

E-mail Address Harvesting on PubMed—A Call for Responsible Handling of E-mail Addresses

To the Editor: PubMed “comprises over 20 million citations for biomedical literature from MEDLINE, life science journals, and online books,” a database that can be queried using the Entrez search engine.¹ Since January 1996, e-mail addresses for first authors, when available, have been added to the MEDLINE record as they appear in the journals.² Search and retrieval of PubMed results may occur through the PubMed Web page and also with use of several Entrez Programming Utilities that “provide access to Entrez data outside of the regular web query interface.”³ With regard to these latter tools, corresponding documentation and an educational course with examples are available to the public online.⁴ *Electronic spam* can be defined as unsolicited e-mail sent to a large number of addresses.⁵ Individuals sending spam harvest e-mail addresses from the Internet using a variety of techniques, including automated use of software to search Web pages for strings of text recognized as e-mail addresses, as well as manual efforts to gain access to large collections of addresses (eg, by subscribing to mailing lists to collect the addresses of other users). Techniques for avoiding spam are many and include avoiding the online publication of e-mail addresses in text form (as opposed to providing an image of the address) and preventing those with malicious intent from accessing large sources of addresses.

PubMed is extremely vulnerable to e-mail address harvesting. When available, e-mail addresses for first authors are included within citations in text form, making them easily retrieved by software in an automated fashion. However, more concerning is the ability to quickly generate listings containing thousands of e-mail addresses using the Entrez Programming Utilities. With regard to this latter vulnerability, having only basic computer programming knowledge, within 30 minutes of discovering the aforementioned utilities, I was able to generate a listing of more than 7000 addresses. Therefore, clearly more responsible handling of e-mail addresses in PubMed is needed, which may be accomplished by eliminating publication of e-mail addresses in text form and restricting the return of e-mail addresses when results are fetched outside of the regular Web query interface.

Brendan Thomas, MD
University of Illinois at Chicago
Chicago, IL

1. US National Library of Medicine; National Institutes of Health. PubMed Help: FAQs. <http://www.ncbi.nlm.nih.gov/books/NBK3827/#pubmedhelp.FAQs>. Accessed March 3, 2011.

2. US Department of Health and Human Services (DHS). The NLM Technical Bulletin. No 287: November-December; 1995. ftp://nlmpubs.nlm.nih.gov/nlminfo/newsletters/techbull/pdf_tb/novdec95.pdf. Accessed February 8, 2011.

3. National Center for Biotechnical Information (NCBI). Entrez programming utilities. NCBI Website. Updated February 16, 2009. http://www.ncbi.nlm.nih.gov/corehtml/query/static/eutils_help.html. Accessed February 8, 2011.

4. Sayers E, Wheeler D; US National Library of Medicine; National Institutes of Health (NIH). Building customized data pipelines using the Entrez Programming Utilities (eUtils). <http://www.ncbi.nlm.nih.gov/books/NBK1058/>. Accessed February 8, 2011.

5. Spam. Merriam-Webster Online Web site. [http://www.learnersdictionary.com/search/spam\[1\]](http://www.learnersdictionary.com/search/spam[1]). Accessed February 8, 2011.

doi:10.4065/mcp.2010.0817

Weighing the Evidence Linking UVB Irradiance, Vitamin D, and Cancer Risk

To the Editor: A recent article concluded that evidence of a beneficial role of vitamin D in reducing the risk of cancer incidence and mortality is not impressive, in part because of a lack of good randomized controlled trials (RCTs).¹

Randomized controlled trials are appropriate for studying pharmaceutical drugs but not necessarily for vitamin D. Most vitamin D is produced from solar UVB irradiance and confounds oral intake in RCTs of vitamin D. Nested case-control studies are less reliable than case-control studies because the relation of a single serum 25-hydroxyvitamin D [25(OH)D] level measurement to subsequent serum 25(OH)D levels declines with the passage of time, and undetectable cancers can grow rapidly in the absence of adequate serum 25(OH)D levels. This could explain why the results of Helzlsouer² and Harbour and Miller³ showed no beneficial effect of vitamin D.

The strongest evidence to date for a beneficial effect of vitamin D in reducing the risk of cancer comes from ecological studies using solar UVB dose indices. A review of such ecological studies found evidence from 3 continents for 13 types of cancer and from 1 or 2 continents for 5 types.⁴ Many of the recent ecological studies included a number of other cancer risk-modifying factors in the analysis, such as smoking, alcohol consumption, socioeconomic status, diet, and ethnic background. The fact that similar results have been found in diverse geographic locations, including Australia, China, Europe, France, Japan, and the United States, as well as in several multinational studies involving up to 175 countries, strongly supports the role of solar UVB. Although some ecological studies can be faulted for using latitude as the index, summertime solar UVB doses in the United States are strongly asymmetric because of variations in surface elevation and stratospheric ozone layer thickness, with highest doses in the southwest and lowest doses in the northeast, a pattern that correlates well with about 15 types of cancer.⁵ No mechanism other than vitamin D production