The Breast 21 (2012) 514–517

Contents lists available at SciVerse ScienceDirect

The Breast

journal homepage: www.elsevier.com/brst

Original article

Breast cancer information on the internet: Analysis of accessibility and accuracy

E.M. Quinn [a,*], M.A. Corrigan [a], S.M. McHugh [b], D. Murphy [c], J. O'Mullane [c], A.D.K. Hill [b], H.P. Redmond [a]

[a] Department of Academic Surgery, Cork University Hospital, Wilton, Cork, Ireland
[b] Department of Surgery, Beaumont Hospital, Dublin, Ireland
[c] Department of Computer Science, University College Cork, Cork, Ireland

ARTICLE INFO

Article history:
Received 12 April 2011
Received in revised form
22 November 2011
Accepted 31 January 2012

Keywords:
Breast
Cancer
Internet
Patient information

ABSTRACT

Studies show internet sourced information often has poor accuracy. However, it is rapidly becoming a major source of patient information. Our aim was to assess accuracy of breast cancer-related information on the internet.

The top five breast cancer-related search terms were identified using the commercial program "Wordtracker". These terms were searched using the search-engine "Google" and the top 100 webpages per topic analysed for applicability and accuracy of information.

Overall 500 webpages were analysed. 42% were inapplicable to the question asked. Applicable accuracy rates were variable amongst the five terms: "breast cancer symptoms" 84%, "breast cancer care" 87%, "breast cancer stage" 88%, "breast cancer survival" 91% and "breast cancer signs" 78%. Educational websites were more likely to be accurate($p < 0.001$) and interest group administered websites less likely to be accurate($p = 0.018$) than other websites.

Finding accurate breast cancer information on the internet is difficult due to large numbers of inapplicable unregulated websites preferentially returned via search engines.

Introduction

Breast cancer affects approximately 2400 women in Ireland and 45,000 in the United Kingdom (UK) per year.[1,2] The internet is an increasing source of healthcare associated information for the general population with a rising number of patients accessing medical information online.[3,4] This has been shown to be especially true with regard to cancer-related information overall and breast cancer in particular.[5,6] In addition to information provision, internet sourced information has also been shown to be a factor influencing patient decisions with regard to their treatment.[4]

Medical-based websites on the internet lack regulation and may potentially provide inaccurate information to patients. Internet search engine results for medical information queries may not be specific to the individual patient's questions. Previous studies have noted poor quality internet information in both head and neck cancer[7] and gastric cancer,[8] as well as non-surgical disciplines, such as rheumatology.[9] Specifically in breast cancer, previous research has also revealed variability in the quality of information provided by voluntary websites in the UK.[10] Information relating to aromatase inhibitors has also been reported as of poor quality, with frequent inaccuracies.[11]

One organisation aiming to provide regulated internet healthcare information is "Health on the Net"(HON).[12] This organisation provides a listing of websites deemed by the Economic and Social Council of the United Nations to be a reliable and useful source of healthcare associated information.[12] However, although websites undergo examination before inclusion, registration with this organisation is entirely voluntary.

The aims of this study therefore were to assess the most commonly searched terms relating to breast cancer on the internet, the appropriateness of the returned websites and the accuracy of information on these sites. In this manner we sought to determine the quality of information provided to breast cancer patients performing internet searches relating to their diagnosis.

Methods

The commercial program 'Wordtracker' (Wordtracker LLP, London, UK, www.wordtracker.com) was used to identify the most commonly searched terms relating to breast cancer via the internet search engine 'Google' (Google Inc, Mountain View, CA, USA, www.google.co.uk) over a one month study period commencing August 2010. The top five most commonly searched terms were captured and each inputted into a separate basic Google search, as would be

* Corresponding author. Tel.: +353 21 4922371; fax: +353 21 4546400.
E-mail address: edelquinn@rcsi.ie (E.M. Quinn).

0960-9776/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.breast.2012.01.020

performed by the general population. The first 100 results returned for each of the five searched terms were recorded for analysis.

Webpages were assessed for their accuracy in relation to the search term by comparison with recognised peer-reviewed sources of information. The sources used were the Royal College of Surgeons in Ireland's Breast Cancer Management guidelines,[13] National Institute for Health and Clinical Excellence (NICE) guidelines on breast cancer,[14] National Cancer Registry of Ireland survival figures,[1] Surveillance Epidemiology and End Results (SEER) breast cancer survival figures[15] and the American Joint Committee on Cancer (AJCC) staging guidelines for breast cancer.[16] Information on each webpage relating to the search term used was identified and this information compared directly to the relevant source information for correlation. Pages were deemed inaccurate if they declared symptoms or signs not contained within either the NICE or RCSI guidelines on breast cancer as important, if survival figures were inconsistent with both survival data sources, if breast cancer care management was recommended which is not advised in either the NICE or RCSI guidelines and if staging quoted did not match that of the AJCC staging guidelines. Data was extracted and compared to source material by one author (EQ).

Websites returned were sub-classified by type into seven categories: governmental sites, education facility related sites (for example a teaching hospital or university), registered charity organisations, commercial company sites, news and media sites, individual sites or group-run interest sites. Websites were also assessed for the presence of the HON standard. Websites returned were excluded if they were simply redirecting sites to another list of search engine results via a different search engine. All other websites were included. Statistical analysis performed using Fisher's Exact Test with $p < 0.05$ considered significant.

Results

Over the one-month study period 368,000 searches were made using the term "breast cancer". Of these searches the most commonly searched for subcategory was "breast cancer symptoms" accounting for 22,200 (6%) searches, followed by "breast cancer care" with 18,100 (4.9%) searches. The remainder of the top five subcategories consisted of "breast cancer stage" ($n = 6{,}600$, 1.7%), "breast cancer survival" ($n = 6{,}600$, 1.7%) and "breast cancer signs" ($n = 5{,}400$, 1.4%).

A standard Google search of the term "breast cancer" at the end of the study returned 103,000,000 webpages. A search of the specific term "breast cancer symptoms" returned 12,300,000 webpages, while "breast cancer care" returned 32,400,000. A search for "breast cancer stage" returned 18,400,000 webpages with "breast cancer survival" returning 123,000,000 and "breast cancer signs" 9,000,000.

For each search term an extrapolated question was constructed in order to analyse the first 100 results returned for each term as to whether they provided sufficient quality information to correctly answer the question[Table 1].

Of the 500 webpages overall 244 (48.8%) accurately answered the posed questions. Overall 19 (3%) webpages were inaccessible. A total of 211 (42.2%) webpages were not applicable in that they did not provide medical information with regard to the breast cancer term searched. Of the remaining 289 webpages the information provided was accurate in 244 (84.7%), with commercial company websites accounting for 152 (52.6%), followed by news and media sites ($n = 38$, 13%), educational sites ($n = 23$, 8%), individual sites ($n = 22$, 7.6%), interest group sites ($n = 25$, 8.6%), charity run sites (16, 5.5%) and government sites ($n = 13$, 4.5%) [Fig. 1]. Educational facility related websites were significantly more likely to be accurate than other websites ($p < 0.001$) whereas interest group-run websites were significantly less likely to be accurate than others ($p = 0.018$) [Table 2].

Table 1
Top five search terms returned from a wordtracker search of "breast cancer". The searches column shows the number of searches using that term on www.google.co.uk in the preceding month. The websites column shows the number of websites returned from the search engine when that search is performed. The question column shows the extrapolated question used to analyse the returned websites.

	Searches	Websites	Question
Breast cancer symptoms	22,200	12,300,000	What are the symptoms of breast cancer?
Breast cancer care	18,100	32,400,000	What is the treatment and follow up care in breast cancer?
Breast cancer stage	6600	18,400,000	How is breast cancer staged/what does my stage of breast cancer mean?
Breast cancer survival	6600	123,000,000	What is the overall (or by stage) survival rate for breast cancer?
Breast cancer signs	5400	9,000,000	What are the signs of breast cancer?

Each of the five topics searched were then analysed for topic specific accuracy of both all webpages returned ($n = 500$) and question applicable webpages ($n = 289$) returned. Topic specific accuracy rates significantly improved for all subgroups except breast cancer signs when inapplicable websites were excluded (Symptoms: 63% vs. 84% $p = 0.002$, Care: 34% vs. 87% $p < 0.001$, Stage: 46% vs. 88% $p < 0.001$, Survival: 32% vs. 91% $p < 0.001$, Signs: 68% vs. 78% $p = 0.139$) [Fig. 2].

Overall 63 (22%) of question applicable websites displayed the HON accreditation. HON positive applicable websites had an accuracy rate of 92% whereas HON negative applicable websites had an accuracy rate of 82%. However, this finding was not statistically significant ($p = 0.075$).

Discussion

Our study confirms that the internet is used as a source of health information with 368,000 Google searches performed on the topic of breast cancer in one month. This high level of usage has previously been reported with McHugh et al. demonstrating that breast cancer is the most commonly researched cancer online[5] Castleton et al. demonstrated that 63% of 500 adult patients with cancer reported searching the internet for information about cancer, with the majority (54%) doing so prior to their first oncology visit.[4] It has also been reported that 71% of patients searched the internet for further information after being given a cancer diagnosis.[3]

Our study shows that there is a large amount of information in relation to breast cancer available on the internet; searching for

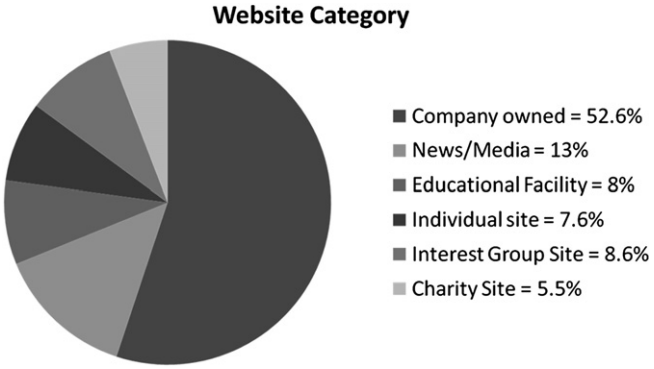

Fig. 1. Breakdown of applicable websites by subgroup ($n = 289$).

Table 2
Accuracy of applicable websites returned ($n = 273$) by subgroup. Figures shown are percentages. Education facility related websites are more likely to be accurate than other websites ($p < 0.001$) whereas interest group-run websites are less likely to be accurate ($p = 0.018$).

Accuracy for:	Breast cancer symptoms	Breast cancer care	Breast cancer stage	Breast cancer survival	Breast cancer signs
Charity	100	100	100	100	100
Company	82	85	81	88	82
Educational	100	100	100	100	100
Government	100	100	100	100	100
Individual	71	75	100	100	20
Interest Group	50	80	100	N/A	55
Media	60	80	83	90	100

"breast cancer" returns over 100 million webpages. The most commonly searched term "breast cancer symptoms" returns 12.3 million webpages. With such a large return of search engine results it can be difficult for patients to find answers to their specific questions. Indeed 42% of returned webpages by the Google search engine were not applicable to the specific question being asked.

Only 22% of 500 webpages assessed in our study displayed confirmation of the validity of their information via the HON accreditation. Previous data has shown that 69% of cancer patients do not have a particular internet site they source information from,[4] reflecting the lack of well promoted, patient friendly, accurate websites.

The variability in quality and accuracy of internet information available is unquestionably source related. Certainly accurate information is available, with governmental sites, formal educational sites and charity sites all having question applicable accuracy rates of 100% as demonstrated in our study. However, accuracy rates for individually run sites range from 20% to 100% depending on the term searched. Interest group-run sites are significantly less likely to be accurate than other sites ($p = 0.018$). This confirms that inaccurate information is prevalent on the internet and certain types of sites should be avoided. Scullard et al. similarly showed that governmental websites have a 100% accuracy rate compared to a 55% accuracy rate for news sites in relation to paediatric health.[17]

The association between HON associated sites and increased accuracy did not quite reach significance. This may however be as a result of low numbers with only 22% of sites assessed in our study HON associated. Similarly low rates of HON usage (39%) have also been reported in previous studies.[7]

A recent study by Meric et al. found that more popular and less popular breast cancer sites did not differ in the quality of information provided or levels of inaccuracy.[18] The content rather than the quality of websites correlated with popularity suggesting that inclusion of popular content can be a way of drawing patients to a particular site. Therefore, when designing or recommending accurate health websites care must be taken to ensure patient friendly, relevant content.

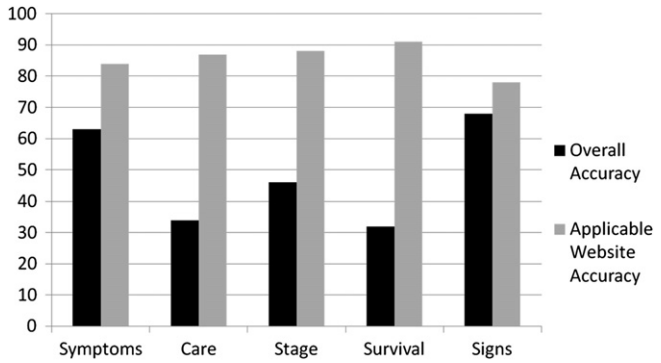

Fig. 2. Accuracy of websites returned by subgroup. Results show overall ($n = 500$) accuracy (symptoms 63%, care 34%, stage 46%, survival 32%, signs 68%) and question applicable ($n = 289$) accuracy (84%, 87%, 88%, 91%, 78% respectively).

The most commonly searched terms relating to breast cancer were broad, with high resultant numbers of search engine results. This may have contributed to the high rate of inapplicable websites returned in our study. It is likely that focused specific search terms could result in smaller numbers of websites returned with a higher applicability rate. Patients performing their own searches should perhaps be encouraged to use narrower search terms if they have a particular question in mind.

Acknowledged limitations of our study include restriction of the search to one search engine www.google.co.uk, as it is recognised that patients can use different search engines. However, this is the search engine that had been shown to have 368,000 breast cancer searches in the preceeding month and hence this search engine was used. Questions the patients may have been seeking answers to had to be extrapolated from the search terms used. It is impossible to know exactly what patients were looking for when they performed these searches. To counter-act this limitation, it was attempted to keep the questions analysed as broad as possible.

Conclusions

Breast cancer-related information on the internet has reasonable accuracy rates once inapplicable websites are excluded. However the accuracy rates vary enormously depending on the type of website accessed. Finding websites with applicable and accurate information using broad internet searches is difficult. Patients wishing to use the internet as a source of health information should ideally be directed to appropriate, accurate websites by their health professionals.

Ethical approval

Ethical approval was not required for this study.

Funding source

The study was funded directly by the authors.

Conflict of interest statement

The authors declare no conflicts of interest.

References

1. National Cancer Registry of Ireland. *Incidence, mortality, treatment and survival*, http://www.ncri.ie/data.cgi/choose-methods.php; 1994-2007 [accessed 15/11/10].
2. Cancer Research UK. *Breast cancer statistics: key facts*, http://info.cancerresearchuk.org/cancerstats/types/breast/#incidenceCancer; 2010 [accessed 15/11/10].
3. Van de Poll-Franse LV, van Eenbergen MC. Internet use by cancer survivors: current use and future wishes. *Support Care Cancer* 2008;**16**(10):1189–95.
4. Castleton K, Fong T, Wang-Gillam A, Wagar MA, Jeffe DB, Kehlenbrink L, et al. A survey of internet utilization among patients with cancer. *Support Care Cancer* 2010 [epub ahead of print] PMID: 20556435.
5. McHugh SM, Corrigan M, Morney N, Sheikh A, Lehane E, Hill ADK. A quantitative assessment of changing trends in internet usage for cancer information. *World J Surg* 2011;**35**(2):253–7.
6. Fogel J, Albert SM, Schnabel F, Ditkoff BA, Neugut AI. Use of the internet by women with breast cancer. *J Med Internet Res* 2002;**4**(2):E9.
7. Ni Riordain R, McCreary C. Head and neck cancer information on the internet: type, accuracy and content. *Oral Oncol* 2009;**45**(8):675–7.
8. Killeen S, Hennessey A, El Hassan Y, Killeen K, Clarke N, Murray K, et al. Gastric cancer-related information on the internet: incomplete, poorly

accessible and overly commercial. *Am J Surg* 2010 [epub ahead of print] PMID: 20851373.
9. Thompson AE, Graydon SL. Patient-orientated methotrexate information sites on the Internet: a review of completeness, accuracy, format, reliability, credibility and readability. *J Rheumatol* 2009;**36**(1):41–9.
10. Ream E, Blows E, Scanlon K, Richardson A. An investigation of the quality of breast cancer information provided on the internet by voluntary organisations in Great Britain. *Patient Educ Couns* 2009;**76**(1):10–5.
11. Beaton C, Codd RJ, Holland PA, Gateley CA. Evaluation of the quality and accuracy of information regarding aromatase inhibitors available on the internet. *Breast J* 2008;**14**(4):366–8.
12. Berland GK, Elliot MN, Morales LS, Algazy JI, Kravitz RL, Broder MS, et al. Health information on the internet: accessibility, quality and readability in English and Spanish. *JAMA* 2001;**285**:2612–21.
13. Clinical Guidelines Committee Royal College of Surgeons in Ireland. *Breast cancer management clinical guidelines*. Royal College of Surgeons in Ireland, http://www.rcsi.ie/files//surgery/docs/20090209032025_Breast%20Cancer.pdf; 2000 [accessed 28/08/2010].
14. National Institute for Health and Clinical Excellence. *Breast cancer (early & locally advanced): diagnosis and treatment*, http://guidance.nice.org.uk/CG80; February 2009 [accessed 28/08/2010].
15. Gloeckler Ries LA, Eisner MP. *Cancer of the female breast, 5-year relative survival rates (%) by histology and AJCC stage*. Ages 20+. 3rd ed. SEER, http://seer.cancer.gov/publications/survival/surv_breast.pdf; 1988–2001 [accessed 28/08/2010].
16. American Joint Committee on Cancer. Breast. In: *AJCC cancer staging manual*. 7th ed. New York: Springer; 2010. p. 347–69.
17. Scullard P, Peacock C, Davies P. Googling children's health: reliability of medical advice on the internet. *Arch Dis Chil* 2010;**95**:580–2.
18. Meric F, Bernstam EV, Mirza NQ, Hunt KK, Ames FC, Ross MI, et al. Breast cancer on the world wide web: cross-sectional survey of quality of information and popularity of websites. *BMJ* 2002;**324**:577–81.