Internet Citations in Oncology Journals: A Vanishing Resource?

An estimated one-third of oncology patients use the Internet to obtain cancer-related information *(1)*. Scientific articles increasingly refer clini-

Table 1. Internet citation characteristics by journal*

Characteristic	Journal						
	CA: A Cancer Journal for Clinicians	*Journal of the National Cancer Institute*	*Biochimica et Biophysica Acta—Reviews on Cancer*	*Advances in Cancer Research*	*Cancer Research*	*Journal of Clinical Oncology*	All journals
No. of issues	3	12	3	3	12	12	45
No. of published works†	13	153	15	17	627	450	1275
No. of citations (endnotes and footnotes)	626	4240	2060	3106	26 435	14 002	50 469
No. of Internet citations	43	46	12	2	162	55	320
Internet citation location							
Text	20	8	12	2	1	7	50
Footnotes	12	2	0	0	161	0	175
Endnotes	11	36	0	0	0	48	95
Inactive Internet citations/all Internet citations							
5 mo	2/25	3/20	0/7	0/2	5/63	4/31	14/148
17 mo	0/4	0/18	0/3	0/0	7/53	2/12	9/90
29 mo	2/14	4/8	1/2	0/0	15/46	5/12	27/82

*All publications from the March and April issues of 2001, 2002, and 2003 were examined.
†Published works included research articles, reviews, news articles, and correspondence.

cians and researchers to information available only on the Internet. Internet information, however, may vanish unexpectedly *(2)*. The ephemeral nature of Internet citations results from each Internet address (uniform resource locator [URL]) specifying both the identifier and location of the electronic content. Internet citations may therefore become inaccessible because of changed locations of cited information without appropriate forwarding links, removal of cited information from the Internet, or equipment failure.

We systematically examined the frequency of use and accessibility of URLs in high-impact oncology journals *(3)* to determine how often readers encounter online roadblocks in accessing cited electronic information (Table 1). We found that the number of articles containing URLs increased annually from 9% in 2001, to 11% in 2002, and to 16% in 2003. The proportion of inactive URLs cited within articles also increased: 9.5%, 10%, and 33% of Internet addresses were inactive 5, 17, and 29 months after publication, respectively. Overall, inactivity was greatest for URLs ending with ".edu" (25%), ".org" (13%), and ".gov" (12%). Searches for inactive URLs using the Internet Archive (http://www.archive.org [last accessed: May 10, 2004]) recovered some information for approximately half of the inaccessible Internet citations. Although archiving resources (e.g., the Internet Archive) may provide information for inactive Internet citations, readers currently have no assurance that any Internet citation continues to cite information originally referenced by the author.

The impermanence of URLs in health sciences literature calls for the prompt attention of readers, authors, editors, and publishers. Although long-term solutions may include instituting systems of permanent information identifiers, such as digital object identifiers (DOIs) or persistent uniform resource locators (PURLs) *(4,5)*, the implementation of these identifiers may be costly and awaits wider acceptance from the online publishing community.

The oncology journals we examined had various Internet citation policies, ranging from no stated policies to specific requirements regarding locations (e.g., in footnotes) and inclusion of Internet citation accession dates. We recommend changes in publishing policies that may offer a more timely answer to Internet citation impermanence, including the adoption of simple guidelines for Internet citations. We recommend that authors be required to 1) include an accession date with any URL, 2) submit all URLs to an electronic archive such as the Internet Archive and provide the date of submission, and 3) maintain a printed copy of the information provided at the URL for future communication until the Internet address and content become available in the electronic archive, which for the Internet Archive is approximately 6 months. The Internet Archive Wayback Machine contains more than 300 terabytes of data (more than the amount of text contained in the world's largest libraries, including the Library of Congress) and may be one of the most effective current means of archiving electronic information and recovering data from inactive URLs.

The rising popularity of citing electronic information in oncology journals demands that publishers adopt new policies to address information management challenges not previously encountered with printed literature.

ERIC J. HESTER
LAUREN F. HEILIG
AMANDA L. DRAKE
KATHRYN R. JOHNSON
CATHERINE T. VU
LISA M. SCHILLING
ROBERT P. DELLAVALLE

REFERENCES

(1) Eysenbach G. The impact of the Internet on cancer outcomes. CA Cancer J Clin 2003;53:356–71.
(2) Dellavalle RP, Hester EJ, Heilig LF, Drake AL, Kuntzman JW, Graber M, et al. Going, going, gone: lost Internet references. Science 2003;302:787–8.
(3) Institute of Scientific Information (ISI). Journal Citation Reports: JCR Science Edition. Institute for Scientific Information, Philadel-

phia (PA); 2001. Available at http://www.isinet.com/products/evaltools/jcr/jcrweb. [Last accessed: May 10, 2004.]
(4) The Digital Object Identifier System. International DOI Foundation; 2003. Available at: http://www.doi.org. [Last accessed: May 10, 2004.]
(5) Persistent URL home page. OCLC Online Computer Library Center: OCLC PURL Service. 2003. Available at: http://purl.oclc.org. [Last accessed: May 10, 2004.]

NOTES

Affiliations of authors: University of Colorado Health Sciences Center, Denver (EJH, LFH, ARD, LMS, RPD); Department of Veterans Affairs Medical Center, Denver, CO (RPD); and George Washington University School of Medicine, Washington, DC (CTV).

Correspondence to: Robert P. Dellavalle, MD, PhD, University of Colorado Health Sciences Center, 4200 E. Ninth Avenue, Box B-153, Denver, CO 80262 (e-mail: robert.dellavalle@uchsc.edu).

Dr. Hester was supported by National Institutes of Health grant T32 AR07411, and Dr. Dellavalle was supported by the National Cancer Institute (NCI) grant number K-07 CA92550. This project was supported in part by research grants from NCI Cancer Education grant number R25 CA49981.

DOI: 10.1093/jnci/djh181