

Major bibliographic errors in PubMed: personal experience among 240 publications and proposed remediation process for errors

Mitchell S. Cappell, MD, PhD

See end of article for author's affiliation.

DOI: <http://dx.doi.org/10.3163/1536-5050.104.1.015>

Although substantial data exist on errors in bibliographic citations in journal articles [1], literature review reveals no data on errors in bibliographic citations in PubMed. Yet bibliographic errors in PubMed are much more important than errors in individual journals. PubMed is a ubiquitous literature search engine. Errors in PubMed could render articles inaccessible to researchers or clinicians performing computerized literature searches.

Despite major contributions to academic medicine, the National Library of Medicine (NLM), like any human institution, is subject to errors, including errors in PubMed. Major errors are herein defined as incorrectly spelling authors' names, deleting authors' names, making mistakes in keywords in article titles, or making mistakes in journal volume, issue, or page numbers. Review of my 240 articles cited in PubMed, 1982–2015, revealed 3 major errors (1.25% rate). The 3 errors comprised misspelling of surnames of authors in 2 cases and deletion of all authors in 1 case [2, 3].

When called about this last error, the PubMed representative said the error was in a reference listed as “in process” and would be corrected during a systematic reference verification. The verified (corrected) reference would then be listed without the “in process” designation. However, this reference verification process can require considerable time. One reference for an article I coauthored was still not verified more than one year after the initial listing of the publication [4].

The following suggestions may be useful. First, PubMed should warn during every literature search that references listed as “PubMed—in process” are preliminary and subject to errors. Second, if contacted about an error, PubMed should expeditiously mark this reference as “reference accuracy is contested”. Third, PubMed should

commit to correcting any error less than four weeks after notification.

This work received an exemption/approval from the Institutional Review Board of William Beaumont Hospital on May 7, 2015.

REFERENCES

1. Hansen ME, McIntire DD. Reference citations in radiology: accuracy and appropriateness of use in two major journals. *Am J Roentgenol (AJR)*. 1994;163:719–23. DOI: <http://dx.doi.org/10.2214/ajr.163.3.8079876>.
2. [Incorrect listing in PubMed]: Cappell MS, Mogrovejo E, Manickam P, Polidori G. Upper gastrointestinal bleeding associated with focal ulceration at the attachment base of a chronically retained endoscopic hemoclip. *Minerva Gastroenterol Dietol*. 2014 Dec;60(4):285–7. PMID: 25384807. [cited 31 Jan 2015]. (corrected 11 Mar 2015).
Correct listing as published in journal: Cappell MS, Mogrovejo E, Manickam P, Polidori G. Upper gastrointestinal bleeding associated with focal ulceration at the attachment base of a chronically retained endoscopic hemoclip. *Minerva Gastroenterol Dietol*. 2014; Dec;60(4):285–7. PMID: 25384807.
3. [Incorrect listing in PubMed: No authors listed]: Case report of working with an institutional review board to approve a study on a particularly sensitive topic in a particularly vulnerable population: investigator negotiation and compromise may be the best strategy. *Gastroenterol Nurs*. 2015 Jan–Feb;38(1):62–3. DOI: <http://dx.doi.org/10.1097/SGA.000000000000090>. PMID: 25636014. [cited 10 Feb 2015]. (corrected 13 Mar 2015).
Correct listing as published in journal: Cappell MS. Case report of working with an institutional review board to approve a study on a particularly sensitive topic in a particularly vulnerable population: investigator negotiation and compromise may be the best strategy. *Gastroenterol Nurs*. 2015;Jan–Feb 38(1):62–3. DOI: <http://dx.doi.org/10.1097/SGA.000000000000090>. PMID: 25636014

4. Jaurigue MM, Cappell MS, Therapy for alcoholic liver disease [review]. *World J Gastroenterol*. 2014;Mar 7;20(9): 2143–58. DOI: <http://dx.doi.org/10.3748/wjg.v20.i9.2143>. PMID: 24605013 [PubMed—in process].

Listed as in process in PubMed when accessed March 11, 2015, more than one year after publication.

AUTHOR'S AFFILIATION

Mitchell S. Cappell, MD, PhD, mscappell@yahoo.com, Chief of Gastroenterology & Hepatology, William Beaumont Hospital, Oakland University William Beaumont School of Medicine, 3535 West Thirteen Mile Road, Royal Oak, MI 48073

Editor's Note: We sent Mitchell S. Cappell's letter to the National Library of Medicine. Here is their response.

National Library of Medicine response

David Gillikin

See end of article for author's affiliation.

DOI: <http://dx.doi.org/10.3163/1536-5050.104.1.016>

In Fiscal Year 2014, the National Library of Medicine (NLM) released more than 981,000 journal citations [1]. Of that total, 94% (926,000) were submitted electronically as eXtended Markup Language (XML) citations by the publisher or a supplier designated by the publisher. The remaining 6% were created by NLM by scanning from the print copy of a journal and using optical character recognition. NLM creates a very small number of missing citations not sent via XML each year. When NLM manually creates a missing citation, we use cut-and-paste mechanisms.

Citation errors can arise most often from two sources: (1) as published incorrectly in the full-text article or (2) as introduced by either the XML supplier (which is the case in the two identified errors in the letter) or by NLM during its processing.

For the former, NLM requires the publication of an erratum notice so that the publication history is transparent for all users. That requires action on the part of the publisher. For example, if an author's name is misspelled in the full-text article, the journal must issue an erratum notice. This NLM policy is explained in our errata fact sheet [1]. NLM processed over 9,000 published errata during Fiscal Year 2014 [2]. For the latter, NLM will enter the correction in our workflow as soon as it is reported or discovered. For example, if the XML supplier incorrectly sent the last name as the forename and vice versa while the author names are correct in the full-text article, NLM

will reverse the name parts once we know about the error.

There is always room for improvement in our processing time for reported error corrections and for our quality assurance in general, and we continually revise our workflow process. We always welcome suggestions.

REFERENCES

1. National Library of Medicine (US). Key MEDLINE indicators [Internet]. Bethesda (MD): The Library; 7 Feb 2002 [updated 17 Dec 2014; cited 20 May 2015]. <http://www.nlm.nih.gov/bsd/bsd_key.html>.
2. National Library of Medicine (US). Fact sheet: errata, retractions, partial retractions, corrected and republished articles, duplicate publications, comments (including author replies), updates, patient summaries, and republished (reprinted) articles policy for MEDLINE [Internet]. Bethesda, MD: The Library; 6 Mar 2002 [updated 30 Apr 2015; cited 20 May 2015]. <<http://www.nlm.nih.gov/pubs/factsheets/errata.html>>.

AUTHOR'S AFFILIATION

David Gillikin, gillikd@mail.nlm.nih.gov, Chief, Bibliographic Services Division, National Library of Medicine, Bethesda, MD 20894-3831