Analyzing and evaluating websites: locating public health within an international multi-disciplinary review of published research

Abstract

**Background:** Little has been written on the types of public health and wider academic research published about the analysis and evaluation of websites. This paper investigates the subject with the aim of highlighting the gaps and opportunities for future research. In so doing, it not only informs academics of potential research areas, it also provides valuable information for undergraduate and masters dissertation students, and their supervisors, on an important source of accessible data.

**Methods:** We undertook a systematic search of published research identifying articles from 2000 to 2017. Articles were eligible for inclusion if their title contained the words website* and analysis, or website* and evaluation. A second, more detailed investigation was conducted on articles in the same sample for 2015-2017. A third round of investigation reviewed 10 articles in more depth, based on those from the highest-ranking journals.

**Results:** We demonstrate that research specifically about websites has increased significantly since the year 2000. Contrary to recent social media-focused research, in our research, health-related articles were found to be the largest group in the field. Research about websites focused on issues such as information accuracy and corporate social responsibility. Other health promotion and public health areas of international concern, such as ‘sustainable development goals’, were less prominent.

**Conclusion:** It is recommended that future research includes a focus on recognised international health promotion and sustainable development priorities.

**Keywords:** Websites, Public health, New Media, Dissertation, Pedagogy
Introduction

Neglected New Media?

Website analysis has the potential to provide insights into the operation of powerful organisations and the thinking of bodies across the statutory, voluntary and for-profit sectors. Important opportunities are available for international comparative research. From a public health perspective, websites can be analysed for health promotion, health services and health protection content. While research using social media data has gained widespread academic attention and there have been analyses on using the internet as a research tool, less is known about trends in research about websites, which is the focus here.

A recent review of social media research anticipated a “dearth of studies specific to health-related research” and indeed Golder, et al. (2017) uncovered only limited research in the health field. On the other hand, previous analysis surveying the use of the internet for research, which, again, is not our focus, found higher rates of usage in health sciences, with the social sciences lagging behind (Hesse-Biber and Griffin, 2012). Therefore, it is apposite to study the distribution of disciplines undertaking research about websites.

This paper investigates the extent and characteristics of website analysis being undertaken by researchers internationally. In doing so, it highlights for academics some potentially untapped research areas.

In addition, it provides a valuable context for dissertation students and their supervisors. Much has been written in the pedagogical literature on different aspects of dissertation supervision. Articles have covered areas from ‘how first and second year undergraduate learning is used in the dissertation’, how to supervise dissertations online, and feedback on drafts, through to dissertation supervision in different academic disciplines, from geography to psychology (Ashwin, et al., 2017; Ross and Sheail, 2017; Basturkmen et al., 2014; Ginn, 2014 and Castello, 2012).

However, this paper provides analysis that aims to highlight for dissertation students, supervisors, as well as other academics, a relatively neglected data source.

Background

The internet is constantly developing. In 2016, it was estimated that there were over one billion websites (Hine, 2016). Of these, most are commercial (Webster, 2014).

What is different about websites?

In contrast to research about websites, the literature on social media research is extensive. Zeller (2017) highlights that social media can be both an object of research and a means of conducting research. Websites also have a dual purpose in research. Obtaining information is one use. But it is websites as a subject of analysis that we are considering here.

Website data are categorised as “non-reactive” and “found-data”, as distinguished from “made-data” (Zeller, 2017). Websites are ‘go-to’ sources for the public and patients that typically represent organisations and they also disseminate information. Thus, because readers seek them, they have addresses, or uniform resource locators (URLs). Nevertheless, Web 2.0 websites are also fast changing and dynamic in content, interactivity and uses.

Important literature on e-government, for example, Twizeyimana and Andersson (2019) and Wirtz and Daiser (2018), and ‘the internet of things’, for example, Nord, Koohang and Paliszkiewicz (2019) also relates to these developments.

There are a number of attractions and justifications for analysing and evaluating websites. The advantages and researchable areas that websites offer are summarised as follows. They provide:

- quick, inexpensive, access to desk research data
- international data (in future, international access may be further
enhanced by developments in translation software (Fantinuoli, 2018))

• lower-carbon, ‘aeroplane-free’, data when geographically disparate research material is sought (while the energy consumption of the internet is high (Cook et al., 2017) some calculations have suggested that it is more energy efficient than air travel (Baliga, et al., 2009))

• a ‘public face’, image, or ‘shopfront’ of organisations and a gateway to their activities, ideas and values

• insight into what the public and patient audiences see

• data subject to particular codes, such as child safety (UKCCIS, 2018).

However, there are also hazards involved in this research. Websites can:

• change, and pages need to be downloaded and referenced

• be difficult to find. It was estimated in 2013 that search engines’ reach was to “search (and thus retrieve from) only about 30% of what exists on the internet” (Krippendorff, 2013)

• have paywalls that prevent access

• provide data requiring benchmarking, corroboration or triangulation of findings. This may impact on the ease of data access

• have different levels of activity, or inactivity

• deliver unequal access to website content from across social groups

• be subject to state censorship, nevertheless, they have also been linked to anti-corruption campaigns (Poell, 2015)

• raise uncharted ethical issues—although research ethics guidance is developing (Zeller, 2017).

Categorising websites demonstrates their research potential in wide-ranging areas. They can be categorised as those for: government; public services (including transport); business (including corporate and multinational, and such areas as shopping, tourism, entertainment, and service industries); campaigning and voluntary bodies; individuals; and others.

Websites are used for: (i) buying and selling; (ii) finding information (including: availability of resources, campaigning, connecting people together, emergency information, entertainment, fundraising, news, research, sexual gratification, therapeutic and health-related, travel); (iii) delivering communications, ‘sounding-off’, being seen to speak; (iv) providing information through links, such as e-government tax return activities; (v) illegal activities (e.g. fake sites stealing information or money). Sub-categories of website content include downloadable documents, blogs, information on the producer, links and screen information (Bryman, 2012).

What can be researched about websites?

There are distinct areas of research within this field. A typology of areas that ‘research about websites’ can cover is provided in Table 1. The first area – content – has various related categories, which are listed separately within this point (i) here.

Thus, we have set out a series of distinct areas that can be researched about websites. Additionally, questions for website evaluation may cover: quality of content, adherence to codes of conduct and the impact of limited codes of conduct, gaps in data, underpinning assumptions, ownership, direction and navigation critiques.

Website analysis

Consideration of the methods used for research about websites can also help us to scrutinise the current research picture and potential research gaps.

Given the ‘found data’ nature of website material, the key methods we might expect...
to see in website analysis relate to textual and visual analysis and investigation of quantitative ‘found data’. Therefore, content, discourse, visual and linguistic analysis are likely to dominate. We anticipated that the methods devised in our sample of articles would aim to address the research areas set out in Table 1, including, for example, rows 5 and 8 i.e. audience research and computer-based technical design issues (Webster et al., 2014). Experimental and longitudinal methods could also be expected.

### Research questions
We have identified that social media and internet research reveal uneven distributions of publications by discipline (Golder, et al., 2017; Hesse-Biber and Griffin, 2012). Thus, we start by asking; what is the balance of papers by academic discipline in our data covering publications on research about websites?

<table>
<thead>
<tr>
<th>Table 1. A Typology of Research About Websites</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Topics investigated in research about websites</strong></td>
</tr>
<tr>
<td>1.</td>
</tr>
<tr>
<td>2.</td>
</tr>
<tr>
<td>3.</td>
</tr>
<tr>
<td>4.</td>
</tr>
<tr>
<td>5.</td>
</tr>
<tr>
<td>6.</td>
</tr>
<tr>
<td>7.</td>
</tr>
<tr>
<td>8.</td>
</tr>
<tr>
<td>9.</td>
</tr>
<tr>
<td>10.</td>
</tr>
<tr>
<td>11.</td>
</tr>
<tr>
<td>12.</td>
</tr>
<tr>
<td>13.</td>
</tr>
<tr>
<td>14.</td>
</tr>
</tbody>
</table>
The research methods used in the papers identified can divide into those that are only analysing data from websites and those that triangulate with other data, such as interview data. Some may use comparative methods, drawing on many websites, and others concentrate on one example. We will consider the extent to which these and other methods, referred to in the previous section on ‘website analysis’, appear in our sample.

For this paper, we wanted to investigate both the extent and nature of research about public health-related websites. We considered high-level international priorities set by the WHO, incorporated into the sustainable development goals (SDGs) (WHO, 2018) and also adopted by national agencies, for example, Public Health England (2016). These priorities include action on climate change. Thus, a question arises: are these priority SDG issues being covered by published research about websites?

Added to this, the categorisation of websites in our background review includes ‘government’ as a high-level website type. A question springs from this, as follows: are government, local government and statutory websites underrepresented in the sample?

Methods

We used a three-stage investigation to gain an overview of the field and assess patterns in website research. These stages were, firstly, a broad sweep of 17 years’ research from 2000 to 2017, secondly, a more focused review of articles from 2015 to 2017 and, finally, following Bryman (2012), whose motivation is to support researchers by providing some case studies, a yet more detailed review of 10 articles from these latter three years.

To aid reproduction and future longitudinal comparisons, and because the numbers of journal articles found via further databases were of the same order of magnitude, it was decided to focus only on a Zetoc database search. This is a multi-disciplinary database that covers all journals listed by the British Library. As such, it is also appropriate because it is not health-biased. The word search criteria were: ‘website*’ AND ‘analysis’ [in Title] as well as, ‘website*’ AND ‘evaluation’ [Title]. These were the words most likely to produce the largest number of relevant results. The keywords ‘Internet’, ‘web’ and e-government’ were not used as this would have attracted too many social media and service delivery articles. Articles were restricted to English language copy. A total of 744 articles were found. These were from 2000 onwards.

For the second research element, three years’ data, from 2015 to 2017, were chosen from the wider group of 744. The sample of articles generated (135) was sufficient to gain a clear insight into patterns of website research. On the 135 articles we recorded the:

- name and H-score/SJR of journal
- primary academic field of the research
- secondary research fields
- methods, including types of questions asked
- types of website being analysed
- number of websites looked at
- extent of international comparison
- country of focus.

The aim of the third stage was a more detailed analysis of the standards and features of high-quality website research through case studies (Bryman, 2012). We selected 10 articles from 2015-2017 with the highest H-index and Scimago Journal Rankings (SJR) (www.scimagojr.com). Overall, the majority of articles in the wider sample did have an SJR ranking. Only 12 out of 94 ‘website* and analysis’ articles were from unranked journals. In the ‘website* and evaluation’ sample, 84% of papers were from SJR ranked journals.
Individual papers from non-SRJ ranked journals may also be of high quality.

Findings

Figure 1 demonstrates the growth in this research area since the year 2000. The number of papers published per year on website analysis and evaluation is provided.

Investigation of the articles for the years 2015 to 2017 shows that 42% of all website analysis articles were on health-related subjects, with only 5 out of 94 (5%) on media or politics topics.

The ‘website* and evaluation’ search indicated a similar pattern of journal areas. The business-orientated journal websites also included some health-related areas, such as, for example, an article on health tourism websites in three countries (Moghavvemi et al., 2017). Other business websites analysed included those from the following industries: food companies, drug companies, retail, banks, luxury brands, the sex industry and auction websites. Tourism was the single largest business website group to be analysed, with eight articles focused on tourism websites (29% of the 28 business articles).

Most of the articles that analysed website content considered more than one website, but around 12% focused on one website only. This percentage increased slightly in the ‘website* and evaluation’ sample, where some articles described the setting up and evaluation of specific sites.

The international orientation of the research was perhaps lower than might be expected, given the opportunities afforded by research about websites. Although, over 21% of articles had an explicitly international focus – looking at websites from a range of countries. In addition, some articles looked at multinational companies’ websites hosted in one country. One example looked at corporate social responsibility (CSR) issues among banking websites in Ghana (Boateng, 2016). Analysis of African and Latin American websites was generally limited. Often websites from two or three countries were compared, although a number of articles, particularly on tourism, extended the geographical range.

In considering our first research question, the single largest number of articles were on health issues and, within these, the majority came from a public health perspective. The health articles were concerned with issues such as the accuracy of the information provided by websites. Of the 39 health journal articles in the ‘website* and analysis’ sample for 2015-2017, 28% were published in health
promotion or public health journals. However, 10 came from journals focused on one medical condition or specific surgical or medical procedure or issue. The types of articles about websites in these issue-specific journals were also often concerned with the quality of website content. Psychiatry, or mental health, appeared, well represented, and indeed a number of nutrition-focused articles could be classed as coming under ‘mental health’. This is an area where we might expect to see high research interest given the potential of the internet to help therapeutically, but also to disrupt mental health, through online bullying (Modecki et al., 2014). Physician-rating websites were analysed by three articles in the 2015-17 sample. However, in answer to our last two research questions, analysis of government and local authority websites was minimal, as was the focus on SDGs and, specifically, environmental concerns.

With reference to our second research question, the types of analysis employed were wide-ranging. Two articles in the ‘website* and evaluation’ sample included ‘systematic review’ in their title (Rogers, et al., 2017; Rangraz et al., 2017). Among the evaluation articles, some discussed and put forward criteria for website evaluation (Rangraz et al., 2017). Within the ‘health’ subset of the ‘website* and analysis’ sample, as has been mentioned, a substantial proportion of articles were involved with assessing the quality of information provided on websites. This was often benchmarked against external standards (McDaniel et al., 2016). Readability was also a concern of the

Table 2: Articles in three-year sample by academic discipline of journal, ‘website* AND analysis’ and ‘website* AND evaluation’ searches

Note: ‘a’ in column headings stands for analysis and ‘e’ for evaluation

<table>
<thead>
<tr>
<th>Journal discipline</th>
<th>2017a</th>
<th>2017e</th>
<th>2016a</th>
<th>2016e</th>
<th>2015a</th>
<th>2015e</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health including:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>health promotion</td>
<td>15</td>
<td>11</td>
<td>14</td>
<td>5</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>specific medical conditions/surgical area</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>psychiatry</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>physician rating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Business including:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tourism, HR, advertising, marketing, retail</td>
<td>10</td>
<td>1</td>
<td>13</td>
<td>6</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Information and Communications Technology</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Media</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Politics</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Other including:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017 linguistics and criminology</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2016 sociology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015 linguistics and visual communication</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duplicate (and 1 not relevant)</td>
<td>4</td>
<td>3</td>
<td>9 (1)</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>35</td>
<td>25</td>
<td>47</td>
<td>13</td>
<td>28</td>
<td>13</td>
</tr>
<tr>
<td>Total (Excluding duplicates)= Analysis 94</td>
<td>31</td>
<td>22</td>
<td>39</td>
<td>9</td>
<td>24</td>
<td>10</td>
</tr>
<tr>
<td>Articles with no SJR ranking</td>
<td>8</td>
<td>3</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
‘health’ subset. The ‘business’ subset was also concerned with attractiveness of websites and navigation attributes. A number of articles in this group compared websites with each other and assessed which were best for users.

Few articles attempted to assess the influence of websites on public opinion and target audiences (Rodgers et al., 2016). Some articles categorised provision of sites to support analysis of their sector. More articles provided different forms of content analysis than assessed website user behaviour or user views, that is, ‘sentiment analysis’. This was to be expected as these latter terms are more associated with social media research.

Forms of content analysis were used in 17 of the 2015, 24 or the 2016 and 18 of the 2017 ‘analysis’ sample. Some articles took a longitudinal sample (Smith, 2017). Others focused on analysing the documents provided on the websites. To give an additional perspective on the types of articles published about websites, we provide an in-depth look at 10 cases in Table 3. This approach follows Bryman (2012) who provides such an analysis of two articles. As can be seen, analysis and evaluation of websites provides a rich source of data for health promotion and public health-related articles taking an international perspective. We have answered research questions concerning the balance of papers by academic discipline, methods used, the priority given to SDG issues and the extent to which government websites are analysed. In so doing, and though our further analysis, we have provided contextual material for those considering research about websites.

**Conclusion**

There has been a rapid growth in website production and associated academic analysis of websites since 2000. Public health issues have been to the fore of the research agenda, driven by concerns that the public may be receiving unsafe or poor-quality health and medical information. There is a trend in analysis seeking to hold corporations to account and demonstrate enforcement of international agreements, such as on product advertising.

There appear to be gaps in the literature and significant public health issues are underserved in the sample, such as critiques of government websites, and analyses of environmental and SDG-related areas. Although just over 21% of articles undertook some international comparison of websites, this figure is perhaps surprisingly low and shows untapped research potential. Even allowing for the search terms used, there was also an unexpectedly low level of research from media and politics disciplines. Thus, we have provided valuable pointers, perhaps to untapped research data. This will be of interest to academics from across a range of disciplines. Not only do websites provide an accessible data source for academics’ own research, but website analysis can also be undertaken by dissertation students. An understanding of this area helps academics in supporting and supervising undergraduate and postgraduate dissertation students.

The opportunities for research that holds power to account were found to be underexploited in this sample. Given the ease with which corporations and governments post on their websites, any lack of information provided by these actors is open to further research scrutiny.

**Acknowledgements**

The authors would like to thank Dr Jean Chapple and Dr Sophie Pettit for their advice on a previous draft of this article. We are also grateful to the anonymous reviewers.

**Competing interests**

The author mentions no competing interests.
### Table 3. Case sample of 10 articles

<table>
<thead>
<tr>
<th>Author, year, title, Journal, H-index/SJR for year</th>
<th>Type of study, Purpose</th>
<th>Sample, Design, Data collection</th>
<th>Key findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ruiz and Barnett 2015 Exploring the presentation of HPV information online: a semantic network analysis of websites Vaccine 148/1.86</td>
<td>Computer-based semantic network analysis. Assesses positive and negative language used about HPV vaccine.</td>
<td>223 websites. Uses ConText software and network statistics to identify clusters of words.</td>
<td>Negative words are most common, describing vaccine side-effects, HPV and cancer. Benefits as well as risks are described. Information presented in more sexual terms for women/girls than for men/boys.</td>
</tr>
<tr>
<td>McDaniel et al 2016 Shared vision, shared vulnerability: a content analysis of corporate social responsibility [CSR] information on tobacco industry websites Preventive Medicine 148/1.81</td>
<td>Review of CSR activity and content. Downloaded content and categorised CSR content. Use of Nivio. Reviews website text that links to the CSR focus.</td>
<td>8 websites from 4 US and 4 multinational tobacco companies. Compares content to WHO guidance.</td>
<td>CSR activities in 58 countries mentioned. WHO treaty banning activity needs to be enforced.</td>
</tr>
<tr>
<td>Liewers et al 2016 A retrospective analysis of real-world use of the eaTracker ® My Goals website by adults from Ontario and Alberta, Canada BMC Public Health 103/1.34</td>
<td>Analysis of use of nutrition and physical activity website. To characterize My Goals user demographics, types of goal set and My Goals tracker use.</td>
<td>Use of one website in one country is analysed. Secondary analysis of anonymous data on goals set. University ethics approval. Use of SPSS. Analysis of 6,460 users goals. Use of tracker and quality of goals assessed.</td>
<td>High drop-out rates found. ‘Write your own goals’ were often of poor quality. More support for users needed.</td>
</tr>
<tr>
<td>Berry and Gribble 2016 Health and nutrition content claims on websites advertising infant formula available in Australia: a content analysis Maternal &amp; Child Nutrition 4/1.56</td>
<td>To assess compliance against international standards. Coding frame aligned with national food standard code.</td>
<td>Australian infant formula advertising websites, 8 websites from 6 parent companies.</td>
<td>Extent of claims for nutrition content assessed. Despite codes prohibiting claims, there has been a failure to monitor and enforce.</td>
</tr>
<tr>
<td>Yun, Sun and Mao 2017 Growth of Integrative Medicine at Leading Cancer Centers Between 2009 and 2016: a systematic analysis of NCI-designated comprehensive cancer center websites J Natl Cancer Inst Monogr 126/5.49</td>
<td>Review of service provision of integrative medicine therapies.</td>
<td>45 U.S. cancer center websites.</td>
<td>Between 2009 and 2016 there was an increase in service provision and the majority of therapies were provided in the same health systems.</td>
</tr>
</tbody>
</table>

### Non-health Journals

<table>
<thead>
<tr>
<th>Author, year, title, Journal, H-index/SJR for year</th>
<th>Type of study, Purpose</th>
<th>Sample, Design, Data collection</th>
<th>Key findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duen-Ren Liu et al 2015 Complementary QA Network Analysis for QA Retrieval in Social Question-Answering Websites J Assoc Information Science and Tech 118/1.31</td>
<td>Describes research on developing more accurate Q&amp;A links. Addresses problem of information overload and providing accurate responses.</td>
<td>Develops new search algorithms. Tests on example of medical questions (eg I have a headache. How can I cure it?).</td>
<td>Author’s new way of searching Q&amp;A data more effectively identified similar Q&amp;As than conventional methods.</td>
</tr>
<tr>
<td>D’Angella and De Carlo 2016 Orientation to sustainability and strategic position of destinations: an analysis of international tourism websites Current Issues in Tourism 50/1.47</td>
<td>Develops a ‘Green D-web score’ of 35 indicators and applies this to tourism websites to show orientation to environmentally friendly tourism.</td>
<td>Official tourist websites from 77 counties.</td>
<td>Provides a summary of cities with high, moderate and low orientation to providing information about ‘towards green score indicators’.</td>
</tr>
<tr>
<td>Mughavvemi et al 2017 Connecting with prospective medical tourists online: a cross-sectional analysis of private hospital websites promoting medical tourism in India, Malaysia and Thailand Tourism Management 143/3.03</td>
<td>Content analysis across 5 dimensions: hospital information and facilities, admission and services, interactive online services, external activities and technical items.</td>
<td>51 private hospital websites, from 3 countries.</td>
<td>Differences between websites noted, providing opportunities for hospitals to improve their promotion through online presence and interactivity.</td>
</tr>
</tbody>
</table>
References


References


