML citation formats in the American Chemical Society research corpus (ACS research corpus) vary between and within documents. The primary formats identified are:

1. Citations with only one article cited,
2. Citations with multiple articles cited,
3. Notes by the author, or
4. Combination of 1, 2 and 3 above.

Table 3.1A illustrates an amalgam of 2 and 3 in that it is a citation containing a note and multiple articles. Table 3.1A also demonstrates the reader's view of the citation in a browser such as Microsoft Explorer. The number '3' in the citation indicates that it is the third citation in this article. Following the citation number 3, is a note from the author, "For recent discussions of protein databases and search algorithms, see the following and references therein." The letters in parentheses that follow (e.g. (a), (b), etc.) indicate sub-citations of citation 3.

BROWSER VIEW:
3. For recent discussions of protein databases and search algorithms, see the following and references therein: (a) Eng, J. K.; McCormack, A. L.; Yates, J. R., III *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976. (b) Yates, J. R., III.; Eng, J. K.; Clauser, K. R.; Burlingame, A. L. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 1089. (c) Jensen, O. N.; Podtelejnikov, A. V.; Mann, M. *Anal. Chem.* **1997**, *69*, 4741.[Full text - ACS][Medline] (d) Reiber, D. C.; Grover, T. A.; Brown, R. S. *Anal. Chem.* **1998**, *70*, 673.[Full text - ACS] (e) Yates III, J. R. *Electrophoresis* **1998**, *19*, 893. (f) Clauser, K. R.; Baker, P.; Burlingame, A. L. *Anal. Chem.* **1999**, *71*, 2871.[Full text - ACS]

Table 3.1A: View of a citation as seen in a browser such as Microsoft Explorer. The number 3 at the beginning indicates it is the third citation in this article; while the letters in parentheses indicate sub-citations.

The citation pre-processing goal was to assign unique identifiers to each of the articles. Unfortunately, the underlying HTML added another level of complexity to the parsing process. Table 3.1B displays the underlying HTML text for the Browser View in Table 3.1A. As with most HTML, the original text was a continuous string; however I have separated each of the sub-citations into separate rows for clarity.

AS TEXT APPEARS IN HTML:
<A NAME="ac0007783b00003"><P>**3.** For recent discussions of protein databases and search algorithms, see the following and references therein:
(a) Eng, J. K.; McCormack, A. L.; Yates, J. R., III <I>J. Am. Soc. Mass Spectrom.</I> <B>1994</B>, <I>5</I>, 976.<citation reflink="journal=J.+Am.+Soc.+Mass+Spectrom.&authorname=Eng&year=1994&volume=5&page=976&reference=Eng, J. K.; McCormack, A. L.; Yates, J. R., III J. Am. Soc. Mass Spectrom. 1994, 5, 976."><ZAZ HREF="APP=ftslink&action=reflink&origin=ACS&version=0.0&coi=1:CAS:528:DyaK2MXhvF2htLw%253D&journal=J.%2BAm.%2BSoc.%2BMass%2BSpectrom.&author=Eng&pubyear=1994&volume=5&startpage=976&reference=Eng%2C%20J.%20K.%3B%20McCormack%2C%20A.%20L.%3B%20Yates%2C%20J.%20R.%2C%20III%20J.%20Am.%20Soc.%20Mass%20Spectrom.%201994%2C%205%2C%20976.&md5=47ffd4deebb2061fd35cadf297949e27"><CASLINK></ZAZ>
(b) Yates, J. R., III.; Eng, J. K.; Clauser, K. R.; Burlingame, A. L. <I>J. Am. Soc. Mass Spectrom. </I><B>1996</B>, <I>7</I>, 1089.<citation reflink="journal=J.+Am.+Soc.+Mass+Spectrom.&authorname=Yates&year=1996&volume=7&page=1089&reference=Yates, J. R., III.; Eng, J. K.; Clauser, K. R.; Burlingame, A. L. J. Am. Soc. Mass Spectrom. 1996, 7, 1089."><ZAZ HREF="APP=ftslink&action=reflink&origin=ACS&version=0.0&coi=1:CAS:528:DyaK28XmsF2ltL8%253D&journal=J.%2BAm.%2BSoc.%2BMass%2BSpectrom.&author=Yates&pubyear=1996&volume=7&startpage=1089&reference=Yates%2C%20J.%20R.%2C%20III.%3B%20Eng%2C%20J.%20K.%3B%20Clauser%2C%20K.%20R.%3B%20Burlingame%2C%20A.%20L.%20J.%20Am.%20Soc.%20Mass%20Spectrom.%201996%2C%207%2C%201089.&md5=e818e48c8f4e16d5c3f12d07091409a2"><CASLINK></ZAZ>

(c) Jensen, O. N.; Podtelejnikov, A. V.; Mann, M. <I>Anal. Chem.</I> <B>1997</B>, <I>69</I>, 4741.<citation reflink="journal=Anal.+Chem.&authorname=Jensen&year=1997&volume=69&page=4741&reference=Jensen, O. N.; Podtelejnikov, A. V.; Mann, M. Anal. Chem. 1997, 69, 4741."><A HREF="http://pubs.acs.org/cgi-bin/citation?ancham/69/i23/html/ac970896z.html">[Full text - ACS]</A><ZAZ HREF="APP=ftslink&action=reflink&origin=ACS&version=0.0&coi=1:CAS:528:DyaK2sXntVagtb4%253D&journal=Anal.%2BChem.&author=Jensen&pubyear=1997&volume=69&startpage=4741&reference=Jensen%2C%20O.%20N.%3B%20Podtelejnikov%2C%20A.%20V.%3B%20Mann%2C%20M.%20Anal.%20Chem.%201997%2C%2069%2C%204741.&md5=6d600b80a927a35c088e33b10a57f35d"><CASLINK></ZAZ><A HREF="http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?db=m&form=6&uid=9406524&Dopt=r">[Medline]</A>
(d) Reiber, D. C.; Grover, T. A.; Brown, R. S. <I>Anal. Chem.</I> <B>1998</B>, <I>70</I>, 673.<citation reflink="journal=Anal.+Chem.&authorname=Reiber&year=1998&volume=70&page=673&reference=Reiber, D. C.; Grover, T. A.; Brown, R. S. Anal. Chem. 1998, 70, 673."><A HREF="http://pubs.acs.org/cgi-bin/citation?ancham/70/i04/html/ac971157l.html">[Full text - ACS]</A><ZAZ HREF="APP=ftslink&action=reflink&origin=ACS&version=0.0&coi=1:CAS:528:DyaK1cXjvV2ksg%253D%253D&journal=Anal.%2BChem.&author=Reiber&pubyear=1998&volume=70&startpage=673&reference=Reiber%2C%20D.%20C.%3B%20Grover%2C%20T.%20A.%3B%20Brown%2C%20R.%20S.%20Anal.%20Chem.%201998%2C%2070%2C%20673.&md5=629c3c004b74173ed1df3825d592fa81"><CASLINK></ZAZ>
(e) Yates III, J. R. <I>Electrophoresis</I> <B>1998</B>, <I>19</I>, 893.<citation reflink="journal=Electrophoresis&year=1998&volume=19&page=893&reference=(e) Yates III, J. R. Electrophoresis 1998, 19, 893.">
(f) Clauser, K. R.; Baker, P.; Burlingame, A. L. <I>Anal. Chem.</I> <B>1999</B>, <I>71</I>, 2871.<citation reflink="journal=Anal.+Chem.&authorname=Clauser&year=1999&volume=71&page=2871&reference=Clauser, K. R.; Baker, P.; Burlingame, A. L. Anal. Chem. 1999, 71, 2871."><A HREF="http://pubs.acs.org/cgi-bin/citation?ancham/71/i14/html/ac9810516.html">[Full text - ACS]</A><ZAZ HREF="APP=ftslink&action=reflink&origin=ACS&version=0.0&coi=1:CAS:528:DyaK1MXjtlWqs78%253D&journal=Anal.%2BChem.&author=Clauser&pubyear=1999&volume=71&startpage=2871&reference=Clauser%2C%20K.%20R.%3B%20Baker%2C%20P.%3B%20Burlingame%2C%20A.%20L.%20Anal.%20Chem.%201999%2C%2071%2C%202871.&md5=4902920942280529d6726978efd489c6"><CASLINK></ZAZ>

Table 3.1B: HTML view of the same browser citation in Table 3.1A. NOTE: Individual sub-citations have been separated for ease of identification purposes only and are not segmented in the original HTML text.

The parser collects citation information from parameters in the hyperlink. The hyperlink starts with <citation reflink= and ends with </A>. The parameters in Table 3.1B for reference (f) are: journal, &authorname, &year, &volume, &page and &ref. One challenge with using the hyperlink is missing citation information instances. For example, in Table 3.1A sub-citation (e) is Yates III, J. R. *Electrophoresis* 1998, 19, 893.

While this appears complete in the Browser view, the hyperlink fails to include the &authorname parameter and is thus excluded from the final database (See Table 3.1C).

The processed data from the HTML text data was subsequently loaded into an Oracle database. Table 3.1C demonstrates the final form of the HTML text from Table 3.1B. The DOCID refers to the document identification number isolated from the HTML text (Table 3.1B) in the initial portion of the citation, prior to the citation number, in this case 3. This document identification number refers to a specific article published in the ACS research corpus. The SECTION ID, AUTO CIT ID, AND SUB CIT ID, were generated from the portions of a multi-part citation and refer to the section of the paper (Bibliography is section12 in this example), automatically generated citation identification (e.g. Citation 3) and a sub-citation identification in which letter identifiers are changed to numbers. The actual citation identification (ACT_CIT_ID) acts as check for the automatically generated identification number. Most important for this research,

AS TEXT APPEARS IN DATABASE								
ALL_CITATIONS								
DOCID	SECTION ID	AUTO CIT ID	ACT CIT ID	SUB CIT ID	CIT JNL	CIT AUTHOR1	CIT PUB YR	CIT VOL
AC0007783	12	3	3	2	J. Am. Soc. Mass Spectom.	Yates	1996	7
AC0007783	12	3	3	3	Anal. Chem.	Jensen	1997	69
AC0007783	12	3	3	4	Anal. Chem.	Clauser	1999	71
AC0007783	12	3	3	5	Anal. Chem.	Reiber	1998	70
AC0007783	12	3	3	6	J. Am. Soc. Mass Spectom.	Eng	1994	5

Table 3.1C: Same citation in Database View (Some categories have been omitted for simplicity). Note the absence of sub-citation (e) from the Browser and HTML Views. This is due to the absence in the HTML of the &author= tag.

are the document identification number, cited journal name, citation author1 (indicating only first authors were recorded in this table), the citation journal's publication year (CIT

PUB YR) and the citation volume (CIT VOL). These will be addressed later Section 3.2: Data Collection.

The text from two journals had publication formatting inconsistencies, but the citation hyperlinks were less affected, so the publication data reflects only 25 of the 27 journals in the ACS research corpus, while the citation data reflects all 27 journals.

The resulting corpus comprises 103,262 articles with 3.2 million total citations. These citations include primarily journal articles, but also monographic citations. To ensure data integrity within the loaded data, the research team, made up of Dr. Blake, myself and two additional master's candidates at the University of North Carolina at Chapel Hill School of Information and Library Science performed two checks. First, to ensure that citation information was collected, and then to ensure that the journal names were correct. To check citation information accuracy we compared citations from100 randomly selected articles with the citations in the database. Only articles in the two error-prone journals showed more than minor differences.

To ensure the journal name accuracy, the research team manually unified titles that (a) occurred in more than 10,000 citations or (b) were in the ACS research corpus. Table 3.2 demonstrates some of the variations identified for just one of the titles in the ACS research corpus. Discrepancies included spelling and abbreviation variations, the use of 'and' or '&,' and the presence or omission of periods after the abbreviations. The variations were collected in a new table, JOURNAL_ CROSS_REFERENCE and associated with a unified journal title. During subsequent table creations, new tables were joined with the JOURNAL_CROSS_REFERENCE table to ensure that journal titles indicated as a match were in fact the same. When joining tables, as you will see in

CITATION_JOURNAL_TITLES	UNIFIED_JOURNAL_TITLE
Crys. Growth Des.	Cryst. Growth Des.
Cryst. Grow. Des.	Cryst. Growth Des.
Cryst. Growth De.	Cryst. Growth Des.
Cryt. Growth Des.	Cryst. Growth Des.
Cryst. Growth Des.	Cryst. Growth Des.
Cryst. Growth. Des.	Cryst. Growth Des.
Crystal Growth Des.	Cryst. Growth Des.
Cryst. Growth Des. 4	Cryst. Growth Des.
Cryst. Growth Design	Cryst. Growth Des.
Cryst. Growth Des. 1	Cryst. Growth Des.
Cryst. Growth Des. 4	Cryst. Growth Des.
Crystal Growth Design	Cryst. Growth Des.
Cryst. Growth . Design	Cryst. Growth Des.
Cryst. Growth & Des.	Cryst. Growth Des.
Crystal Growth and Design	Cryst. Growth Des.
Crystal Growth & Design	Cryst. Growth Des.

Table 3.2 Sample of original journal title variation and unified journal title from the journal cross reference table in the ACS research corpus database.

the following section, it was especially important to include the JOURNAL_CROSS_ REFERENCE table when using the ALL_CITATIONS table as it contains the greatest number of inaccuracies.

3.2 Data Collection

Data was collected with the primary goal of analyzing the value of current ISI publication, citation and impact factor data for academic library collection managers. The simulation of ISI formats act as a baseline for the development of more institutional based data not currently available from ISI.

Data for this study was collected in two phases. The preliminary research was performed as an independent study in the fall of 2006, while the second portion covers the direct development of my master's thesis. The independent study results were recalculated based on some original errors in execution; however, the processes for both represented the same research goals. I have broken the methodology that follows into subsections based on my collection of publication, citation and ISI Impact Factor data.

3.2.1 Collecting Publication Data

I compared the five journals in the ACS research corpus where faculty most often publish (hereafter referred to as "publication") with ISI publication data for these same journals. I calculated publication frequencies per university, using chemistry faculty from Duke University (Duke), North Carolina State University (NCSU) and the University of North Carolina at Chapel Hill (UNC). In addition to the ISI data, I will also do an inter-university analysis to identify possible need for institution specific data. These titles were drawn from 103,262 articles published in the 25 ACS research corpus journals between 2000-2004.

The journal titles were isolated using abbreviated wildcard versions of faculty surnames to allow for variations in publication name (e.g. Smith% was used rather than Smith, John to allow for author names such as Smith, J. or Smith, J. T.) I then manually matched each author name and faculty member at each university, including previous names where applicable. Three tables (See Figure 3.1) were produced in the Oracle database containing author name, publication journal, publication volume, publication issue, document identification number for article, author identification number and the article publication year. These tables are DUKE_AUTHORS, NCSU_AUTHORS, and UNC_ AUTHORS and are resources for both publication and citation identification. In this schema, Unified_Journal_Title describes the publication article journal title. The join between CORPUS_AUTHORS and ALL_AUTHORS made it unnecessary to use JOURNAL_CROSS_REFERENCE because the Publication_Journal_Title in

CORPUS_AUTHORS was derived directly from the ACS research corpus HTML journal title and only contains the 25 journals in the ACS research corpus. CORPUS_AUTHOR only contains each article once (through the use of the first author only), whereas ALL_AUTHOR creates a separate entry for each credited author. ALL_AUTHOR allows for capturing individual university chemistry faculty who appear as second author or beyond.

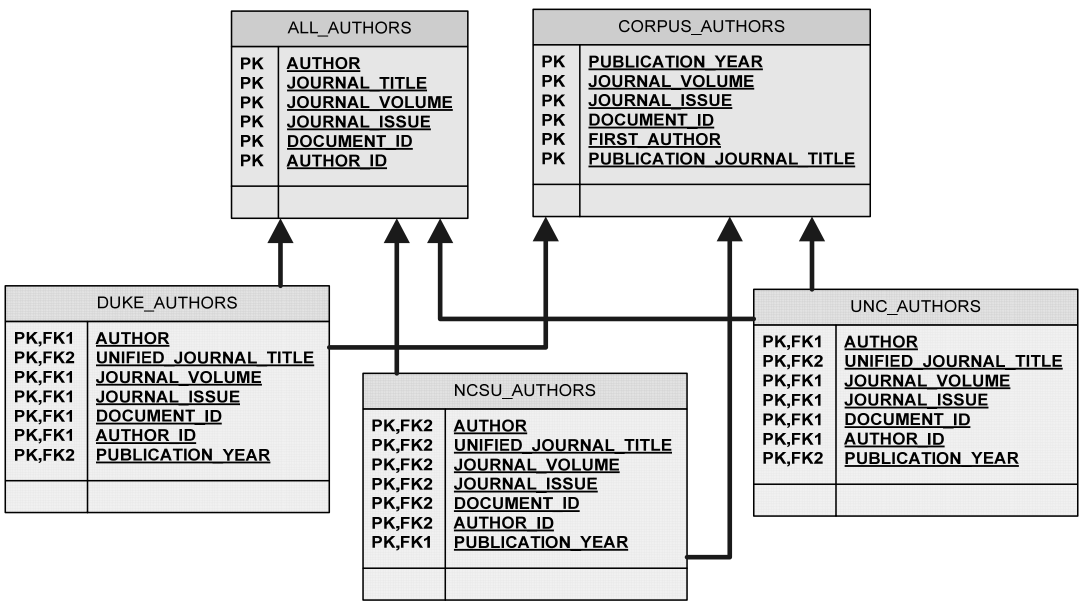

Figure 3.1: Schema for the creation of university specific author tables.

Next I queried the database to calculate publications frequencies for each university, each journal, and in each year. The tables DUKE_PUBLICATION_COUNT, NCSU_PUBLICATION_COUNT and UNC_PUBLICATION_COUNT capture these statistics. Table 3.3 draws on data from these tables to identify the top five publication

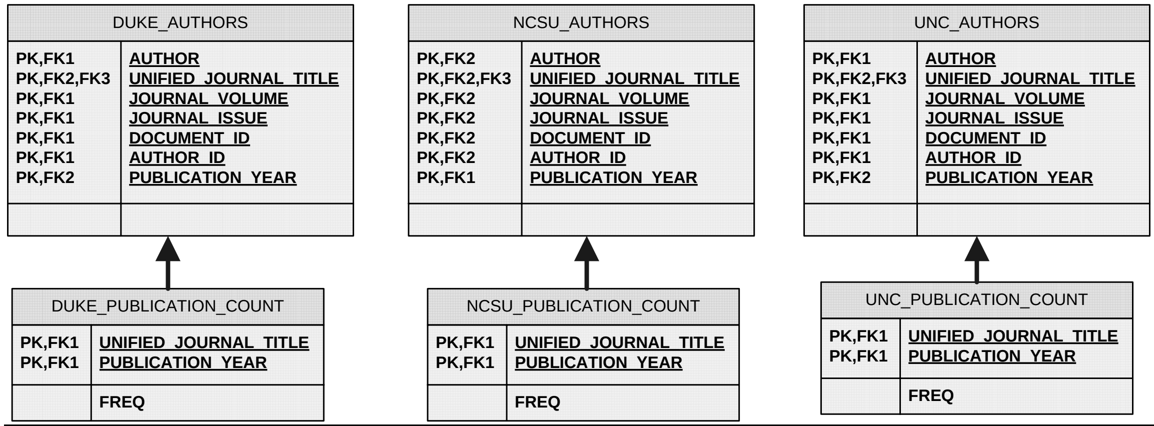

Figure 3.2: Schema for the creation of university specific publication frequency tables.

journals by institution. The ACS research corpus covers the years 2000-2004[1], and therefore publication data results correspond to these years. The query results were then transferred to Excel for format manipulation to mirror that of the ISI publication data. The resulting table is shown in Table 3.3:

DUKE						
JNL	**Total Publications All Years**	**2004**	**2003**	**2002**	**2001**	**2000**
J. Am. Chem. Soc.	105	32	26	18	12	17
J. Phys. Chem. B	57	10	17	11	11	8
Biochemistry	51	15	9	9	13	5
Inorg. Chem.	36	2	7	5	11	11
Org. Lett.	27	9	5	4	4	5

NCSU						
JNL	**Total Publications All Years**	**2004**	**2003**	**2002**	**2001**	**2000**
J. Am. Chem. Soc.	61	12	23	8	11	7
Inorg. Chem.	47	10	10	9	11	7
J. Org. Chem.	46	15	4	5	8	14
J. Phys. Chem. B	29	9	6	5	6	3
Anal. Chem.	23	4	5	4	8	2

UNC						
JNL	**Total Publications All Years**	**2004**	**2003**	**2002**	**2001**	**2000**
J. Am. Chem. Soc.	161	27	37	33	33	31
Anal. Chem.	59	9	13	9	11	17
Macromolecules	54	7	10	9	19	9
J. Phys. Chem. A	48	16	8	14	3	7
J. Phys. Chem. B	47	6	12	11	13	5

Table 3.3: Top five journal titles by university chemistry faculty publication in the American Chemical Society research corpus

I collated these tables into one comprehensive table to identify differences between the universities' top five publication journal results (See Table 3.4, differences highlighted).

JNL	DUKE	NCSU	UNC
Anal. Chem.	22	23	59
Biochemistry	51	8	38
Inorg. Chem.	36	47	43
J. Am. Chem. Soc.	105	61	161
J. Org. Chem.	25	46	15
J. Phys. Chem. A	15	7	48
J. Phys. Chem. B	57	29	47
Macromolecules	25	10	54
Org. Lett.	27	18	29

Table 3.4: Combined top five publication journals by university, differences highlighted.

3.2.2 Collecting Citation Data

I compared the five journals in the ACS research corpus that faculty most often cited (hereafter referred to as "citation") with ISI citation data for these same journals. I calculated citation frequencies per university, using chemistry faculty from Duke University (Duke), North Carolina State University (NCSU) and the University of North Carolina at Chapel Hill (UNC). In addition to the ISI data, I will also do an inter-university analysis to assess the need for institution specific data. These titles were drawn from 103,363 articles published in the 27 journals in the ACS research corpus between 2000-2004.

The journal titles were isolated using the three previously described ACS research corpus database author tables, DUKE_AUTHORS, NCSU_AUTHORS and UNC_AUTHORS (See Figure 3.1).

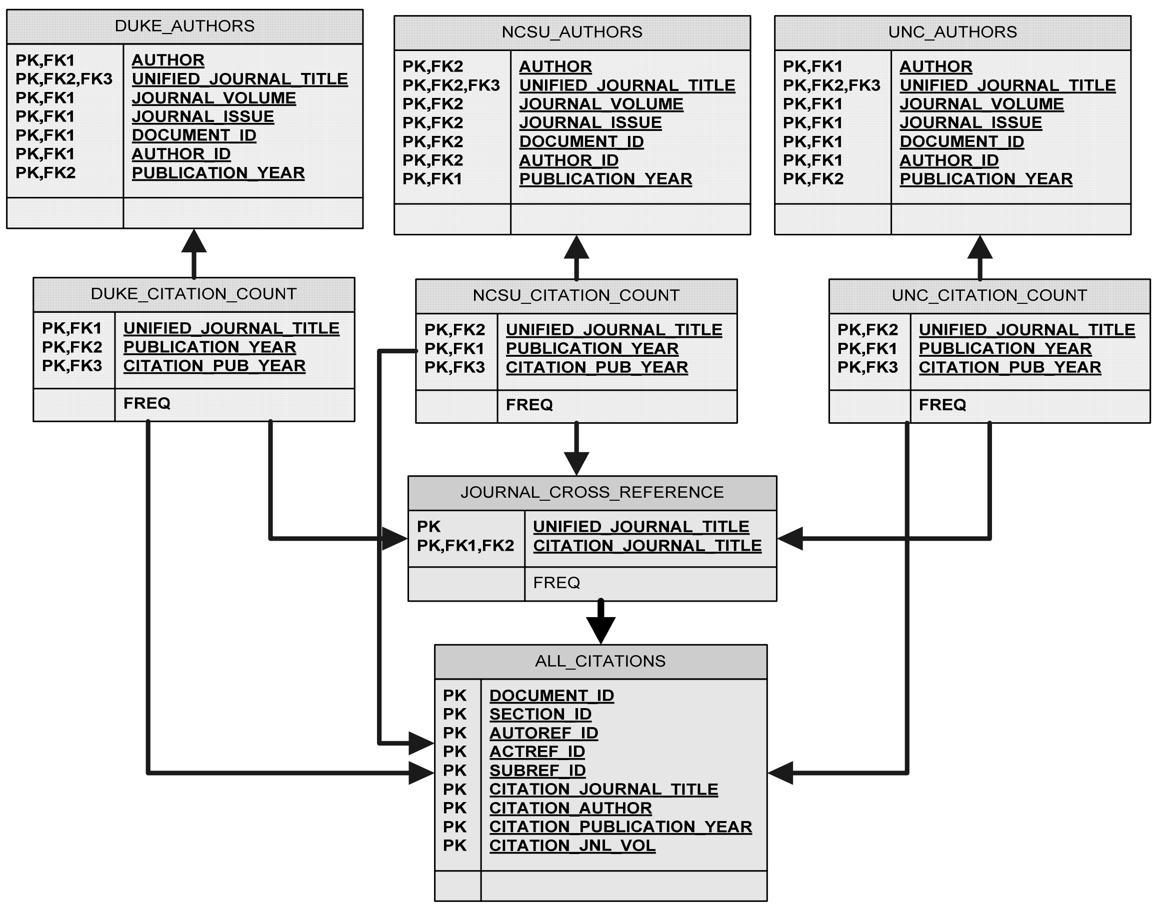

Figure 3.3: Schema for the university citation count tables. Each table is created with a join of the university author table and the ALL_CITATIONS table through JOURNAL_CROSS_REFERENCE to ensure unified journal titles.

Next I created three new citation frequency tables corresponding to each university by joining the university specific author table with the ALL_CITATIONS table connecting on the Document_ID and with the JOURNAL_CROSS_REFERENCE table on Citation_Journal_Title to ensure journal name unification. (See Figure 3.3 for schema)

The resulting tables, DUKE_CITATION_COUNT, NCSU_CITATION_COUNT AND UNC_CITATION_COUNT, include the article publication year, the unified citation journal name, the citation publication year and a frequency count. Each table was queried to identify the top five citation journals by institution. Because the data in the ACS research corpus covers the years 2000-2004, the citation data results are delineated by article publication year. The query results were transferred to Excel for format manipulation to mirror that of the ISI publication data. The resulting table is shown in Table 3.5:

DUKE						
JNL	**Total Citations All Years**	**2004**	**2003**	**2002**	**2001**	**2000**
Biochemistry	513	136	90	68	149	70
J. Am. Chem. Soc.	1634	492	305	217	333	287
J. Biol. Chem.	362	78	62	75	99	48
J. Org. Chem.	375	100	67	48	83	77
Science	438	146	96	75	62	59

NCSU						
JNL	**Total Citations All Years**	**2004**	**2003**	**2002**	**2001**	**2000**
Anal. Chem.	233	52	48	43	53	37
Inorg. Chem.	389	41	100	61	118	69
J. Am. Chem. Soc.	1346	218	411	205	315	197
J. Org. Chem.	610	162	181	50	92	125
Tetrahedron Lett.	241	35	77	30	51	48

UNC						
JNL	**Total Citations All Years**	**2004**	**2003**	**2002**	**2001**	**2000**
Anal. Chem.	733	143	191	90	137	172
Inorg. Chem.	754	124	129	115	194	192
J. Am. Chem. Soc.	2888	572	600	519	681	516
J. Phys. Chem.	538	102	98	141	103	94
Macromolecules	580	116	118	104	178	64

Table 3.5: Top five journal titles by university chemistry faculty citation in the ACS research corpus.

I collated these tables into one comprehensive table to identify differences between the universities' top five citation journal results (See Table 3.6, differences highlighted).

Journal Title	DUKE	NCSU	UNC
Anal. Chem.	244	233	733
Biochemistry	513	175	390
Inorg. Chem.	261	389	754
J. Am. Chem. Soc.	1634	1346	2888
J. Biol. Chem.	362	70	181
J. Org. Chem.	375	610	299
J. Phys. Chem.	212	121	538
Macromolecules	252	138	580
Science	438	200	385
Tetrahedron Lett.	266	241	268

Table 3.6 Combined top five citation journals by university, differences highlighted.

Once the top five titles were designated, I ran an additional query on the individual university citation count tables to determine the years from which these citations were drawing. The query results are delineate by article publication year, as in the previous example, but additionally the citation publication year creating a more comprehensive data set for analysis. (See Chapter 4.2 Citation Analysis)

3.2.3 Collecting ACS Research Corpus Impact Factor Simulations

3.2.3.1 Impact Factor

I implemented the ISI Impact Factor calculation using the CORPUS_AUTHOR, ALL_CITATIONS and JOURNAL_CROSS_REFERENCE, previously described. (See Figure 3.4)

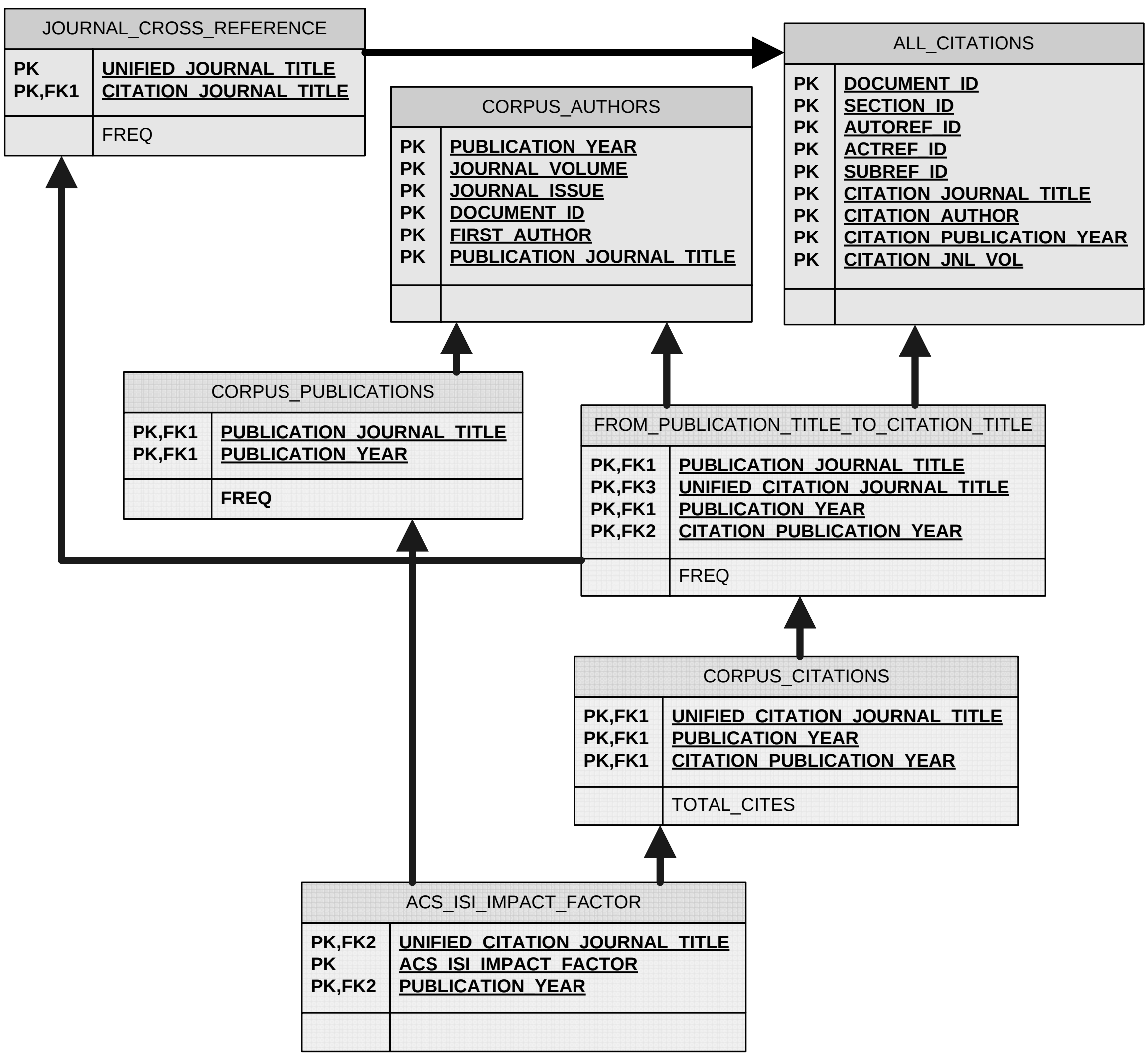

Figure 3.4: Schema used for creation of ACS research corpus Impact Factor simulation.

CORPUS_AUTHOR differs from ALL_AUTHOR used in the creation of the university author tables (e.g.. DUKE_AUTHORS) in that CORPUS_AUTHOR only contains each article once (through the use of the first author only), whereas ALL_AUTHOR creates a separate entry for each author credited in an article. While using ALL_AUTHOR was important for identifying faculty from each university associated with articles, the duplication of articles has a negative effect on the simulation of the ISI impact formula.

I created CORPUS_PUBLICATIONS by querying CORPUS_AUTHORS to isolate the publication journal name, publication year and a frequency count. (See Figure 3.5) The resulting table contains the 25 ACS research corpus journal titles delineated by publication year. CORPUS_AUTHOR was not joined with JOURNAL_CROSS_ REFERENCE as titles in CORPUS_AUTHOR were created directly from the ACS research corpus HTML journal title and only contain the 25 journals in the ACS research corpus. Two copies of CORPUS_PUBLICATIONS comprise the denominator of the ACS research corpus ISI Impact Factor formula simulation (ACS-IF-Sim).

```
CREATE TABLE CORPUS_PUBLICATIONS
AS SELECT Publication_Journal_Title, Publication_Year, COUNT(*) AS Freq
FROM FIRST_AUTHOR
GROUP BY Publication_Journal_Title, Publication_Year;
```

Figure 3.5: SQL CREATE statement for CORPUS_PUBLICATIONS used in the denominator of the ACS research corpus ISI Impact Factor simulation.

Next I created FROM_PUBLICATION_TITLE_TO_CITATION_TITLE (hereafter referred to as the “FROM_TO” table) by joining CORPUS_AUTHOR with ALL_CITATIONS on Document_ID and ALL_CITATIONS with JOURNAL_ CROSS_REFERENCE on Citation_Journal_Title. (See Figure 3.6 for SQL)

FROM_TO providS citation frequencies for every publication journal-citation journal combination by publication year and citation year. In order to obtain a total cites to a journal in a year, regardless of the citing journal, I created an additional table, the CORPUS_CITATIONS table which summed the frequencies for each journal

```
CREATE TABLE FROM_PUBLICATION_TITLE_TO_CITATION_TITLE
AS SELECT B.Publication_Journal_Title, C.Unified_Journal_Title AS
          Unified_Citation_Journal_Title, B.Publication_Year,
          A.Citation_Publication_Year, COUNT(*) AS Freq
FROM ALL_CITATIONS A, CORPUS_AUTHORS B,
     JOURNAL_CROSS_REFERENCE C
WHERE A.Document_ID=B.Document_ID
AND A.Citation_Journal_Title=C.Citation_Journal_Title
GROUP BY B.Publication_Journal_Title, C.Unified_Journal_Title, B.Publication_Year,
         A.Citation_Publication_Year;
```

Figure 3.6: SQL CREATE statement for FROM_TO

based on the Unified_Citation_Journal_Title, Publication_Year and Citation_Year. Two copies of CORPUS_CITATIONS comprises the numerator of the ACS-IF-Sim.

To create the ACS-IF-Sim, I then performed a sum on the frequencies in two copies of CORPUS_CITATIONS, where CORPUS_CITATIONS one (C1) represents citations to articles of the previous year (i.e. one year ago) and CORPUS_CITATIONS two (C2) represents citations to articles of the year previous to that (i.e two years ago), over a sum of the frequencies in two copies of CORPUS_ PUBLICATIONS, where CORPUS_PUBLICATIONS one (P1) represents articles published in the previous year (i.e. one year ago) and CORPUS_ PUBLICATIONS two (P2) represents the articles published in the year prior to that (i.e. two years ago). For clarity see the example in Figure 3.7, where IF is the ISI Impact Factor simulation. The formula in Figure 3.7 allows me to simulate ISI Impact Factors for those journals within the ACS research corpus publication year scope of 2000-2004.

<u>IF year = 2004</u>

C1 = citations to articles published in 2003
C2 = citations to articles published in 2002
P1 = published articles in 2003
P2 = published articles in 2002

$$IF_{2004} = \frac{C1_{(2003)} + C2_{(2002)}}{P1_{(2003)} + P2_{(2002)}}$$

Figure 3.7: ISI Impact Factor formula (Journal Citation Reports)

Following the same general format in Figure 3.7, the 2004 ISI Impact Factor calculation for *Crystal Growth & Design* is as follows:

<u>*Crystal Growth & Design* IF year = 2004</u>

C1 = citations to articles published in 2003 = **365**
C2 = citations to articles published in 2002 = **252**
P1 = published articles in 2003 = **124**
P2 = published articles in 2002 = **92**

$$IF_{2004} = \frac{C1_{(2003)} + C2_{(2002)}}{P1_{(2003)} + P2_{(2002)}} = \frac{365 + 252}{124 + 92} = \frac{617}{216} = \mathbf{2.856}$$

Figure 3.8: ISI Impact Factor calculation for *Crystal Growth & Design*. (Journal Citation Reports)

Figure 3.9 demonstrates the complexity of the SQL for ACS-IF-Sim (ACS_ISI_IMPACT_FACTOR in Figure 3.4).

The creation of this SQL statement was an iterative process and ultimately checked through manual calculation of the Impact Factor for the ACS-IF-Sim of one of the titles in the ACS research corpus. The manual calculation was then compared to the ACS-IF-Sim number for this same journal. The resulting numbers were found to be the same.

```
CREATE TABLE ACS_ISI_IMPACT_FACTOR
AS SELECT C1.Unified_Citation_Journal_Title, C1.Publication_Year,
          SUM(C1.Total_Cites+C2.Total_Cites)/SUM(P1.Freq+P2.Freq) as
          ACS_ISI_Impact_Factor
FROM CORPUS_CITATIONS C1, CORPUS_CITATIONS C2,
     CORPUS_PUBLICATIONS P1, CORPUS_PUBLICATIONS P2
WHERE C1.Publication_Year=C1.Citation_Publication_Year+1
AND C1.Publication_Year=C2.Publication_Year
AND C2.Publication_Year=C2.Citation_Publication_Year+2
AND C1.Unified_Citation_Journal_Title=C2.Unified_Citation_Journal_Title
AND C1.Unified_Citation_Journal_Title=P1.Publication_Journal_Title
AND C1.Publication_Year=P1.Publication_Year+1
AND C1.Unified_Citation_Journal_Title=P2. Publication_Journal_Title
AND C1.Publication_Year=P2.Publication_Year+2
GROUP BY C1.Unified_Citation_Journal_Title, C1.Publication_Year;
```

Figure 3.9: SQL Create statement for ACS research corpus ISI Impact Factor simulation

ACS_ISI_IMPACT_FACTOR was then queried for ACS-IF-Sim numbers for 2003 and imported into Excel alongside the ISI Impact Factor numbers for 2003 for analysis.(See Section 3.2.4 Collecting ISI Data) Table 3.7 illustrates the result of this process.

Journal Title	ISI Impact Factor 2003	ACS_IF_ Simulation 2003
Acc. Chem. Res.	15	4.3691
Anal. Chem.	5.25	1.2029
Biochemistry	3.922	0.7340
Biomacromolecules	2.824	0.7486
Biotechnol. Prog.	1.488	0.2466
Chem. Mater.	4.374	1.0988
Cryst. Growth Des.	2.742	0.6755
Environ. Sci. Technol.	3.592	0.8229
Ind. Eng. Chem. Res.	1.317	0.3241
Inorg. Chem.	3.389	1.0977
J. Agric. Food Chem.	2.102	0.6064
J. Am. Chem. Soc.	6.516	2.0998
J. Chem. Info. Comput. Sci.	3.078	0.9112
J. Comb. Chem.	4.2	1.6337
J. Med. Chem.	4.82	0.8489
J. Nat. Prod.	1.849	0.3453

Journal Title	ISI Impact Factor 2003	ACS_IF_ Simulation 2003
J. Org. Chem.	3.297	0.9755
J. Phys. Chem. A	2.792	0.8926
J. Phys. Chem. B	3.679	1.0127
J. Proteome Res.	5.611	*
Macromolecules	3.621	1.1101
Nano Lett.	5.65	1.8700
Org. Lett.	6.144	1.3176
Org. Proc. Res. Dev.	4.092	0.1688

Table 3.7: 2003 ISI Impact Factors and ACS-IF-Simfor journals in the ACS research corpus. *Insufficient data in ACS research corpus prevented the ACS-IF-Sim calculation for the *Journal of Proteome Research*.

3.2.3.2 Self-Citation Omission-Impact Factor (ACS-SCO-IF)

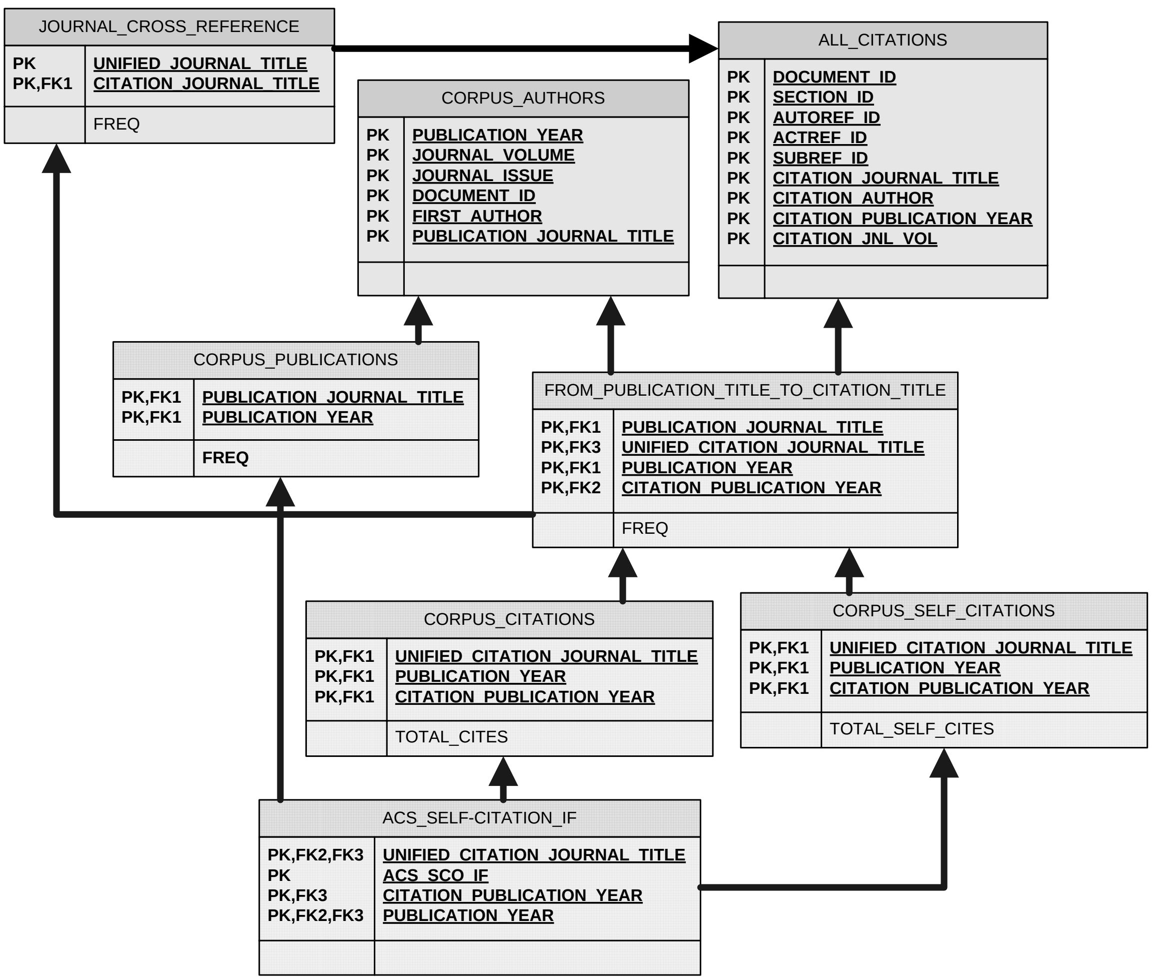

Figure 3.10: ACS research corpus Self-Citation Omission-Impact Factor schema.

To explore the influence of self-citation on the ISI Impact Factor I calculated the number of times that each journal contained self-citations. Figure 3.10 displays the new schema with the addition of the self-citation tables while Figure 3.11 displays the SQL for this calculation.

```
CREATE TABLE ACS_SCO-IF
AS SELECT C1.Unified_Citation_Journal_Title, C1.Publication_Year,
                SUM(C1.Total_Cites+C2.Total_Cites-S1.Self_Cites-
                S2.Self_Cites)/SUM(P1.Freq+P2.Freq) as ACS_SCO-IF
FROM CORPUS_CITATIONS C1, CORPUS_CITATIONS C2,
         CORPUS_PUBLICATIONS P1, CORPUS_PUBLICATIONS P2,
         CORPUS_SELF_CITES S1, CORPUS_SELF_CITES S2
WHERE C1.Publication_Year=C1.Citation_Publication_Year+1
AND C1.Publication_Year=C2.Publication_Year
AND C2.Publication_Year=C2.Citation_Publication_Year+2
AND C1.Unified_Citation_Journal_Title=C2.Unified_Citation_Journal_Title
AND C1.Unified_Citation_Journal_Title=P1.Publication_Journal_Title
AND C1.Publication_Year=P1.Publication_Year+1
AND C1.Unified_Citation_Journal_Title=P2. Publication_Journal_Title
AND C1.Publication_Year=P2.Publication_Year+2
AND C1.Publication_Year=S1.Publication_Year
AND C1.Unified_Citation_Journal_Title=S1.Unified_Citation_Journal_Title
AND C1.Citation_Publication_Year=S1.Citation_Publication_Year
AND C2.Publication_Year=S2.Publication_Year
AND C2.Unified_Citation_Journal_Title=S2.Unified_Citation_Journal_Title
AND C2.Citation_Publication_Year=S2.Citation_Publication_Year
GROUP BY C1.Unified_Citation_Journal_Title, C1.Publication_Year;
```

Figure 3.11: SQL for the creation of the ACS Self-Citation Omission Impact Factor (ACS-SCO-IF).

In the creation of ACS-SCO-IF I eliminated self-citation for the resulting set. For clarity see the following example, where ACS-SCO-IF is the Self-Citation Omission Impact Factor created with the ACS research corpus:

SCO-IF year = 2004

C1 = citations to articles published in 2003
$C1_{self\text{-}cite}$ = journal self-citation in 2003
C2 = citations to articles published in 2002
$C2_{self\text{-}cite}$ = journal self-citation in 2002
P1 = published articles in 2003
P2 = published articles in 2002

$$SCO\text{-}IF_{2004} = \frac{(C1_{(2003)} - C1_{self\text{-}cite}) + (C2_{(2002)} - C2_{self\text{-}cite})}{P1_{(2003)} + P2_{(2002)}}$$

Figure 3.12 Altered Impact Factor formula (Journal Citation Reports) for self-citation elimination

Following the same generalized format above, the 2004 Self-Citation Omission Impact Factor calculation for *Crystal Growth & Design* is as follows:

Crystal Growth & Design AIF year = 2004

C1 = citations to articles published in 2003 = **365**
$C1_{self\text{-}cite}$ = journal self-citation in 2003 **= 73**
C2 = citations to articles published in 2002 = **252**
$C1_{self\text{-}cite}$ = journal self-citation in 2003 **= 47**
P1 = published articles in 2003 = **124**
P2 = published articles in 2002 = **92**

$$AIF_{2004} = \frac{(C1_{2003} - C1_{self\text{-}cite}) + (C2_{2002} - C2_{self=cite})}{P1_{2003} + P2_{2002}}$$

$$\frac{(365 - 73) + (252 - 47)}{124 + 92} = \frac{497}{216} = \mathbf{2.301}$$

Figure 3.13: The calculation for the ISI Impact Factor excluding journal self-citation. Though the self-citation data is available on the ISI Citation Reports website (i.e. C1 and C2), the Self-Citation Omission Impact Factor is not available. Calculations for this data are left to the individual (Journal Citation Reports)

This formula only produces ACS-SCO-IFs for those journals within the ACS research corpus and is limited by the publication year scope of 2000-2004. The resulting ACS-

SCO-IFs were then added to Table 3.7, which includes ISI Impact Factors gathered from the JCR and ACS-IF-Sim described in Section 3.2.2.1.

3.2.4 Collecting ISI Data

I gathered the ISI citation, publication and Impact Factor data for 2000-2004 from the ISI Journal Citation Reports (JCR) via a subscription purchased by the University Libraries at the University of North Carolina at Chapel Hill. These data sets are readily available by using specific journal names. The ISI publication and citation tables below (Table 3.8 and Table 3.9) indicate data collected on the collated top five publication journals by faculty at each university (Table 3.8) and the collated top five citation journals by faculty at each university (Table 3.9). The JCR is not university specific, but act as a guideline for results analysis in the ACS research corpus collected institutional statistics.

ISI Citation Reports						
JNL	**Total Publications All Years**	**2004**	**2003**	**2002**	**2001**	**2000**
Anal. Chem.	4694	1007	966	906	883	932
Biochemistry	8656	1687	1637	1754	1755	1823
Inorg. Chem.	5358	1146	1186	978	1107	941
J. Am. Chem. Soc.	13401	3167	2994	2680	2327	2233
J. Org. Chem.	7153	1399	1549	1439	1370	1396
J. Phys. Chem. A	7302	1419	1426	1519	1439	1499
J. Phys. Chem. B	9236	2570	1863	1671	1625	1507
Macromolecules	6409	1366	1372	1261	1170	1240
Org. Lett.	5885	1252	1276	1189	1087	1081

Table 3.8: ISI Citation Report total publication articles for combined Duke, NCSU and UNC top five publication journals in the ACS research corpus

ISI Citation Reports						
JNL	Total Citations All Years	2004	2003	2002	2001	2000
Anal. Chem.	21597	891	4465	5737	5478	5026
Biochemistry	29372	1244	6066	7525	7443	7094
Inorg. Chem.	15111	747	3633	3842	3745	3144
J. Am. Chem. Soc.	75023	4254	18741	20429	16482	15117
J. Biol. Chem.	171523	8233	37324	45031	44177	36758
J. Org. Chem.	20485	1059	4956	5388	4588	4494
J. Phys. Chem.						
Macromolecules	22311	899	4517	5746	5326	5823
Science	114978	6235	23412	31885	28516	24930
Tetrahedron Lett.	21544	1243	4741	6110	5158	4292

Table 3.9: ISI Citation Report total cited articles for combined Duke, NCSU and UNC top five citation journals in the ACS research corpus. NOTE: There is no available data from ISI for J. Phys. Chem. due to a title change in 1996.

ISI data was unavailable for the Journal of Physical Chemistry because of a title change in 1996 that split this journal into the Journal of Physical Chemistry A and the Journal of Physical Chemistry B. ISI does not continue to report citation data for titles that are no longer published, as a lack of publication data eliminates the possible creation of an ISI Impact Factor for these titles.

Table 3.10 shows the ISI Impact Factor from the JCR, the calculated ACS-IF-Sim and ACS-SCO-IF data along with the percent change between the ACS-IF-Sim and ACS-SCO-IF. As with the publication and citation data, the ISI Impact Factors are not university specific.

Journal Title	ISI Impact Factor 2003	ACS_IF_ Simulation 2003	SCO-IF 2003	% Change
Acc. Chem. Res.	15	4.3691	4.2575	3%
Anal. Chem.	5.25	1.2029	0.3177	74%
Biochemistry	3.922	0.7340	0.2350	68%
Biomacromolecules	2.824	0.7486	0.3541	53%
Biotechnol. Prog.	1.488	0.2466	0.0938	62%
Chem. Mater.	4.374	1.0988	0.6548	40%
Cryst. Growth Des.	2.742	0.6755	0.2553	62%

Journal Title	ISI Impact Factor 2003	ACS_IF_ Simulation 2003	SCO-IF 2003	% Change
Environ. Sci. Technol.	3.592	0.8229	0.1288	84%
Ind. Eng. Chem. Res.	1.317	0.3241	0.0818	75%
Inorg. Chem.	3.389	1.0977	0.3596	67%
J. Agric. Food Chem.	2.102	0.6064	0.0417	93%
J. Am. Chem. Soc.	6.516	2.0998	1.2697	40%
J. Chem. Info. Comput. Sci.	3.078	0.9112	0.2245	75%
J. Comb. Chem.	4.2	1.6337	0.7733	53%
J. Med. Chem.	4.82	0.8489	0.3501	59%
J. Nat. Prod.	1.849	0.3453	0.1095	68%
J. Org. Chem.	3.297	0.9755	0.5236	46%
J. Phys. Chem. A	2.792	0.8926	0.3699	59%
J. Phys. Chem. B	3.679	1.0127	0.4804	53%
J. Proteome Res.	5.611			
Macromolecules	3.621	1.1101	0.2919	74%
Nano Lett.	5.65	1.8700	1.8700	0%
Org. Lett.	6.144	1.3176	0.8440	36%
Org. Proc. Res. Dev.	4.092	0.1688	0.0605	64%

Table 3.10: ISI Impact Factors, ACS-ISI Impact Factor Simulation and ACS research corpus Self-Citation Omission Impact Factor

4 Results and Analysis

4.1 Publication Analysis

Of the 25 journal titles analyzed for publication, faculty from all three universities published most frequently in *The Journal of the American Chemical Society*. *The Journal of Physical Chemistry B* was the only other title that occurred in all three university top five lists. At first glance, the institutional level data does not appear to be required. The ISI Journal Citation Reports (JCR) subject specific publication data, ranks *The Journal of the American Chemical Society* and *The Journal of Physical Chemistry B* first and second respectively. Although the ACS research corpus comprises only 25 publication journals, four are in the top ten publication journals in the JCR, which suggests that our corpus is well suited to this study of chemists. The other top publishers present are large publisher such as Elsevier and the Royal Society of Chemistry. The latter publishers offer title-by-title subscriptions at a higher per journal cost, or journal bundle options, at a lower per journal cost. The bundled journal options contain publisher-determined journal lists.

Of the 429 ISI designated chemistry titles falling outside the JCR top ten, selection and deselection requires more specific university data**.** For instance, in the JCR publication ranking, the Journal of Chemical Education, published by The American Chemical Society ties for 94^{th} with Carbohydrate Research, an Elsevier journal. If using LJUR data, a collection manager might discover their faculty publish more frequently in

Carbohydrate Research, for example, and rarely publish in the Journal of Chemical Education. In this instance, the collection manager could possibly deselect the Journal of Chemical Education. The JCR would not supply the critical information offered by the LJUR.

Out of the nine top five university publication titles represented in the Table 3.3, five (*Biochemistry*, *Journal of Organic Chemistry*, *Journal of Physical Chemistry A*, *Macromolecules* and *Organic Letters*) are associated with only one university. Table 3.4 represents the collation of Table 3.3 highlighting differences between the universities top five publication journal results. In some instances publication focus is distinct as in UNC publications in the Journal of Physical Chemistry A (48), where the publications by UNC chemistry faculty are more than triple that of Duke chemistry faculty (15) and almost seven times that of NCSU chemistry faculty (7). The JCR ranking for *Journal of Physical Chemistry A* is 6th for 2004. As the 6th highest publication journal, a collection manager might make the decision to select or continue purchase of this title. The data in Table 3.4 would not support this decision for NCSU.

Table 3.3 and 3.4 demonstrate the importance of university specific data as suggested by Line and Pan, “no measure of journal use other than one derived from a local use study is of any significant practical value to librarians.”(Line & Pan, 1978, p. 313) As a collection development tool, the differences between university top five publication titles might support the idea that consortia agreements benefit all members. A group purchase of all ACS titles supplies access of the *Journal of Physical Chemistry A* to UNC and the *Journal of Organic* Chemistry, a title unique to NCSU’s top five, while maintaining the bundled rather than per journal cost. The ACS research corpus data supports the utility of the ISI publication data. The JCR publication data supplies a broad view of journal performance, while LJUR data set is a university specific evaluation tool for academic library journals. Using the JCR in lieu of LJUR as a collection management tool risks purchasing titles superfluous to faculty research.

The LJUR provides university faculty publication and citation data as separate database queries used in conjunction. Despite the strengths of university publication data cited here, reviewing publication and citation data jointly produces a more accurate picture of university journal use. The increasing electronic access to serials, separate purchase options for access to archival and current journal issues, title-by-title and bundled purchasing options, the increasing number of journals, changes in research focus and cross-disciplinary research, necessitates more specific data from the ISI LJUR citation data. Supplying the article publication year coupled with journal citation counts is insufficient for effective journal collection evaluation.

4.2 Citation Analysis

In their citation study of *The Journal of the American Chemical Society*, Gross and Gross indicate a need for separate citations based on year range. (Gross,P.L.K. 1927) Their logic still holds true today,

> It must be realized that a periodical which has been in existence for only ten years, having let us say, but half as many references as one which has been published continuously for fifty years would be more desirable, dollar for dollar invested, than the latter, assuming the cost per year to be comparable in the two cases. It is also possible that a journal may have been of such quality for a long period of years that it is now little used and that in later years its quality may have improved or the nature of its material changed in such a way that is now a very valuable journal. The reverse change is even easier to imagine. (p. 387)

The UNC data listed in Table 4.1 is similar in format to the LJUR data from ISI; however the trends in rising journal cost, the options of print and electronic formats and a necessity to also provide journal titles to support faculty research demand that librarians have additional information in order to consider citation data as a tool for selection and deselection of journals.

UNC						
JNL	**Total Citations All Years**	**2004**	**2003**	**2002**	**2001**	**2000**
Anal. Chem.	733	143	191	90	137	172
Inorg. Chem.	754	124	129	115	194	192
J. Am. Chem. Soc.	2888	572	600	519	681	516
J. Phys. Chem.	538	102	98	141	103	94
Macromolecules	580	116	118	104	178	64

ISI Citation Reports						
JNL	**Total Citations All Years**	**2004**	**2003**	**2002**	**2001**	**2000**
Anal. Chem.	21597	891	4465	5737	5478	5026
Inorg. Chem.	15111	747	3633	3842	3745	3144
J. Am. Chem. Soc.	75023	4254	18741	20429	16482	15117
J. Phys. Chem.						
Macromolecules	22311	899	4517	5746	5326	5823

Table 4.1: UNC top five citation journals by year. ISI Citation Reports citation data for these same top five journals.

Performing this additional date query on the citation data allows the collection manager to know if the citation journals are covered by a “current” subscription to electronic access or if the citation falls within a “backfile” subset of the entire journals publication history.

This point deserves a more thorough explanation. Consider a choice between a subscription to an electronic version that covers articles in a journal, from 2000 through the present and another subscription that covers from 1997 through the present. While this access may cover the bulk of current faculty research needs, there is likely additional access available from the publisher with coverage since the journals inception through the beginning year of the "current" subscription (i.e from 1890 through 1997). In Table 4.2, the number under each year refers to the total number of citations made by an institutions' faculty during that year. In my data, there is an additional table for each title that includes the citation years. For example the most frequently cited title by Duke, NCSU, and UNC is *The Journal of the American Chemical Society*, which is cited 492, 218, and 572 times respectively during 2004. In its current format, the ISI data would be as follows:

University	Total Citations All Years	2004	2003	2002	2001	2000
DUKE	**1634**	492	305	217	333	287
NCSU	**1346**	218	411	205	315	197
UNC	**2888**	572	600	519	681	516

Table 4.2: ACS research corpus citation data by university for the Journal of the American Chemical Society.

Figure 4.1 displays the recommended format for the inclusion of the additional information gathered for citation years. Using 1995-present as the "current" subscription range (above the arrow), you can see that 377 citations were made to *The Journal of the American Chemical Society* to articles included in the "current" subscription timeframe.

Journal	**2004**
J. Am. Chem. Soc.	572

Citation Year	Number of Citations
2004	12
2003	53
2002	64
2001	56
2000	46
1999	34
1998	27
1997	32
1996	25
1995	28
1994	2
1993	25
1992	13
1991	13
1990	17
1989	12
1988	8
1987	13
1986	4
1931-1985	134

Figure 4.1: ACS research corpus Duke Journal of the American Chemical Society journal citation data with additional data for citation year.

In addition to “current” subscriptions, many publishers offer electronic archives. Archives have journal or publisher specific titles such as “ACS Legacy Archives,”[2] “Elsevier Backfiles,”[3] or the Institute of Physics “Historic Archive;”[4] however almost all refer to electronic access available beyond the current subscription. For the *Journal of the American Chemical Society*, the “legacy” files contain articles from 1879-1995. In Figure 4.1, the total number of citations that fall within this date range is 195 (below the arrow). While this indicates fewer citations per year than during the 1995-present “current” subscription, it reveals the degree to which researchers use portions of the current subscription and their need for the electronic archives of this journal. This analysis enables a collection manager to map citations directly to the purchasing options available at any given time. By controlling this data, the collection manager can respond quickly to changing publishing trends.

Most chemistry-centric institutions will have archival print copies of *The Journal of the American Chemical Society*; however, this will not be the case for every journal cited by faculty and the additional information of citation year, allows a, collection manager to make educated decisions based on the changing citation information.

The JCR cited journal table provides the cited year statistics covered in Figure 4.1 for all citations within a journal, but to this researcher's knowledge, there is no corresponding LJUR data. Though a collection manager can access citation journal-citation year combinations for publication specific citation, an institution specific version is unavailable. I contacted Thomson Scientific's ISI division on several occasions during the course of my research, and received no response on this matter.

In February 2004, The Chronicle of Higher Education reported the "mass resignation" of the editorial board of the Journal of Algorithms, "a scholarly journal popular with computer scientists and mathematicians." (Read, 2004) According to the article, the editorial board left to begin a new publication. While this is extreme, it exemplifies the changing face of scholarly publication. Changes in editorial boards, electronic journal and backfile availability, new and emerging journal titles, title-by-title versus bundled purchases, and trends toward faculty multidisciplinary publishing; suggest an ever changing publication and citation climate for collection managers. The citation data format found in the LJUR and ISI Citation Reports no longer provides the level of detail required to match faculty usage patterns with the wide array of choices available to collection managers.

4.3 Impact Factor Analysis

4.3.1 ISI Impact Factor and ACS Impact Factor Simulation Analysis

To begin analyzing the ACS research corpus simulation of the ISI Impact Factor (ACS-IF-Sim), I collected Impact Factor statistics from the JCR licensed to the University of North Carolina at Chapel Hill.

Journal Title	ISI Impact Factor 2003	ACS_IF_ Simulation 2003
Acc. Chem. Res.	15	4.3691
Anal. Chem.	5.25	1.2029
Biochemistry	3.922	0.7340
Biomacromolecules	2.824	0.7486
Biotechnol. Prog.	1.488	0.2466
Chem. Mater.	4.374	1.0988
Cryst. Growth Des.	2.742	0.6755
Environ. Sci. Technol.	3.592	0.8229
Ind. Eng. Chem. Res.	1.317	0.3241
Inorg. Chem.	3.389	1.0977
J. Agric. Food Chem.	2.102	0.6064
J. Am. Chem. Soc.	6.516	2.0998
J. Chem. Info. Comput. Sci.	3.078	0.9112
J. Comb. Chem.	4.2	1.6337
J. Med. Chem.	4.82	0.8489
J. Nat. Prod.	1.849	0.3453
J. Org. Chem.	3.297	0.9755
J. Phys. Chem. A	2.792	0.8926
J. Phys. Chem. B	3.679	1.0127
J. Proteome Res.	5.611	*
Macromolecules	3.621	1.1101
Nano Lett.	5.65	1.8700
Org. Lett.	6.144	1.3176
Org. Proc. Res. Dev.	4.092	0.1688

Table 4.3: ISI Impact Factors, ACS-ISI Impact Factor Simulation and ACS research corpus Self-Citation Omission Impact Factor . *Due to the absence of data in the ACS research corpus no ACS-ISI Impact Factor Simulation is available for this title.

Due to the much smaller corpus of journals available in the ACS research corpus these numbers allow for possible correlations only. The ISI Impact Factor and ACS-IF-Sim for 2003 are listed in Table 4.3.

The ACS-IF-Sim are between 61.10% and 95.88% lower than ISI-IF. Several factors could contribute to this difference. ISI contains a larger set of data than is included in the ACS research corpus. The ACS research corpus has a slightly higher publication count due to the inclusion of editorials, letters and additional publication materials; however, the number of citations in the ACS-ISI Impact Factor considers only 27 journals, whereas the ISI Impact Factor uses approximately 5,900 journals (ISI JCR website). The 27 journals in the ACS capture 29% , of the ISI impact factor, which suggests that ACS journals play a central role in chemistry research.

The ACS corpus has two additional limitations: (1) the American Chemical Society did not provide every articles within the 2000-2004 time frame and (2) pre-processing errors as described in the Pre-Processing section of the methodology.

4.3.2 ACS Impact Factor Simulation and ACS Self-Citation Omission Impact Factor Analysis

One way that a journal editor can increase the journal's ISI Impact Factor is to require authors to make self-citations to their journal (Cameron, 2005) Table 4.8 includes the percent change between the self-citation inclusive ACS ISI Impact Factor Simulation (ACS-IF-Sim) and the Self-Citation Omission Impact Factor (ACS-SCO-IF) that excludes self-citation in its calculation. These percentages range from 3% for Accounts of Chemical Research to 93% for Journal of Agriculture and Food Chemistry. In my original hypothesis, I suggested self-citation omission would adjust for publisher manipulation of the Impact Factor. If a journal requires a specific number of self-citations, it is possible to increase their impact factor. The small scope of the ACS research corpus does not allow us to effectively evaluate this assertion. It is possible that the Journal of Agriculture and Food Chemistry covers a much narrower field of study than that of Accounts of Chemical Research. If this is the case, self-citation may be inevitable due to a lack of relevant journals in which to publish.

In Table 4.8 I estimated the impact on ISI Impact Factors if self-citations were prevalent to the same degree in the ACS corpus.

Journal Title	ISI Impact Factor 2003	Adjusted ISI IF with % Change	ACS_IF_Simulation 2003	SCO-IF 2003	% Change
Acc. Chem. Res.	15	**14.617**	4.3691	4.2575	3%
Anal. Chem.	5.25	**1.387**	1.2029	0.3177	74%
Biochemistry	3.922	**1.256**	0.7340	0.2350	68%
Biomacromolecules	2.824	**1.336**	0.7486	0.3541	53%
Biotechnol. Prog.	1.488	**0.566**	0.2466	0.0938	62%
Chem. Mater.	4.374	**2.607**	1.0988	0.6548	40%
Cryst. Growth Des.	2.742	**1.036**	0.6755	0.2553	62%
Environ. Sci. Technol.	3.592	**0.562**	0.8229	0.1288	84%
Ind. Eng. Chem. Res.	1.317	**0.333**	0.3241	0.0818	75%
Inorg. Chem.	3.389	**1.110**	1.0977	0.3596	67%
J. Agric. Food Chem.	2.102	**0.145**	0.6064	0.0417	93%
J. Am. Chem. Soc.	6.516	**3.940**	2.0998	1.2697	40%
J. Chem. Info. Comput. Sci.	3.078	**0.758**	0.9112	0.2245	75%
J. Comb. Chem.	4.2	**1.988**	1.6337	0.7733	53%
J. Med. Chem.	4.82	**1.988**	0.8489	0.3501	59%
J. Nat. Prod.	1.849	**0.586**	0.3453	0.1095	68%
J. Org. Chem.	3.297	**1.770**	0.9755	0.5236	46%
J. Phys. Chem. A	2.792	**1.157**	0.8926	0.3699	59%
J. Phys. Chem. B	3.679	**1.745**	1.0127	0.4804	53%
J. Proteome Res.	5.611	**5.611**			
Macromolecules	3.621	**0.952**	1.1101	0.2919	74%
Nano Lett.	5.65	**5.650**	1.8700	1.8700	0%
Org. Lett.	6.144	**3.935**	1.3176	0.8440	36%
Org. Proc. Res. Dev.	4.092	**1.467**	0.1688	0.0605	64%

Table 4.4: ISI Impact Factor, Adjusted ISI Impact Factor reflecting the % change calculated from ACS-IF-Sim and the ACS-SCO-IF *Due to the absence of data in the ACS research corpus no ACS-ISI Impact Factor Simulation is available for this title.

While the percent change had a minor effect on some titles, such as Accounts of Chemical Research, other titles were affected much more acutely such as Macromolecules.

Though ISI Citation Reports provide graphical and numerical data on the number of self-citations in a journal in a given year, they do not supply a corresponding Impact Factor for comparison. This lack of information makes it difficult to establish the validity of the ranked lists provided by ISI. The addition of the ACS-SCO-IF from ISI would allow the collection manager to compare ranked lists in analyzing for selection or deselection of journals.

More important than the elimination of self-citations and recommended for future research is the absence of university specific Impact Factors. Determining the criteria for such a factor is complex, but with university specific citation and publication data, a university-specific Impact Factor would be comparable to the isi Impact Factor. While not completely replacing the need for citation and publication reports such as the LJUR, a university specific ISI Impact Factor created from university specific publication and citation data, would allow collection managers to make decisions based on data resulting directly from their faculty's research.

5 Recommendations

Collection managers in an academic setting are often forced to make difficult decisions regarding the selection and deselection of journal titles with decreasing budgets and increasing costs. Although collection managers already use publication, citation and impact factors as one of the many tools to aid in their journal subscription decisions, the specificity of the data does not map directly with the available purchasing choices. When combined with other data, these tools are relatively effective; however, data available to ISI could be compiled to provide additional or improved university specific tools, such as the citation date specific tool in Figure 4.1 or an LJUR impact factor.

Though the titles in the ACS research corpus are only available through one publisher – the ACS, new publication mechanisms and new ways to bundle packages of journals are becoming increasingly prevalent. Determining the value of an individual title is paramount to deciding which of the purchasing choices will be most cost effective. For example, in Table 3.4 NCSU faculty cited the Journal of Physical Chemistry A only seven times in the ACS research corpus. When considering a bundle of ACS titles, NCSU might choose to purchase on a title by title basis for only those ACS titles supporting faculty research as evidenced through LJUR publication and the proposed LJUR citation update. Thus although the cost per journal titles is more, they can redirect finances to titles that are more likely support faculty research.

Finally, libraries require university specific data such as the LJUR rather than the generic citation and publication data available through JCR. LJUR provides data collected on research faculty, the academic library's primary patron. Expanding citation data to include date and a university specific Impact Factor are just two examples of how collection managers can better map citation data to the purchasing options that they have available

In this paper, I have considered three data sources publication, citation, and impact-factors. Although beyond the scope of this paper, I advocate a multi-faceted approach to collection management that may include faculty and student interviews, and usage statistics as data sources to feed into an informed decision. (Brown, E.A.R., 2005)

6 Conclusions

This analysis suggests that changes to both the ISI citation data and the ISI Impact Formula would better support the needs of academic library collection mangers. Additional university specific data aids informed decision making for the selection and deselection of journal titles in academic libraries by informing package versus title by title purchase, the need for archival material specifically addressing faculty needs. With the increasing journal costs, decreasing budgets and journal title bundling trend by vendors; more detailed data is imperative.

The current format for citation reports found in the university specific Local Utilization Journal Reports (LJUR) offered for purchase by ISI acts as an aid to many university collection managers. Following the same format as the publication piece of this product is insufficient to address the citation year for the changing needs of collection managers and academic libraries in addressing publishing trends. Adding the date cited to this data provides crucial information when determining to purchase electronic backfiles.

Journal titles such as Science, which according to OCLC's WorldCat is owned by 4593 "Libraries Worldwide," are often part of a bound volume collection in an academic institution; however with the changing multi-disciplinary research trends, journals previously considered unimportant to the mission of the library have suddenly become necessary for supporting new and developing programs. In cases such as this, libraries may need to purchase electronic backfiles and adding the citation date to the existing

LJUR citation data would allow collection managers to determine efficiently if faculty citation journals are covered by their "current" subscription or if additional backfiles should be purchased.

Providing a Self-Citation Omission Impact Factor in conjunction with the ISI Impact Factor would allow collection managers to compare impact factors and make strategic decisions regarding their journal trends. The comparison of these two factors would also promote insight into the availability of journals in a specific field. If for instance a journal's Impact Factor is significantly reduced by the subtraction of self-citations, a collection manager may take this as an indicator of specialized content scope and through additional faculty inquiry determine the journal's impact to their collection.

Creating a university specific Impact Factor based on the publication and citation data currently available through the purchase of the LJUR would further aid a collection manager in prioritizing journal selections. Although a good collection manager is aware of journals in their specialty, a localized Impact Factor would aid in their identification of new and emerging fields more efficiently than citation tools currently available. It is difficult for the academic collection manager to be intimately knowledgeable about every faculty member's research area. A localized Impact Factor would bring unrecognized research support titles to the forefront.

Combining the LJUR publication data, the recommended changes to the LJUR citation data, the creation of a self-citation omission Impact Factor and adding a localized Impact Factor would aid the collection manager in the decisive determination of journal selection and deselection. While an ongoing relationship with faculty is essential to an effective collection, the tools described herein, would facilitate efficient identification of collection gaps, data on current or archival faculty needs, aid in budget control through a vest value evaluation of title by title or package purchases, while also accelerating the recognition of emerging faculty research needs.

7 Notes

[1] For most titles.
[2] ACS Legacy Archives, http://pubs.acs.org/rates/institutions/options.html. Last accessed February 20, 2006.
[3] Elsevier ScienceDirect, http://info.sciencedirect.com/content/journals/backfiles/. Last accessed February 20, 2006.
[4] Institute of Physics, http://ej.iop.org/pdf/historic_archive.pdf. Last accessed February 20, 2006.

8 References

Bauer, K., & Bakkalbasi, N. (2005). An examination of citation counts in a new scholarly communication environment [computer file]. *D-Lib Magazine, 11*(9), 1.

Bollen, J. et.al. (2005). Toward alternative metrics of journal impact: A comparison of download and citation data. *Information Processing and Management*. 41, 1420.

Black, S. (2001). Using citation analysis to pursue a core collection of journals for communication disorders. *Library Resources & Technical Services, 45*(1), 3-9.

Brown, E.A.R. (2005) *Citation Analysis: Literature review and simple analysis of chemistry faculty publications and citations at three research universities* [Unpublished Manuscript]

Brown, K. C. (2002). How many copies are enough? using citation studies to limit journal holdings. *Law Library Journal, 94*(2), 301-314.

Calhoun, J. C. (1995). Serials citations and holdings correlation. *Library Resources & Technical Services, 39*, 53-77.

Cameron, B. D. (2005). Trends in the usage of ISI bibliometric data: Uses, abuses, and implications. *Portal, 5*(1), 105-125.

Davis, P. M. (2002). Where to spend our E-journal money? defining a university library's core collection through citation analysis. *Portal, 2*(1), 155-166.

Edwards, S. (1999). Citation analysis as a collection development tool: A bibliometric study of polymer science theses and dissertations. *Serials Review, 25*(1), 11-20.

Fuseler-McDowell, Elizabeth. (1990). Collection evaluation and development using citation analysis techniques. *IAMSLIC at a crossroads.* International Assn. of Marine Science Libs. & Information Centers. (99-108)

Garfield, E. (1955). Citation indexes for science. *Science, 122*(3159), 108-111.

Garfield, E. (1994). The ISI Impact Factor. Thomson Scientific, http://scientific.thomson.com/free/essays/journalcitationreports/impactfactor. Retreived March 25, 2006.

Garfield, E. (2005). The agony and the ecstasy-The history and meaning of the journal impact factor. International Congress on Peer Review and Biomedical Publication. Chicago: 2005. .Retrieved March 20, 2006 from http://www.garfield.library.upenn.edu/papers/jifchicago2005.pdf.

Gooden, A. M. (2001). Citation analysis of chemistry doctoral dissertations: An ohio state university case study {computer file}. *Issues in Science & Technology Librarianship,*

Gross, P. L. K., & Gross, E. M. (1927). College libraries and chemical education. *Science, 66*(1713), 385-389.

Hoeffel, C. “Journal impact factors” [letter]. *Allergy* 53(12): 1225 (December, 1998).

Journal Citation Reports. [University of North Carolina at Chapel Hill online subscription.] Available from http://scientific.thomson.com/products. Retrieved March 31, 2006.

LaBonte, K. B. (2005). Citation analysis: A method for collection development for a rapidly developing field [computer file]. *Issues in Science & Technology Librarianship*, (43), 1.

Line, M. B., & Pan, E. (1978). Rank lists based on citations and library uses as indicators of journal usage in individual libraries. *Collection Management, 2*(4), 313-316.

OCLC WorldCat. [University of North Carolina at Chapel Hill online subscription.] Available from http://www.oclc.org/worldcat/

Thomson Scientific, Journal Citation Reports, http://thomsonscientific.com/products/jcr/. Retreived April 4, 2006.

Thomson Scientific, Research Services, http://scientific.thomson.com/products/rsg/. Retrieved March 31, 2006.

Thomson Scientific, Local Journal Utilization Report, http://scientific.thomson.com/products/ljur/. Retreived March 31, 2006.

Vaughan, L., & Shaw, D. (2005). Web citation data for impact assessment: A comparison of four science discipline. *Journal of the American Society for Information Science and Technology, 56*(10), 1075-1087.

Vinkler, P. (2004). Characterization of the impact of sets of scientific papers: The garfield (impact) factor. *Journal of the American Society for Information Science and Technology, 55*(5), 431-435.

Walcott, R. (1994). Local citation studies--a shortcut to local knowledge. *Science & Technology Libraries, 14*, 1-14.

Wible, J. G. (1990). Comparative analysis of citation studies, swept use, and ISI's impact factors as tools for journal deselection. *IAMSLIC at a crossroads. international assn. of marine science libs.* (pp. 109-116)& Information Centers; International Assn. of Marine Science Libs. & Information Centers.