PubMed vs. HighWire Press: A head-to-head comparison of two medical literature search engines

Thomas E. Vanhecke*, Michael A. Barnes, Janet Zimmerman, Sandor Shoichet

Department of Internal Medicine, William Beaumont Hospital, 3061 W. 13 Mile Rd., Royal Oak, MI 48073, USA

Abstract

PubMed and HighWire Press are both useful medical literature search engines available for free to anyone on the internet. We measured retrieval accuracy, number of results generated, retrieval speed, features and search tools on HighWire Press and PubMed using the quick search features of each. We found that using HighWire Press resulted in a higher likelihood of retrieving the desired article and higher number of search results than the same search on PubMed. PubMed was faster than HighWire Press in delivering search results regardless of search settings. There are considerable differences in search features between these two search engines.
© 2006 Elsevier Ltd. All rights reserved.

Keywords: Internet; Search engine; PubMed; HighWire Press; Medical literature

1. Introduction

There is little doubt that the internet has changed modern medicine forever. A computer with a fast connection to the internet has become as important to a physician as the stethoscope and the reflex hammer. What used to take countless hours of digging through stacks of bound journals in a medical library now takes mere seconds behind a computer with a fast internet connection.

Despite this new technology, navigating these vast resources can be a daunting task. Thankfully, the National Library of Medicine created Medline, the all inclusive medical literature database. Various interfaces have been developed which search and access the Medline database. However, very few of these sites are available without cost to the user. PubMed and High-Wire Press are two medical literature search engines that offer free access to Medline and various other medical literature databases and are readily available to anyone with an internet connection [1].

PubMed, the most widely used free medical literature search engine, is a service of the National Library of Medicine and can be accessed without charge at http://pubmed.gov [2]. PubMed has been available to the public since 1995 and boasts usage in upwards of 70,000,000 searches per month [2].

A new and increasingly popular search engine is HighWire Press, a creation of Stanford University that can be freely accessed at http://highwire.org [3]. HighWire Press is also a no-cost, web-based, medical literature search engine. Besides offering seamless access to Medline, HighWire Press declares it "hosts the largest repository of free, full-text, peer-reviewed content, with 999 journals and 1,460,993 free, full-text articles online." [3]. Despite being released to the public in 2003, HighWire Press remains relatively new and obscure compared to PubMed. Vanhecke described many of the search features of HighWire Press in a recent review [4]. Usage statistics provided by HighWire Press demonstrates growth to approximately 10,000,000 searches conducted per month through its web servers [3]. With the ever increasing popularity of High-Wire Press, an objective measurement of this tool is needed.

2. Research design

In the fast-paced life of clinical medicine, practitioners rely on the "quick search" features of these internet search portals to rapidly locate articles of interest. To date, there has

---

* Corresponding author. Tel.: +1 248 259 6025; fax: +1 248 551 8880.
  E-mail address: tvanhecke@gmail.com (T.E. Vanhecke).

not been a study which has directly compared these two search engines against each other using the quick search features. In the present study we sought to directly compare HighWire Press with PubMed for use in quick, clinically-focused medical searches. Main study outcomes include search efficiency, amount of time it takes to conduct searches, thoroughness, and features of both HighWire Press and PubMed using a set of experimental searches using the quick search features of each search engine.

3. Methods and procedures

All experimental searches were conducted from March 5th, 2006 through April 22nd, 2006. To simulate real-life usage, these experiments were conducted at three different locations in the vicinity of Royal Oak, Michigan. The first location was a residential location, the second location was a public coffee shop, and the third location was within William Beaumont Hospital in Royal Oak, Michigan. A wireless connection was used to connect to the internet from within the residential location and the public coffee shop. A T1 Ethernet connection was used to connect to the internet from within William Beaumont Hospital. To ensure equal internet connectivity, bandwidth speed measurements were taken before and after each experimental session. A Lenovo X-41T ThinkPad laptop PC with Microsoft Internet Explorer was used in all experiments. The homepage of each search engine was analyzed using commercially available software (Web Page Analyzer—0.961) to assess page size, composition, and download times. A commercially available software "stopwatch" program (Virtual Stopwatch v3.12a) was used to measure the amount of time it took to conduct searches. Results were expressed as mean $\pm$ standard deviation or counts with proportions as appropriate. Chi-square analysis was performed for trend as appropriate. A p-value < 0.05 was considered statistically significant.

3.1. Search engine accuracy

The ability of PubMed and HighWire Press to locate a particular article was measured with a set of experimental searches involving 49 different research publications. The 49 publications used in this experiment were taken from an article published in the Journal of the American Medical Association in 2005 [5]. In this article, Ioannidis presented a table [5] of 49 different publications within various respected medical journals which had received more than 1000 reference citations apiece. In this table, the author presented a list of these publications along with a small description of each study. We used the study descriptions provided in Ioannidis' table as keywords in these experimental searches. With the exception of omitting commonly used words such as "the" or "to", the study descriptions were entered exactly into the keyword search fields of each search engine, and a search was conducted. For example, one of the publications identified by Ioannidis was the North American Symptomatic Carotid Endarterectomy Trial published in the New England Journal of Medicine in 1991 [6]. Correspondingly, the author's description of this publication was: "Carotid endarterectomy in

high-grade stenosis" [5]. Therefore, the word "in" was omitted and "carotid endarterectomy high-grade stenosis" was entered into the search field of each search engine and a search was then performed with these keywords. Searches in PubMed and HighWire Press were conducted on all 49 publications listed in the table. The overall frequency of deriving the specified article within the first 25 and first 200 results was recorded for each search engine. All searches were conducted with the simple quick search input field located on the home page of each search engine. All searches on HighWire Press used the "all results, including PubMed" search option with the default best match search feature.

3.2. Search engine speed

The time it takes to complete a search was measured for both PubMed and HighWire Press. These experiments used Virtual Stopwatch v3.12a to measure the time taken to complete a given search after pressing the search key. Three different searches were conducted using the following keyword search terms: "obesity", "flow-mediated dilation", and "asymmetric dimethyl arginine". Since both PubMed and HighWire Press offer users the option to change the number of results which are received per page, time measurements were taken on every different "results per page" setting for both HighWire Press and PubMed. Each experiment was repeated three different times and results were averaged and reported with standard deviation. The overall number of results supplied by each search engine was also recorded.

Both PubMed and HighWire Press offer the option to obtain full-text access of a particular article if it is available. The amount of time taken to obtain the full-text article in PDF format from a known available source was evaluated. This experiment was conducted utilizing three different articles published in respected journals in 1999 or 2000 [7–9]. The full-text PDF document of each of these articles can be directly accessed through a link provided in the search results for both PubMed and HighWire Press. Virtual Stopwatch v3.12a was used to measure the time it took to download the entire full-text PDF. Each experiment was repeated three times and results were averaged and reported with standard deviation. These experiments were conducted on the same computer at three different locations, as described above.

Author name searches were conducted on a set of author names. In the same manner as above, the author names that were used in this set of experimental searches were provided by a recent publication [5]. The first author of each publication listed in Ioannidis' table [5] was used in an author name search using both PubMed and HighWire Press. The full last name and first and middle initials, when available, were entered into the search fields and a search was conducted. The numbers of results generated were recorded.

3.3. Search engine features

Both PubMed and HighWire were evaluated for their search capabilities and options to modify the display of results.