Accuracy of References in Five Biomedical Informatics Journals

DOMINIK ARONSKY, MD, PhD, JOEL RANSOM, MBA, MPH, KEVIN ROBINSON, BS

Abstract Objective: To determine the rate and type of errors in biomedical informatics journal article references.

Methods: References in articles from the first 2004 issues of five biomedical informatics journals, Journal of the American Medical Informatics Association, Journal of Biomedical Informatics, International Journal of Medical Informatics, Methods of Information in Medicine, and Artificial Intelligence in Medicine were compared with MEDLINE for journal, authors, title, year, volume, and page number accuracy. If discrepancies were identified, the reference was compared with the original publication. Two reviewers independently evaluated each reference.

Results: The five journal issues contained 37 articles. Among the 656 eligible references, 225 (34.3%) included at least one error. Among the 225 references, 311 errors were identified. One or more errors were found in the bibliography of 31 (84%) of the 37 articles. The reference error rates by journal ranged from 22.1% to 40.7%. Most errors (39.0%) occurred in the author element, followed by the journal (31.2%), title (17.7%), page (7.4%), year (3.5%), and volume (1.3%) information.

Conclusion: The study identified a considerable error rate in the references of five biomedical informatics journals. Authors are responsible for the accuracy of references and should more carefully check them, possibly using informatics-based assistance.


Introduction

Authors of publications in scientific journals are expected to provide the readership with references to related and previously published work. Most medical journals follow the “Uniform Requirements for Manuscripts Submitted to Biomedical Journals,” which follow the Vancouver style for formatting references. The requirements specify that authors are expected to check the accuracy of references against the original article. Accurate references allow interested readers to easily locate additional publications that are relevant to the subject of a specific journal article. Unless an error occurs in a critical element of the reference, such as journal, title, year of publication, volume, or first page number, most errors would not hinder their retrieval. However, if references are cited incorrectly, their location may become time-consuming, frustrating, or even impossible. Thus, the accuracy of references is critical.

Previous investigators have reported on reference accuracy in various general and specialist biomedical journals and found error rates ranging from 3% to 60%. High error rates (4.1%–40.3%) were even found among leading biomedical journals that check references prior to publication. Most errors affect the names of authors, followed by errors in the title element of a reference. The reference accuracy in biomedical informatics journals, the aim of this study, has not been previously examined.

Materials and Methods

The references in articles of the first standard, nonspecial 2004 issues of five biomedical informatics journals were selected: Journal of the American Medical Informatics Association (J Am Med Inform Assoc), Journal of Biomedical Informatics (J Biomed Inform), International Journal of Medical Informatics (Int J Med Inform), Methods of Information in Medicine (Methods Inf Med), and Artificial Intelligence in Medicine (Artif Intell Med).

Adapting the methodology of previous studies, we included all references to published journal articles that were indexed in MEDLINE, such as research articles, reviews, editorials, and letters to the editor. Each reference listed in an article’s bibliography was checked against its corresponding MEDLINE citation. If a reference did not exactly match the MEDLINE citation, it was compared with the actual full-text original publication that was referenced. We excluded references to books, proceedings (even if indexed in MEDLINE), abstracts, technical reports, dissertations, theses, Web sites, public media, manuals, or other types of reports, for which the original publication may not be easily accessible for verifying the accuracy of the reference.

We defined a reference to a publication as incorrect if one or more errors were identified in any of the following six elements: journal, authors, title, year, volume, and page numbers. We defined a journal error as the full journal name or abbreviation differing from the official MEDLINE version, which follows the style used in the List of Journals Indexed in...
Index Medicus.\textsuperscript{28} We further classified inaccurate author information by errors involving the author’s last name, the author’s first and middle initials, and whether the error involved the first or subsequent authors. We also noted missing, additional, or an incorrect order of authors. The “Uniform Requirements for Manuscripts Submitted to Biomedical Journals” recommends an abbreviated list of authors using et al. after the sixth author.\textsuperscript{1,29} Some biomedical journals, however, use et al. after the third author. In this study, a missing author error was noted if et al. was used after the first or second author. For references using et al. after the third author, we verified author information for listed authors. An error in the title element was present if the title included misspelled, changed, omitted, or additional text. Page numbers were checked for the accuracy of the first and last page referenced. In addition, we noted punctuation errors in any part of the reference and the listing of duplicate references within the bibliography of an article.

Each included reference was independently checked by two reviewers. Disagreements among the two reviewers were resolved by consensus involving the three participating reviewers. We used descriptive statistics for the overall error rate and the rate in each error category. Punctuation errors and duplicate references were not included in the calculation of the overall error rates.

Results
The five journal issues included 37 publications with 1,356 references, of which 656 (48.4%) met the inclusion criteria. We identified 225 (34.3%) references that contained a total of 311 errors in the journal, authors, title, year, volume, or page number element of a reference. A single error occurred in 170 (75.6%) and multiple errors in 55 (24.4%) references (two errors in 51 and three errors in four references). Six (16%) of the 37 articles had no errors as defined in the study. Table 1 summarizes the type and number of errors. We found 50 punctuation errors in any part of the 656 study references and one duplicate reference; neither error type was included in the overall error rate.

The author element was the most frequent error category and contained 121 (39.0%) errors in 99 references. An error in the first author’s name was found 39 times (eight errors in the last name and 31 in the first or middle initial). An error in the list of coauthors was found 52 times (16 errors in the last name and 36 in the first or middle initial). A missing or additional author or an incorrect order of authors was found 30 times. Inaccurate journal titles accounted for 97 (31.2%) errors. An incorrect journal abbreviation was listed 93 times, and an incorrect journal name appeared four times. One reference listed Vis Graphics Image Procest instead of Comput Med Imaging Graph, one listed Int J Antimicrob Chemother instead of Int J Antimicrob Agents, one listed a reference to a submitted manuscript, and one mixed the authors and the title of a book with the journal, year, and volume of a journal issue. The title element included 55 (17.7%) errors, the year element 11 (3.5%), and the volume element four (1.3%). In 18 references, an inaccurate page listing accounted for 23 (7.4%) errors (four in the first page, nine in the last page, and five in both pages).

Discussion
In this study of five biomedical informatics journals, we found error rates between 22.1% and 40.7%. These error rates are within the range of reference accuracy rates found in other biomedical journals.\textsuperscript{2–26} We were, however, surprised by the high error rate, particularly because authors publishing in biomedical informatics journals are expected to be more familiar with and to more frequently use computerized tools that support the retrieval, management, and preparation of references and bibliographies. Authors publishing in the field of biomedical informatics should realize the importance of correct references. However, a previous study that examined the error rate in the library literature, in which lower error rates could also be expected due to author familiarity with bibliographies and references, did not identify a lower overall reference error rate between biomedical and library publications.\textsuperscript{5}

Our study identified a large number of errors in the abbreviation of journal names. For readers who are unfamiliar with the biomedical informatics field, it is more difficult to identify the correct journal. For example, at the time of this study in summer 2004, a MEDLINE query for “Int J Med Inform [jour]” (International Journal of Medical Informatics) returned no results. At that time, a look up in MEDLINE’s journal database returned the applicable MEDLINE abbreviation Int J Med Inf. The official abbreviation for “informatics” is “inform,” which was corrected in MEDLINE’s electronic journal database after the completion of this study.

Table 1 ■ Reference Errors in Five Biomedical Informatics Journals

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total articles</td>
<td>10</td>
<td>6</td>
<td>7</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>Total references</td>
<td>141</td>
<td>130</td>
<td>223</td>
<td>131</td>
<td>75</td>
</tr>
<tr>
<td>Included references</td>
<td>181</td>
<td>175</td>
<td>145</td>
<td>109</td>
<td>46</td>
</tr>
<tr>
<td>References with errors</td>
<td>40</td>
<td>70</td>
<td>59</td>
<td>38</td>
<td>18</td>
</tr>
<tr>
<td>Reference error rate (%)</td>
<td>22.1</td>
<td>40.0</td>
<td>40.7</td>
<td>34.9</td>
<td>39</td>
</tr>
<tr>
<td>Total errors</td>
<td>56</td>
<td>105</td>
<td>73</td>
<td>52</td>
<td>25</td>
</tr>
<tr>
<td>Journal title errors</td>
<td>11</td>
<td>19</td>
<td>43</td>
<td>16</td>
<td>8</td>
</tr>
<tr>
<td>Authors errors</td>
<td>27</td>
<td>49</td>
<td>19</td>
<td>21</td>
<td>5</td>
</tr>
<tr>
<td>Title errors</td>
<td>12</td>
<td>22</td>
<td>7</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Year errors</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Volume errors</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Page number errors</td>
<td>4</td>
<td>9</td>
<td>1</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
Errors in the author’s name may only demonstrate a lack of attention to detail. However, the names of authors are not only critical for retrieving articles but are also relevant when author citations are used to measure research productivity. For example, the Science Citation Index® from Thomson ISI allows users to examine the citation frequency of an author’s publications. Misspellings and variations in an author’s name may prevent the program from retrieving all relevant citations and influence bibliometric measures and information retrieval approaches.

Many journals, including four considered in this study, publish papers in electronic format and provide readers with a direct link to the abstract or the original publication of the cited references. If errors exist in references, successful linking does not occur in online publications, Furthermore, unsuccessful linking increases the visibility of errors to the readers.

Our study examined only one issue of each journal, and it is possible that other issues contain fewer errors in references. In addition, our study did not examine MEDLINE’s reference error rate, which has been reported to be 1.5% in a study analyzing publications from 1976 to 1998. We would have missed errors if authors incorporated references from MEDLINE that were incorrectly indexed. Such an error is present in reference 20 of the present article, in which MEDLINE lists the third author’s name as “Kee WD,” whereas the original article lists the correct name of the author as “Ngan Kee WD” (same author as reference 26). This study included only references to journal articles that were indexed in MEDLINE; error rates for excluded references may differ from our results and need to be examined in a separate study.

Proposed strategies to minimize error rates from an author’s perspective include the use of reference management software and careful verification with the original publications. Possible editorial office strategies include manual or automated in-house reference checking, asking reviewers to check references, requesting the submission of the first page of a cited reference at the time of manuscript submission, or limiting the number of references.

**Conclusion**

Many references in biomedical informatics journals include errors. Inaccurate references are frustrating and may reflect poorly on the biomedical informatics field. Authors should review and validate the accuracy of references included in submitted manuscripts through verification with the original article.

**References**


