ABSTRACT

Background and Objectives: References for medical articles are not always retrievable. This eliminates the ability to check on the validity of statements, methodologies, data collection, and conclusions.

Methods: References of review, scientific, and research articles published in the 2019 and 2020 Journal of the Society of Laparoscopic & Robotic Surgeons were evaluated for ability to retrieve the reference cited.

Results: Ninety-five articles with 2,424 references were evaluated. There were 1,025 (1,025/2,424 = 42.3%) open access versus 1,399 (1,399/2,424 = 47.7%) paywall articles. There were 357 (14.7%) citations unavailable (misreference) due to bad, broken, or nonexistent links and wrong article, abstract or citation only, and missing citations.

Conclusion: Loss of reference existence or retrievability is a scientific hazard. Science is self-correcting but is doomed to not knowing what was said or discovered when references are no longer available.

Key Words: Citation, Link rot, References.

INTRODUCTION

References are fundamental to scientific communication. They acknowledge and credit prior work influencing the article presented providing insight to methodologies, data, and conclusions. This gives readers information about who, where, and when the reference was made and how to retrieve the source and learn more. Assessment of reference availability of medical articles after publication where all references to the original articles were available, read, and analyzed compared to what an author said, showed that nine to 25 percent (%) of statements cited were incorrectly quoted and misrepresented as “facts.” These errors and distortions become imbedded and magnified over time and need to be accessed for what was actually said or how data was interpreted at the time of publication. This requires access to the references. Before the internet, libraries and personal subscriptions were the source for checking references. As the cost of journals increased along with the volume of articles written, digital access formats through the internet quickly diminished. Not being able to get a reference is a misreference that is unsubstantiated and the potential of misinformation. Original article statements should be compared for misstatements and misrepresentations. A statement that cannot be checked may represent the truth, bend it, or reconstruct it. It cannot be validated because it is missing. How often does this happen? This study assesses reference availability and accuracy of articles published in Journal of the Society of Laparoscopic & Robotic Surgeons (JSLS) in 2019 and 2020 post-publication.

MATERIALS and METHODS

Articles identified as “scientific papers,” “research articles”, and “reviews” as published online at https://www.ncbi.nlm.nih.gov/pmc/journals/1326/ for JSLS had their references checked for availability and citation accuracy using the author information or the hyperlink provided by PubMed, MEDLINE, and PubMed Central at the end of the reference citation for each article identified as “PubMed,” “CrossRef”, or “Google Scholar” hyperlinks.

RESULTS

For the two years 2019 and 2020 there were 95 articles with 2,424 references. Open access versus paywall articles was 42.3% versus 57.7% (1,025 versus 1,399). Fourteen-
point seven percent (14.7%) of the citations were unavailable or misreferenced (357/2,424) (Tables 1 and 2). There were 38 (1.6%) non-English references. Linking to PubMed, Google Scholar, or CrossRef gave different results; such as, free at one and behind a paywall at another and found on one and not another.

DISCUSSION

Studies of medical and specialty journal article references after publication assessing retrievability and checked for accuracy and wording interpretation from the original report ranges from 9%,1 10%,2 14.5%,3 19.7%,4 20%,5 to 25.4%6 error rate. Errors include failing to substantiate statements, wording unrelated to the statement, contradictions to the reference assertion, oversimplification, and overgeneralization of what was originally stated.5 Internet digital references that are unavailable make legitimacy of claims to that reference unknown or questionable.

A reference is used by an author to show validity and scholarship but may be no more “than the trappings of credibility”

Table 1.
Journal of the Society of Laparoscopic and Robotic Surgeons Internet Reference Findings

| Volume # 23, 2019 | # Articles Assessed | # of References | PM Free | PM behind Paywall/Paywall Cost | GS Free | GS Paywall/Paywall Cost | Mis References PM/GS |
|-------------------|---------------------|------------------|---------|--------------------------------|---------|------------------------|----------------------|
| (1)               | 20                  | 530              | 194     | 270/$ 11,242.44               | 215     | 264/$ 11,070.63        | 35/42                |
| (2)               | 14                  | 354              | 14      | 175/$ 7,245.94                | 136     | 191/$ 7,859.04         | 29/25                |
| (3)               | 7                   | 229              | 60      | 130/$ 4,973.24                | 61      | 138/$ 5,340.19         | 28/29                |
| (4)               | 9                   | 242              | 75      | 145/$ 6,022.12                | 76      | 153/$ 6,455.83         | 18/8                 |
| total             | 50                  | 1,355            | 463     | 720/$ 29,483.74               | 488     | 746/$ 30,685.69        | 110/104              |

| Volume # 24, 2020 | # Articles Assessed | # of References | PM Free | PM behind Paywall/Paywall Cost | GS Free | GS Paywall/Paywall Cost | Mis References PM/GS |
|-------------------|---------------------|------------------|---------|--------------------------------|---------|------------------------|----------------------|
| (1)               | 8                   | 264              | 86      | 137/$ 5,513.29                | 84      | 147/$ 5,871.99         | 17/25                |
| (2)               | 11                  | 244              | 83      | 126/$ 5,150.50                | 81      | 145/$ 6,016.80         | 12/11                |
| (3)               | 10                  | 227              | 93      | 95/$ 3,967.99                 | 97      | 98/$ 3,919.99          | 19/21                |
| (4)               | 16                  | 334              | 146     | 149/$ 5,911.35                | 164     | 98/$ 5,726.00          | 20/18                |
| total             | 45                  | 1,069            | 408     | 507/$ 20,543.13               | 426     | 494/$ 21,534.78        | 68/75                |

PM, PubMed; GS, Google Scholar.

Table 2.
Most Common Reasons for Hypertext Transfer Protocol Error Codes and Unsuccessful Universal Resource Locator Access Attempts

| Error Code          | Description                                                                 |
|---------------------|-----------------------------------------------------------------------------|
| 404 Not found       | Internal link not found                                                     |
| 403 Forbidden       | Internal server error                                                       |
| 400 Bad request     | Link not found                                                              |
| 401 Unauthorized    | Page not found                                                              |
| 410 Gone            | Site cannot be reached                                                      |
| 500 Internal server error | Sorry, we can’t find the article                                          |
| 401 Unauthorized    | The link you requested no longer exists                                     |
| An error has occurred | This account has been suspended                                             |
| Bad gateway         | This site can’t be reached                                                  |
| Cannot find server  | Timeout                                                                      |
| Cannot obtain       | Wrong link                                                                   |
| Content not found   | Wrong site                                                                   |

with sources being misquoted, not retrievable, inappropriate, “unreliable and occasionally even imaginary.”7 References are the lingua franca of science communication. For successful scientific communication references must be accessible. Without reference availability retracing methodologies, data, conclusions, and replicate or advance findings
cannot be evaluated. Inability to retrieve a reference makes statements unsupported and unverified, i.e., a misreference.

One of the tasks of a peer reviewer evaluation is assessing the accuracy of reference statements and if important references are missing. "The peer reviewer ensures that references are used appropriately, cited accurately, formatted correctly and that none are missing." It is unknown how often statements, quotations, and inferences are checked during peer review against the submitting authors references and what was in the original article, but it is felt by editor-in-chiefs that it is infrequently to rarely done, otherwise quotation and statement inaccuracies would be corrected and bad or missing links to articles would be called out prior to publication. With a documented 9% – 25% misquote/misstatement rate this is not the case.1-6 Reviews of articles after publication evaluating for quotation accuracy and statement paraphrasing, strongly suggests that either statements are not checked or were misunderstood by the reviewer.1-6 This represents incorrect or false information to readers as “fact.” Authors give reference citation information as part of their manuscript, listing the names(s) of reference author(s), article title, and where and when it was published. Authors do not send the reference itself or when and if it was retrieved. Just supplying reference information does not mean the reference(s) were obtained or read. It is estimated that only 20% of authors read the original article.9 An error cannot be detected from a source if the source is not available at the time of review. References to nonfunctioning links leave everyone in the dark and without a way to corroborate statements.

If a reference was checked by a reviewer in the evaluation process and an error noted an explanation or correction is asked for. We all make mistakes. Having them pointed out and correcting them is a basic attribute of science and a duty of peer reviewers and editor-in-chiefs. Accuracy of citation quotations and inferences is the responsibility of the author(s) citing them. Checking them is the responsibility of a reviewer and reader.

Since a high rate of misquotes and misstatements are not found during peer review with references held hostage by a paywall 58% of the time and 14% of references unavailable, they are infrequently or not checked at all. Even sources of repositories of manuscript assessments, like Editorial Manager, have suggested readings that when attempted to be retrieved are not available. This is a problem for all medical journal scholarship regardless of being open access or behind a paywall. Medical information should be as accurate and available as possible. A 9% – 25% misquote or citation inaccuracy rate is not scientific communication it is scientific miscommunication. Not being able to check references for how they are used and represented in an article to support its integrity and accuracy is an impediment to science.

When web resources as hypertext transfer protocol (HTTP), uniform resource identifier (URI), and universal resource locator (URL) are used, medical articles have a 70% reference rot (RR) rate.9 Another assessment found one out of five science, technology, and medicinal articles have reference rot, meaning it is impossible to revisit the web context surrounding the reference prior to publication after publication.10 When references were printed on paper they referenced other paper-based documents. What was printed on paper did not change. Web-based references still include paper references, but URLs dominate, which allow immediate access. An unintended consequence of the Internet is that references deteriorate, change, or disappear. Reference rot has two components: link rot and content drift. Link rot identifies the resource by a URL that no longer exists and content drift means the content at the URL has changed from insignificant to being unrecognizable and very different from the original, i.e., revisionist history.

Authors judged the references cited in their article as relevant and important to include them. Editors and peer reviewers agreed to publish the article including these references. It is not known if journals check for validity and availability of references at any step in the vetting process other than plagiarism modeling. Plagiarism algorithms check for words in specific sequence, not words out of sequence quotation, whether intentionally or by mistake. Submitted papers have reference linkage using HTTP, URI, and URL that may or may not have an associated digital object identifier (DOI) and after journal acceptance are given a hyperlink by PubMed prior to publication and listing on PubMed. While PubMed Central (PMC) is a free full-text archive of biomedical and life sciences journal literature at the U.S. National Institutes of Health’s National Library of Medicine (NIH/NLM) https://www.ncbi.nlm.nih.gov/pmc/ when using this site for articles the references they hyperlink to have link rot and drift making references of those articles unavailable. PubMed does not check if a reference is retrievable or has rotted. With $140 billion each year of taxpayer money supporting research11 it is incongruous that outcomes from this research includes references that rot and that permanent archiving is not required.

To combat journal article link rot a DOI can be assigned to an article, along with a URL, allowing it to be found even if the location changes. References cited to the web
at large can have their content updated making the material at the time of an article’s use malleable and if not originally archived inconsistent with its current appearance. Rewriting, modifying, and deletions result in loss of the original meaning. Evolution of science is necessary for progress, but fixity of knowledge is important to understanding scientific events contextually.

URL link reference rot is reported from 3.8% 3 months after publication to 10% at 15 months, 13% by 27 months, and a median life of 9.3 years12 with 27%–28% of URIs lost forever.13,14 What is found on the web at large at any one time is liable to have changed or to have simply disappeared without a trace when time has elapsed and the scholarly statement is published and read.15 The extent of degradation of references was examined over a 15-year period that included over 3.5 million articles from arXiv, Elsevier, and PMC having over one million web resources. Table 3 shows the progression of increasing reference rot from 1997 to 2012.10 The longer the time, the greater the reference rot. Between January 1, 2010 – December 31, 2014 Emergency Medicine Australasia16 had 34.1% RR, between January 1, 2013 – December 31, 2017 the Irish Medical Journal17 had 34.0% RR, and in 2013 the International Urogynecology Journal18 had 27.3% RR.

Articles accepted for publication, but not yet published (in press), had an inaccessible rate to 9.9%,19 at three years post-publication 27% unavailable,12 and at five years 45%–78% unobtainable.20 Journals, editors, readers, and authors must assist and collaborate in eliminating loss of references to maintain the integrity of medical communication and ensure availability of readers and researchers to have access to the original articles. Authors should attest to retrieving, reading, and archiving all their references as part of submitting an article. A few webpage archiving sites are: perma.cc (https://perma.cc), Wayback Machine (www.archive.org/web, https://archive.org), Stop link rot: how to preserve webpages (https://law.tamu.libguides.com/linkrotguide), and archive.today (https://archive.ph). Article references should be listed with their URL and DOI identifiers in the reference listing along with author(s) name(s), article title, journal, date, and any other identifier (DOI). Journals need to check reference availability and accuracy before sending an article for peer review. Peer reviewers should demand access to all references during the review period without charge even if behind a paywall. Prior to publication, journals should check all references again for retrievability and not depend

| Year Since Publication | arXiv | Elsevier | PubMed Central |
|------------------------|-------|----------|----------------|
| 2012                   | 13%   | 22%      | 14%            |
| 2005                   | 18%   | 41%      | 36%            |
| 1997                   | 34%   | 66%      | 80%            |

Table 3.

Example of Reference Rot Progression over 15 Years (1997 – 2012) in More Than 3.5 Million Articles Showing an Increase Over Time10

What Authors, Journals, and Readers Should Do to Combat Reference Rot: Reference Rot = Link Rot Plus Content Drift

| Authors Prior to publication: | Journals Not accept a submission without an attestation that all references are archived and saved by the author | Readers Call out and notify journals and editor-in-chiefs when it occurs |
|-------------------------------|-------------------------------------------------|--------------------------|
| Save all references to a hard drive with a backup, to the cloud and thumb drive | All URLs in references should be checked prior to going to a peer reviewer for their ability to be retrieved (not rotted) | Demand that journals assure that references are archived |
| Archive all references and validate they were archived | Make all references available to peer reviewers | |
| To every reference put authors names, name of paper, date of reference, Journal or place obtained, year, pages, date accessed, http URL identifiers, DOI and archived information | Re-examine all articles periodically to check for reference rot and correct broken URLs | |

| Authors Once accepted: | Journals | Readers |
|------------------------|----------|---------|
| Save the article to a hard drive, cloud, thumb drive and backup | All URLs in references should be checked prior to going to a peer reviewer for their ability to be retrieved (not rotted) | |
| Re-examined periodically to check for reference rot and correct broken URLs | Make all references available to peer reviewers | |

Table 4.

What Authors, Journals, and Readers Should Do to Combat Reference Rot: Reference Rot = Link Rot Plus Content Drift

URL, universal resource locator; DOI, digital object identifier.
on PubMed. Preprint self-archiving of a submitted article including references by authors does not infringe any copyright since it hasn’t been accepted and the copyright exchanged. Post article acceptance journals should archive the article and its references. Reference archiving is an assurance for authors and journals of reference reliability. It is retrievable, free, and specific to the date of archiving. A reference that cannot be recovered is useless. Table 4 summarizes author, journal, and reader archiving necessities.

Internet availability and instantaneous access is convenient, but can fade, morph, or disappear. A solution to reference retrievability for journals and peer reviewers is to require authors to submit their entire set of references with their article. Access to these articles would be only to the reviewers. This at least demonstrates that the submitting author(s) had the total article they are referencing. It doesn’t mean they read any of it or understood it, but at least they had the opportunity. Once an article is accepted or rejected the references supplied would be deleted or returned. Reviewers would not have to look at a reference, but would be able to in a free and unencumbered fashion to make a judgement, assessment, and recommendation.

An alternative to authors sending the whole set of references is to have a reference recovery fee charged to the authors to cover the cost of paywall access. It is unreasonable to expect peer reviewers who are doing their due diligence for free to have to pay to see or corroborate a reference or what it said.

The World Wide Web does not have the permanence of a hard copy. Lack of site payment, repurchase of websites, removed content, website removal of linked pages, and lapse or change of domain name or content filters are some of the reasons for link rot. The complacent approach to reference retrieval using the Internet puts science at risk.

It is scientist’s responsibility to preserve their research, references they use to make their arguments, and relevant findings. Every participant in the publishing process, from authors obtaining references, citing, not archiving; journals and reviewers not checking internet availability; repositories creating hyperlinks that do not go to the resource; and readers not demanding accessibility to cited works; all pointing to others and absolving themselves of the reference integrity mess of rot and drift. These events are not possibilities they are realities. Articles with references that are not retrievable are ships without anchors; therefore, all medical articles should be read with a degree of skepticism of things said, misquoted, misrepresented, lost and unavailable.

**CONCLUSION**

Medical readers deserve to know the validity of article content and their sources. Medical articles have deficits other than methodology, statistical errors, wrongful conclusions, and bias. Most errors are innocent, but must be corrected so that mistakes do not become entrenched. This includes access to references, integrity of citations, and accurate quotation representation. Truthfulness and accuracy of science should not be held hostage to reference rot. Archiving references in toto at the time of use is the solution, what is necessary is the will to do it.

**References**

*All references have been downloaded, read and archived on the WayBack Machine at https://archive.org/web/.*

1. Choi J, Gupta A, Kaghazchi A, Htwe TS, Baicocchi M, Spain DA. Citation inaccuracies in influential surgical journals. *JAMA Surg*. 2021;156(8):791–792. doi:10.1001/jamasurg.2021.1445. PMID: 34037684; PMCID: PMC8156172. A snapshot was captured. Visit page: https://web.archive.org/web/20211019173538/https://jamanetwork.com/journals/jamasurgery/article-abstract/2780288. Accessed 9/2/2021.

2. de Lacey G, Record C, Wade J. How accurate are quotations and references in medical journals? *BMJ (Clin Res Ed).* 1985;291(6499):884–886. doi: https://doi.org/10.1136/bmj.291.6499.884. PMID: 3931753. A snapshot was captured. Visit page: https://web.archive.org/web/20211019173343/https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1416756/pdf/bmjcred00467-0046.pdf. Accessed 6/23/2021.

3. Mogul S. Accuracy of cited “facts” in medical research articles: a review of study methodology and recalculation of quotation error rate. *PLoS ONE* 2017;12(9):e0184727. doi: https://doi.org/10.1371/journal.pone.0184727 A snapshot was captured. Visit page: https://web.archive.org/web/20211019173135/https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0184727. Accessed 6/22/2021.

4. Mertens S, Baethge C. The virtues of correct citation: careful referencing is important but is often neglected even in peer reviewed articles. *Dtsch Arztebl Int.* 2011;108(33):550–552. doi: https://doi.org/10.3238/arztebl.2011.0550 PMID: 21912574. A snapshot was captured. Visit page: https://web.archive.org/web/20211019173906/https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3167934/pdf/Dtsch_Arztebl_1Int-108-0550.pdf. Accessed 7/3/2021.

5. Wager E, Middleton P. Technical editing of research reports in biomedical journals. *Cochrane Database Syst Rev.* 2008;4:MR000002. doi: 10.1002/14651858.MR000002.pub3 A snapshot was captured. Visit page: https://web.archive.org/web/20211019174600/https://www.cochranelibrary.com/cdrt doi/10.1002/14651858.MR000002.pub3.full. Accessed 9/7/2021.

6. Jergas H, Baethge C. Quotation accuracy in medical journal articles systematic review and metaanalysis. *Peer J.* 2015;3:e1364.
