Archived at the Flinders Academic Commons: http://dspace.flinders.edu.au/dspace/

This is the authors’ version of an article published in the Australian Journal of Primary Health. The original publication is available by subscription at: http://www.publish.csiro.au/nid/261/aid/11548.htm

doi:10.1071/PY13030

Please cite this article as:


Journal compilation © La Trobe University 2013. All rights reserved. Please note that any alterations made during the publishing process may not appear in this version.
The objective of this systematic review was to identify types of approaches and methods used to evaluate the effectiveness of healthcare information websites. Simple usage data may not be sufficient to assess if the desired healthcare outcomes were achieved or to determine the relative effectiveness of different web resources on the same health topic. To establish the state of the knowledge base on assessment methods used to determine the effectiveness of healthcare websites, a structured search of the literature was conducted in Ovid Medline resulting in 1,611 articles retrieved, of which 240 met the inclusion criteria for this review. Results of this review found that diverse evaluation methods were used during development, prior to release, and after release. Economic assessment was rare and most evaluations looked at content issues such as readability scores. A number of studies did try to assess the usefulness of websites but few studies looked at behaviour change or knowledge transfer following engagement with the designated health website. To assess the effectiveness of the knowledge transfer of healthcare information through the online environment, multiple methods may need to be used to evaluate healthcare websites and may need to be undertaken at all stages of the website development process.
24 Summary Statement (n=95)

25

26 What is known about the topic?
27 There is increasing use of online resources such as health websites.
28 More studies on website usage and assessments of their impact are being published.
29 This research can be used to improve the functionality and contribution of these
30 resources.

32 What does this paper add?
33 This paper establishes an evidence base of evaluation methods used to assess the
34 effectiveness of healthcare websites.
35 The evaluation methods identified in this review were diverse and applied at different
36 stages of development and production.
37 There is a limited evidence base relating to economic assessment and behaviour
38 change.
Introduction (N=734)

Given the diversity of information needed by health professionals and health consumers within a diversity of care settings, the online environment offers a powerful means in which to disseminate and maintain information currency as well as encourage engagement with healthcare knowledge. Information seekers now have an unprecedented ability to access vast amounts of information on any health issue via a keyboard or touch screen. Online healthcare information dissemination is powerful because of its reach, its relative cost advantages and its immediate availability. As a result, there is growing interest in the role of e-health and telehealth within the healthcare system, particularly within primary healthcare (Australian Government, 2010).

Statistics provide evidence of the rapid uptake and use of the Internet by all age groups, geographic regions and countries for accessing information, including healthcare information. Figures from the Internet World Stats website showed that in June 2012, there were 2.45 billion people, or 34.3% of the world’s population, with access to the Internet (Internet World Stats, 2013). According to the Australian Bureau of Statistics in Australia, during the same time period there were over 12 million Internet users in Australia, and the number of users is increasing at around 10% annually (Australian Bureau of Statistics, 2012). A significant proportion of this Internet activity relates to health issues. A study by Eysenbach and Kohler examined the prevalence of health related searches on the web by analysing search terms entered into popular search engines. Results of this research showed that an estimated 4.75% of all searches were health related (Eysenbach and Kohler, 2003). The Pew Internet & American Life Project report, Health Online 2013, indicated that 72% of Internet users have looked online for health information of one kind or another within the past year. Three quarters of these searches began at a search engine such as Google or Bing (Fox and Duggan, 2013).

This demand for online health information highlights the importance of information being derived from credible sources. While the technological capabilities to create online resources are becoming increasingly easier, developing high quality content that is readily accessible draws upon a complex range of skills and knowledge. The types of skills and knowledge required can be most readily seen in the US usability guidelines (US Department of Health and
Human Services, 2006) or the Australian Government’s web publishing guidelines (Australian Government, 2012). An integral part of determining the “value” of a website is to formally investigate the resource in a structured, purposeful manner. This form of investigation, or evaluation, can support design integrity, successful development and delivery, and appropriate modification and recognition of the website (or other form of online information dissemination).

Patton has described evaluation as:

The systematic collection of information about the activities, characteristics, and outcomes of program services, policy or processes, in order to make judgements about the program or process, improve effectiveness, and/or inform decision about future development (Patton, 2008, p. 27).

Evaluation activities in the online environment are particularly important for enhancing the functionality of resources for users and maximising the contribution of online information to outcomes in the health system. By doing so, evaluation can demonstrate the value of the online information and the delivery platform to policy makers, funders and health organisations. Nevertheless, there are challenges in evaluating online health resources. These challenges include different users having different requirements and perspectives on what is a successful outcome (Pawson et al 2011, Greenhalgh & Russell 2010). In addition, there are human, organisational and technical factors relating to system design, development and use that may also impact on appropriate evaluation (Yusof, et al. 2008; Pagliari, 2007, Catwell &Sheikh, 2009).

Regardless of these obstacles, however, there is an increasing recognition of the value of such studies both to the individual resource and to the body of knowledge that supports online resource development (Eysenbach, 2011). However, the landscape of approaches used in evaluating health information websites during design, development and production remains unclear. This systematic review maps the different evaluation approaches being used to assess the effectiveness of healthcare websites and, in so doing, provides baseline knowledge not only of the evaluation methods being used but also of aspects of website and information development that appear to be under-evaluated.

Methods (n=416)
The study was conducted between January and April 2012.

A literature search of the Ovid SP database was conducted on 7 February 2012. The search comprised two constructs: website (Internet/ OR website$.mp.) and evaluation (Evaluation Studies as a Topic/ OR "Outcome and Process Assessment (Health Care)="/ OR assessment.mp. OR "Outcome Assessment (Health Care)="/ OR Quality Indicators, Health Care/ OR Quality Assurance, Health Care/ OR “Quality of Health Care”/ OR quality.mp. OR Quality Improvement/). The following limits were applied: English language, yr=1995-current. Retrievals were restricted to the following Publication Types: evaluation studies OR case reports OR clinical trials OR comparative study OR meta analysis OR randomized controlled trial. The literature search retrieved 1,611 references that were downloaded into the electronic reference manager Endnote for screening.

Included articles needed to meet the following inclusion criteria: discusses a website; describes a study; the study relates to evaluation; and the website provides healthcare information. Exclusion criteria were: not being an online resource; not being a website (social media platforms were excluded); describes online healthcare professional education; not being a study; not relating to health; or the article not being able to be retrieved.

Abstracts of the retrieved articles were screened by a research assistant (SB) against the inclusion and exclusion criteria. Three sets of randomly selected articles (n=70) were screened by a second rater (JT) to determine inter-rater reliability. Inter-rater reliability ranged from 76% to 92% for the three sets. After each exercise, the two raters met to discuss the sources of variability and refine decision making.

Two hundred and forty articles met the criteria for inclusion. A flow chart of retrievals, exclusions and included studies is outlined in Figure 1.

The Data Extraction Form comprised a set of commonly recorded items such as author and citation details, abstract, name of online resource and web address (where provided), healthcare content area, intended audience, year study conducted, description of the evaluation, methods of data collection, and details of any statistical analysis undertaken. In addition, four specific review categories were developed.
They are described below. A trial extraction of data from a random selection of 10% of included studies was jointly undertaken (JT, SB) to finalise the Data Extraction Form and to determine its workability. Weekly meetings were held across the project allowing review of coding issues in data extraction.

Four additional extraction categories were created to provide a further basis for organising the included studies:

Website Type: This categorisation indicated whether the included study looked at one or more websites and whether it was evaluating the same or different content areas at a single or multiple time points.

Measurement Focus: This item categorised the included study by using the focus of measurement reported in the article. The categories represented specific components of evaluation activity and focus across the design, development, implementation, and maturity phases of a website's life cycle. The initial categories were developed from reports in the literature that covered developmental and technical perspectives and user needs (Cunliffe, 2000; Calero et al, 2005; Elling et al, 2007; Tankelevištė & Damasevičius, 2009). In a joint data extraction exercise, the categories and descriptors for the Measurement Focus items were expanded, refined and documented. The categories reflect the outcome measures being examined in the evaluation method (e.g., usability testing, web metrics, behaviour change, and economic assessment).

Stage of Development: This categorisation relates to phase of website development with 3 pre-release categories and 3 post-release categories. The Australian Government Information Office’s range of Better Practice Checklists & Guidance materials highlights the need for consideration of the lifecycle of a website from concept to decommissioning (Australian Government Information Office, 2013).

Evaluation Emphasis: This categorisation utilises an evaluation framework developed to guide evaluation activities for the CareSearch website (www.caresearch.com.au). The framework was developed following a program logic activity to outline the role of evaluation within the development of the CareSearch website (Tieman & Martin, 2009). The three levels of emphasis relate to enhancing access, measuring use, and assessing usefulness or impact.

Descriptive statistical analysis was conducted using SPSS, Version 19.0.01 (2010).
Results (n=262)

Results of the review found that over half of the studies evaluated a particular group of websites on a single topic (e.g. breast cancer information) at single time point (59.2%, 142/240). Forty-four studies compared the performance of a group of websites across different healthcare topics (18.3%) and ten studies looked at a particular group of websites (e.g. mental health advice) at different time points (4.2%). Fifteen studies looked at one website at a single time point (6.3%) and a further 31 looked at a single website at multiple time points (12.9%).

Cancer, mental health, reproductive health and paediatrics were the most common health content areas of the websites included in evaluation studies.

When appraising the measurement focus of the evaluation studies, assessment of content quality areas such as accuracy and currency of the information (68.8%, 165/240) and structural elements such as menu systems, navigation or hyperlinks (43.3%, 104/240) were the two most commonly assessed areas as can be seen in Table 1. Project evaluation and economic assessment were the least common types of measurement focus.

Most of the studies in this review were conducted after the release of the designated website (80.8%, 194/240). Approximately 30% looked at studies prior to release during the concept analysis/needs analysis phase (14.2%, 34/240) or the development phase (17.5%, 42/240) as illustrated in Table 2.

Though more than 80% of the studies (199/240) addressed issues relating to enhancing access to the website, less than a third attempted to assess the usefulness of the resource in terms of whether it had actually changed practice or outcomes (Table 2).

Discussion

Throughout the screening and analysis, it became clear that the evaluation methods being described were extensive and variable. Understanding the range and nature of approaches being used, and possible issues associated with implementing particular
approaches, is important to ensure that the most appropriate types of studies are undertaken to meet different evaluation purposes, and that they reflect the different stages of website development. Eysenbach (2011) has highlighted the need to improve and standardise the quality of reporting in this field to facilitate the use and dissemination of this published research.

The significant number of articles reporting on the findings of evaluations of health information websites suggest that evaluation of online resources is feasible and of interest. Even though there was considerable diversity in terms of the focus of the measurement of the studies, very few studies addressed the economics of online health information provision, or the costs and benefits of information transfer within this environment independently or compared to other forms of information transfer. Most studies also described work that was conducted at a single time point limiting the ability to assess the impact of changes to individuals or to websites over time. Studies around the assessment of the quality of the information content provided in the website and readability levels, reports on changes to structural aspects of the website and issues in search and search engine retrieval were much more commonly published than studies addressing behaviour change or knowledge transfer. While ensuring functionally accessible websites is extremely important, more emphasis is needed on assessing the impact that engagement with these online resources has on individuals and on the health system.

Relatively few studies focused on program evaluations. Hence, there is only a limited amount of published material available to web developers and project managers in assessing the performance of the online resource against specified criteria from funding agencies or policy makers. The apparent paucity of evaluation research in this field may also reflect the reality of decision making processes involved in committing to creating an online resource which may restrict the time and capacity to articulate and define the purposes and intended contributions of the resource to health outcomes. For groups commissioning or creating websites or webpages, the review would indicate that more needs to be done in articulating desired outcomes of the project not just outlining the product specification. Partners in web planning and development need to consider who the intended users are, the capacities and experience of these users, and the circumstances in which they will seek, find and use the provided health information. The information needs of the intended audience should guide planning and inform decision making about formats, presentation, design and navigation to enhance knowledge transfer and knowledge use.
There were some indications that different methods have been used at different stages of the website development process. User testing activities and usability studies were reported in a number of studies suggesting that formative evaluation prior to launch has been seen as a valuable aspect of the web development process. However, it is worth noting that although there were many studies looking at issues around the quality and accessibility of the online content, very few explicitly explored issues around access for marginalised groups such as online options for non-English speaking groups or enhancements for use of information by intellectually disabled groups. Previous research on a "digital divide" highlights the importance of not only providing resources in appropriate forms but of facilitating and supporting access by the whole community (Choi et al, 2013; Kruse et al, 2012; Reinfeld-Kirkmann et al, 2010). Given that many see online information provision as a remedy that enables equitable distribution of health information, this remains an area for further study.

Most post-release studies reported on visitor numbers and usage statistics, or provided the results of visitor surveys and user satisfaction scales. However, there was often little interpretation of the possible meaning of these metrics other than as trend indicators of use. The possibility of web metrics acting as surrogates of individual and health system actions needs to be explored. Further research around the meaning of usage patterns could add great value to these readily available metrics. For example, commercial enterprises will assess the relationship between product views and orders, and then use these web metrics to assess the impact of marketing strategies or product releases. Identifying and evaluating possible relevant metrics for health information would be a valuable piece of research. For example, does time of use correlate with different environmental circumstances for users such as no colleagues available for advice during night duty? Or what number of page views correlate to actual visitor action such as booking a GP appointment? Glyn et al (2012) have already used a web metric system, Google Insights, to show a relationship between an annual breast cancer awareness campaign and online breast cancer activities. Similarly, comparative data on usage rates and penetration are needed for health websites to provide background information against which to assess the effectiveness of strategies and approaches used in website development and delivery.

It is interesting to note that many of the evaluation studies were not undertaken by the agencies responsible for the online resource. This suggests that the study of online health information has become an area of research interest in its own right. For many health agencies, assessing the quality of online resources is important as patients,
carers and families are using this information for self-diagnosis, decision making about treatment options or as part of their engagement with health professionals (Boucher, 2010; Sillence et al 2007; McMullan 2006; Rubenstein, 2012).

Finally, this review developed several evaluation variables to differentiate between the focus measures being used, the website types under review, the stage of website development and the actual emphasis of the evaluation. This highlights the need for a common language to describe not only the design characteristics of evaluation studies but the contribution of the evaluation research in terms of stage of development and the focus of the evaluation with respect to the information users, funders and the health system.

Limitations to this review

This review only included articles published after 1995 from a single biomedical bibliographic database. It is likely that there are substantial numbers of unpublished reports and conference presentations looking at evaluations of health websites. It should be noted that no quality assessment of the individual studies was undertaken. The focus of the analysis was on the purpose of the study not on the conduct of the study.

This study employed several evaluation schemas developed for the review. These categories have not been independently validated and assessed as website evaluation scales; so, additional examination of the value of these schemas is warranted.

This study only looked at health information provided through a website. The reality is that the web is no longer a vast library of web pages accessed through discrete websites; it is a complex mix of information sources and formats, online interfaces, searching tools and brokers, and participation and management gateways. It is unknown whether similar approaches to evaluation can be used for other types and forms of online information dissemination and exchange.

Conclusion

Healthcare information is no longer the providence of the local doctor or nurse. Just as the nature of medical technology has changed, so too has provision of healthcare information. The online environment has changed the ways in which health consumers and health professionals seek and engage with health information, but our understanding of how effectively information is being provided and used through this medium is still limited. There is a need for further research that looks beyond the creation and access of
health websites to the impact that health websites have on outcomes for health consumers and their effects on health professionals and health services.

Evaluation activities and studies of evaluation processes are an essential part of the process of understanding the contribution these resources can make. Evaluation activities undertaken during planning and development can assist in developing accessible and usable websites. Evaluation undertaken following release of a website can help not only to demonstrate use of these online resources but also help to assess the effect of these online resources on individuals and, potentially, on services and organisations and the health system. Commissioning agencies need to ensure that web developers and content providers are aware of best practice requirements and encourage research to direct the effective preparation and use of web-based healthcare information.

Acknowledgements: The authors would like to acknowledge the contribution of Amanda Adams who created the SPSS file and was responsible for data entry, data cleansing and data outputs. Dr Tieman and Ms Bradley are part of the CareSearch Project staff. CareSearch is funded by the Australian Government Department of Health and Ageing.
References (n=457)

Australian Bureau of Statistics (2012) Internet Activity, Australia, (Australian Bureau of
Statistics: Canberra) Available at
http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/8153.0Main+Features1Jun%202012
[Verified 21 February 2013).

2013).

Australia’s First National Primary Health Care Strategy Department of Health and Ageing
Available at http://www.yourhealth.gov.au/internet/yourhealth/publishing.nsf/Content/report-
primaryhealth/$File/NPHCS-Foreword-intro.pdf [Verified 21 February 2013]

Australian Government Information Office (2013) Better Practice Checklists & Guidance
Available at http://agict.gov.au/policy-guides-procurement/better-practice-checklists-
guidance/ [Verified on 27 July 2013]

Boucher J (2010) Technology and Patient-Provider Interactions: Improving Quality of
Care, But Is It Improving Communication and Collaboration? Diabetes Spectrum
23(3):142-144

Online Information Review 29(3): 227-248

Systemic Evaluation PLOS 6(8):e1000126

Choi NG, Dinitto DM (2013) The digital divide among low-income homebound older
adults: Internet use patterns, eHealth literacy, and attitudes toward computer/Internet
use. Journal of Medical Internet Research 15(5):e93

Cunliffe, D. (2000) Developing usable websites- a review and a model Internet Research
10(4):295-308


a Research-Based Tool for Evaluating Informational Websites Lectures in Computer
Science. 4656:293-304

Eysenbach G, Kohler C. (2003) What is the prevalence of health-related searches on the
World Wide Web? Qualitative and quantitative analysis of search engine queries on the

Eysenbach G. (2011) CONSORT E-HEALTH: Improving and standardizing evaluation
reports on web-based and mobile health interventions. Journal of Medical Internet
Research 13 (4), 126.


Figure 1: Schematic representation of articles screening

Articles Retrieved: N=1,611

Duplicates: N=13

Articles after De-duping N=1598

Excluded: N=1,358

Included Articles: N=240
<table>
<thead>
<tr>
<th>Measurement Focus</th>
<th>No.</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feasibility</td>
<td>29</td>
<td>12.1%</td>
</tr>
<tr>
<td>Formative evaluation identifying need, consideration of audience, items for inclusion etc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heuristic</td>
<td>22</td>
<td>9.2%</td>
</tr>
<tr>
<td>Systematic inspection of a user interface design for usability by an expert</td>
<td></td>
<td></td>
</tr>
<tr>
<td>User testing</td>
<td>32</td>
<td>13.3%</td>
</tr>
<tr>
<td>Feedback of the prototype website by intended users</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text content assessment</td>
<td>56</td>
<td>23.3%</td>
</tr>
<tr>
<td>Includes readability assessment, literacy testing, text analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content Quality</td>
<td>165</td>
<td>68.8%</td>
</tr>
<tr>
<td>Studies of the accuracy, currency and quality of the website content. Can include automated assessment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structural Elements</td>
<td>104</td>
<td>43.3%</td>
</tr>
<tr>
<td>Measures structural elements of the website such as navigability, menu systems, hyperlinks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visuals/Graphic Identity</td>
<td>47</td>
<td>19.6%</td>
</tr>
<tr>
<td>Studies looking at presentation of the website (e.g. inclusion of high quality pictures)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metric Analysis</td>
<td>37</td>
<td>15.4%</td>
</tr>
<tr>
<td>Retrieval and analysis of site metrics such as visitor numbers, referrals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Search engine optimisation</td>
<td>61</td>
<td>25.4%</td>
</tr>
<tr>
<td>Studies that assess the effectiveness of page tagging, search term analysis etc that lead a user to the website</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visitor Satisfaction Surveys</td>
<td>25</td>
<td>10.4%</td>
</tr>
<tr>
<td>Online/offline surveys of satisfaction with the resource</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge Transfer</td>
<td>29</td>
<td>12.1%</td>
</tr>
<tr>
<td>Studies that assess that whether access and engagement with the website has led to an increase in knowledge or understanding by the web visitor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behaviour change</td>
<td>32</td>
<td>13.3%</td>
</tr>
<tr>
<td>Studies that assess whether visitor ‘s health behaviours have changed due to engagement with a website (eg stopped smoking, anxiety reduced etc)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project Evaluation</td>
<td>7</td>
<td>2.9%</td>
</tr>
<tr>
<td>Assessment by funders, policy makers of the value of their website project</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Economic Assessment</td>
<td>7</td>
<td>2.9%</td>
</tr>
<tr>
<td>Cost benefit analysis, economic analysis, cost pricing of an individual website</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 2: Number and percentage of studies by stage of website development and by evaluation emphasis

<table>
<thead>
<tr>
<th>Stage of Website Development</th>
<th>No</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept Analysis, Needs Assessment</td>
<td>34</td>
<td>14.2</td>
</tr>
<tr>
<td>Development Phase</td>
<td>42</td>
<td>17.5</td>
</tr>
<tr>
<td>Release/Launch</td>
<td>8</td>
<td>3.3</td>
</tr>
<tr>
<td>Post Release Effectiveness</td>
<td>194</td>
<td>80.8</td>
</tr>
<tr>
<td>Iterative Enhancements</td>
<td>6</td>
<td>2.5</td>
</tr>
<tr>
<td>Redesign</td>
<td>5</td>
<td>2.1</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>0.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Evaluation Emphasis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access (Facilitating the ability of users to be able to access the resource)</td>
</tr>
<tr>
<td>No.</td>
</tr>
<tr>
<td>Use (Tracking if, and how, the resource is being used)</td>
</tr>
<tr>
<td>No.</td>
</tr>
<tr>
<td>Usefulness (Addressing whether the resource has made a difference to use or practice)</td>
</tr>
</tbody>
</table>