# How The Public Access Health Information On The Internet: A Study Involving Public Internet Access Points

V Scaria

**Citation**

V Scaria. *How The Public Access Health Information On The Internet: A Study Involving Public Internet Access Points*. The Internet Journal of Medical Informatics. 2003 Volume 1 Number 2.

**Abstract**

Background: Internet is used to access a variety of information. Health information is obviously one of them. Only a few studies[1,2,3] have been conducted to analyse how consumers access health information on the Internet. To date no similar study have been reported from India.

Objectives: To analyse the actual occurrence of access to health information on the Internet in the general population in a natural and non-obtrusive setting and to quantify how information is accessed on the Internet.

Design: A quantitative study of the Internet access pattern in a natural setting and analysis of data generated from the initial observation.

Methods: Analysis of the History of Internet access on ten personal computers over a period of seven days, and further analysis of the data generated.

Setting: Calicut, Kerala, India.

Results: We analysed the Internet access history of seven computers over a period of seven days and retrieved information on 10,087 page views corresponding to 5316 webpages and 4095 websites. Only 7150 page views were amenable for analysis, out of which 274 page views corresponding to 228pages contained obvious health information. Of the health webpages accessed, 57 were accessed via search engines.

Conclusions: Further studies utilizing advanced softwares to accurately track Internet access patterns of users need to be employed for accurate analysis.

## INTRODUCTION

Health related websites are frequently accessed on the Internet. Nothing is much known on how users access health information on the Internet in a natural setting. Only a few studies have been conducted to analyse the behavior of Internet users in accessing health information on the Internet [1,2,3].

Kerala, the southern state of India is the leading state in Healthcare and Education. The ‘Kerala model' of Healthcare and Education, widely discussed as a model for developing countries has created an enviable health scenario and a health conscious population. This study is important in the latest setting in which the state is embarking on a mass e-literacy drive under the Kerala IT Mission Project.

This study aims to analyse the actual occurrence of access to health information on the Internet, how they access health information in an un-obtrusive natural setting.

## MATERIAL AND METHODS

Setting: Calicut, India

Material: Ten personal computers in five public Internet access centers in Calicut city were selected for the study. All these computers were Pentium based systems working on Microsoft Windows 98 Operating System and having Netscape Navigator as the default browser software. The computers had access to the Internet via high-speed

Integrated Service Digital Network (ISDN) lines or high-speed optic fiber cables. Before commencement of the study, the browsers were programmed to retain the history of web pages accessed for a period of 10 days and to cache the pages accessed on the hard disk.

Data Collection: The data on web pages accessed were retrieved periodically from the computers off working hours. The URLs and the cached files were also saved for further analysis. The user login and logout times were accessed from the logbooks maintained at these centers. These documents contained no user information other than the computers used and the respective login and logout times. Data from three computers were excluded from further analysis as one of them had to undergo repair and two of them had their history data removed by the users.

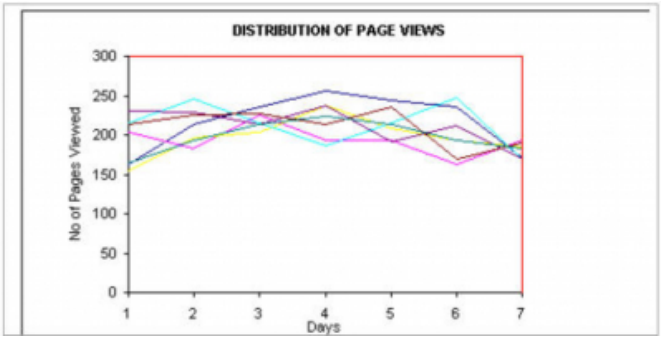
Analysis: Revisiting the pages from the cache, we performed analysis of webpage content. The password-protected pages were excluded from the study, as they were not amenable to analysis. All the pages that contained obvious health related information and health websites were later saved for analysis. The access history was later revisited to analyse the source of the hyperlink to pages.

## RESULTS

The access history revealed a total of 10,087 page views [Figure1] corresponding to 5316 web pages and 4095 websites. Only 7150 page views were amenable to analysis, as the information on the other pages were not cached or were password-protected pages as in e-mail accounts. A total of 318 users accessed the computers during the study period of one week, (figure1) corresponding to a mean 70.65minutes and a mean of 31.72pages per user. 73 users (22.96% of all) accessed health information on the Internet.274 page views corresponding to 228 pages (4.29%of all pages) and 152 websites (3.71%of all websites) contained obvious health information and each user accessed a mean of 3.75 pages.

### Figure 1

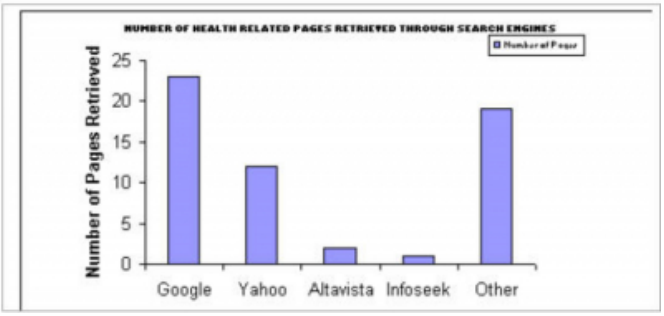
Figure 1: Distribution of Page views during the period of study




Out of the 228 webpages containing Health related information, 57 pages were accessed via search engines [Figure2] as evidenced by a visited hyperlink in the search results page in that login session, while another 86 websites were accessed via hyperlinks from health websites and 5 from non-health websites as evidenced by visited hyperlinks in pages preceding that page in that login session. Of the search engines, Google [5] and Yahoo [6] together referred more than half of the health information webpages. None of the health specific search engines were employed by users to search for Health information on the Internet.

### Figure 2

Figure 2: Number of pages containing Health related information retrieved via Search Engines (N=57)




Out of the 5088 non-health webpages, 1385 pages were accessed through search engines as evidenced by a visited hyperlink in the search results page preceding the visit in that particular login session.

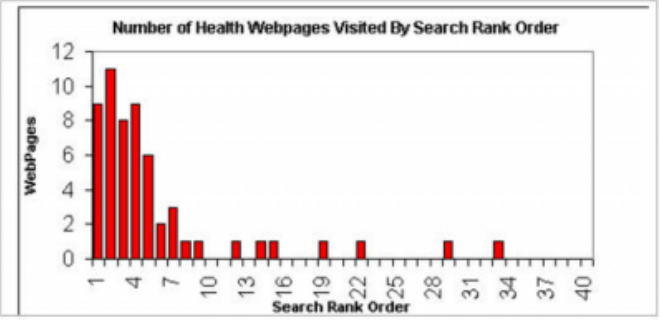
The Health websites were later analyzed by visiting the websites entirely for the presence of a trustmark or seal related to Health Information and we came across 20 websites that displayed the HON seal [4], out of which 12 were retrieved via search engines.

The search results pages were also analyzed and the number

of health webpages users accessed by rank order was plotted [Figure3] and the data revealed that a majority of health webpages visited (89.47%) by users was from the first ten ranks of search results. Only 7 out of the total 57 websites retrieved belonged to an order greater than 10.

### Figure 3

Figure 3: Number of pages containing health information visited by rank order in search results (N=57)




## DISCUSSION

The main outcome of this study is that access patterns of Health information on Internet could be quantified in an un-obtrusive, natural setting. The analysis outcome that 25% of health information was accessed via Search engines puts forward the necessity of employing a midstream filtering of health information. The finding that a referring webpage to a number of health information webpages accessed could not be retrospectively found out may mean either the users targeted specific webpages, rather than browsing endlessly. This information could be possibly derived from other sources like periodicals, books or advertisements, which need to be exactly quantified by structured interviews or surveys, which was impossible in this setting.

Most of the websites that displayed the HON seal were visited via search engines. This may be indirectly linked to the quality of these websites [7] as good quality websites have a higher chance to be 'cited' by other websites [8].

The study also revealed that a majority of health webpages visited (89.47%) by users was from the first ten ranks of search results, which implies search rank was clearly an important factor deciding the access to any particular health web page. This finding has a striking similarity with the pioneering study conducted in a different setting [1] Though different search engines employ different algorithms to assign ranks, the number of incoming links or web 'citations' is an important factor in determining the rank order in many search engines like Google [9]. The finding that users for searching for health information employed none of the Health search engines is probably due to the lack of awareness of these utilities.

## CONCLUSIONS

Further studies employing advanced technology need to be conducted to absolutely quantify Internet access patterns. In addition, these need to be supplemented by structured interviews or surveys to qualify data from observational studies in a user perspective. We hope that this paper would kindle interest for further research in this field.

## ACKNOWLEDGEMENTS

I would like to thank the staff at the Internet access centers for their support and assistance they rendered. I would also like to thank my colleagues and friends, for motivating me to undertake this study.

Funding: None

Competing interests: None

## CORRESPONDENCE TO

Vinod Scaria Perumcheril-33/4711,Malaparamba Calicut 673009,Kerala India E-Mail: vinodscaria@yahoo.co.in

## FOOTNOTES

The study involved the analysis of History files [on the computer] and Log Files [maintained at the Access Centers]. The Log files contained no data other than the Login and Logout times of each visitor on the respective Computers and thus no personal information was analysed.

### References

1. Eysenbach G, Kohler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. BMJ 2002 Mar 9;324(7337):573-577. http://bmj.com/cgi/content/full/324/7337/573?view=full&pmid=11884321
2. Rozic-Hristovski A, Hristovski D, Todorovski L Users' information-seeking behavior on a medical library Website. J Med Libr Assoc. 2002 April; 90 (2): 210-217 http://www.pubmedcentral.gov/articlerender.fcgi?artid=100766
3. Larner A Prevalence of internet access and use for medical information URL: http://bmj.com/cgi/eletters/324/7337/577#20976 [Accessed on 2/03/03]
4. Health on Net Code of Conduct Principles URL:http://www.hon.ch/HONcode/[Accessed on 2/10/02]
5. Google URL http://www.google.com/ {Accessed on 2/10/02]
6. Yahoo URL: http://www.yahoo.com/ [Accessed on 2/10/02]

7. Fallis D., Fricke M. (2002). Indicators of Accuracy of Consumer Health Information on the Internet: A Study of Indicators Relating to Information for Managing Fever in Children in the Home. J. Am. Med. Inform. Assoc. 9: 73-79 http://www.jamia.org/cgi/content/full/9/1/73
8. Hernández-Borges AA, Macías-Cervi P, Gaspar-Guardado MA, Torres-Álvarez de Arcaya ML, Ruiz-Rabaza A, Jiménez-Sosa A. Can Examination of WWW Usage Statistics and other Indirect Quality Indicators Help to Distinguish the Relative Quality of Medical websites? J Med Internet Res 1999;1(1):e1 http://www.jmir.org/1999/1/e1/
9. Google Technology: Page Rank Explained URL:http://www.google.com/technology/index.html [Accessed 2/10/02]

## Author Information

**Vinod Scaria**
Centre for Cybermedicine and Internet Research