

Adding dimensions to the analysis of the quality of health information of websites returned by Google. Cluster analysis identifies patterns of websites according to their classification and the type of intervention described.

Article (Accepted Version)

Ghezzi, Pietro and Yaqub, Mubashar (2015) Adding dimensions to the analysis of the quality of health information of websites returned by Google. Cluster analysis identifies patterns of websites according to their classification and the type of intervention described. *Frontiers in Public Health*, 3. a204. ISSN 2296-2565

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/56221/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Adding dimensions to the analysis of the quality of health information of websites returned by Google. Cluster analysis identifies patterns of websites according to their classification and the type of intervention described.

Pietro Ghezzi and Mubashar Yaquab

Journal Name:	Frontiers in Public Health
ISSN:	2296-2565
Article type:	Original Research Article
Received on:	13 Apr 2015
Accepted on:	10 Aug 2015
Provisional PDF published on:	10 Aug 2015
Frontiers website link:	www.frontiersin.org
Citation:	Ghezzi P and Yaquab M(2015) Adding dimensions to the analysis of the quality of health information of websites returned by Google. Cluster analysis identifies patterns of websites according to their classification and the type of intervention described.. <i>Front. Public Health</i> 3:204. doi:10.3389/fpubh.2015.00204
Copyright statement:	© 2015 Ghezzi and Yaquab. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY) . The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

This Provisional PDF corresponds to the article as it appeared upon acceptance, after rigorous peer-review. Fully formatted PDF and full text (HTML) versions will be made available soon.

Adding dimensions to the analysis of the quality of health information of websites returned by Google. Cluster analysis identifies patterns of websites according to their classification and the type of intervention described.

Mubashar Yaqub E-mail: M.Yaqub1@uni.bsms.ac.uk; myaqub80@gmail.com

Pietro Ghezzi* E-mail: p.ghezzi@bsms.ac.uk

Brighton & Sussex Medical School, Falmer, Brighton BN19RY, United Kingdom.

*Corresponding author.

Abstract

Background and aims: Most of the instruments used to assess the quality of health information on the Web (e.g. the JAMA criteria) only analyze one dimension of information quality (IQ), trustworthiness. In this study we analyzed the type of intervention that websites describe, whether supported by evidence-based medicine (EBM) or not, to provide a further dimension of IQ, accuracy, and correlated this with the established criteria.

Methods: We searched Google for “migraine cure” and analyzed the first 200 websites for: 1) JAMA criteria (authorship, attribution, disclosure, currency); 2) class of websites (commercial, health portals, professional, patient groups, no-profit); and 3) type of intervention described (approved drugs, alternative medicine, food, procedures, lifestyle, drugs still at the research stage). We used hierarchical cluster analysis to identify different patterns of websites according to their classification and the information provided. Subgroup analysis on the first 10 websites returned was performed. **Results:** Google returned health portals (44%), followed by commercial websites (31%) and journalism websites (11%). The type of intervention mentioned most often was alternative medicine (55%), followed by procedures (49%), lifestyle (42%), food (41%) and approved drugs (35%). Cluster analysis indicated that health portals are more likely to describe more than one type of treatment while commercial websites most often describe only one. The average JAMA score of commercial websites was significantly lower than for health portals or journalism websites, and this was mainly due to lack of information on the authors of the text

and indication of the date the information was written. Looking at the first 10 websites from Google, commercial websites are under-represented and approved drugs over-represented. **Conclusions:** Analyzing the type of therapies/prevention methods provides additional information to the trustworthiness measures such as the JAMA score, and could be a convenient and objective indicator of websites whose information is based on EBM.

KEYWORDS

Internet, IQ, websites, migraine, information, online.

Background

As more patients search for health information on the Internet, many studies have analyzed the quality of health information available on the web for different pathological conditions. There is a concern that, because the Internet is practically not controlled or regulated, this might expose the public to misinformation, particularly those with low information literacy, which is the ability to critically appraise the information (Berland et al., 2001;Eysenbach and Kohler, 2002). This could potentially result in patients turning to non-approved therapies, whose efficacy (and risks associated) has not been scientifically proven. Furthermore, in recent years there have been an increasing number of commercial websites selling

counterfeit medicines, with additional risks for vulnerable patients (Liang and Mackey, 2009).

Although there are a number of specialized health websites, most patients will use generic search engines, such as Google, to search for health-related information (Fox and Duggan, 2013), and several studies have tried to address the quality of health information available on the Internet using various methods.

The most used are the Health-on-the-Net (HON) seal of approval , the JAMA criteria (Silberg et al., 1997) and the DISCERN criteria (Charnock et al., 1999). The latter two are, in fact, instruments to quantitatively measure the IQ by assigning a score based on the website matching some requirements. In particular, these criteria consider whether a website provides information about authorship, ownership/financial interests, advertising, contact details or date of update (currency). Another parameter that is also evaluated when assessing health information is its readability. For health information to be accessible to lay people, it has to be comprehensible by the average reader, and it was suggested that a patient information leaflet should be aimed at a reading ease of 6th grade in order to be understandable by 75% of the population (Brosnan et al., 2012). Studies have shown that the average reading level of health websites is between 10th and 13th grade (Kaphingst et al., 2006;Thompson and Graydon, 2009), and thus may not be accessible to individuals with lower literacy. Of note, readability and the trustworthiness indicators above do not always correlate, and one study reported that readability might be lower in websites displaying health seals of approval (Kaicker et al., 2010).

All of the criteria for evaluating health websites mentioned above do not take into consideration the content of the website. Therefore, they are mainly evaluating the transparency/trustworthiness of the website rather than the overall IQ, in that they do not consider the content of the site and whether the information provided is scientifically correct and evidence-based.

There are several dimensions of IQ as originally described by Wang and Strong in their seminal study (Wang and Strong, 1996). The study identified almost 200 data quality attributes and 18 dimensions of IQ. This and other studies group the various attributes of IQ in 4 categories of IQ: accessibility IQ, representational IQ (that includes understandability, format, appearance), contextual IQ (including completeness and timeliness), and intrinsic IQ (that include accuracy, correctness and objectivity) (Lee et al., 2002). In the context of health information, probably intrinsic IQ components should include the correctness of the description of the scientific basis of the information described, that would be missing, for instance, in a website reporting that AIDS is not due to a viral infection. Assessment of these intrinsic dimensions of IQ will require an in-depth analysis of the content to be done by a panel of experts. A few studies have analyzed the content of the websites. For instance, Peterlin et al. analyzed the IQ available on cluster headache in 40 websites found using the search engine MetaCrawler (Peterlin et al., 2008). In that study the authors used a scoring system for the technical component that evaluated the correctness of the information on aspects such as epidemiology, risk factors, diagnosis, pathophysiology, treatment, prevention. A similar approach has been used to assess the IQ in 114 websites returned by MetaCrawler or MSN searching for retinopathy of prematurity (Martins and Morse, 2005).

In this study we aim analyzing the type of treatment reported, an alternative and simpler strategy to add an intrinsic quality dimension to the health IQ, beyond the first layer of contextual and representational IQ measured by the JAMA score.

We searched the words 'migraine cure' because we aimed at looking at the information that would be returned from a search done by a layperson. In fact, studies have shown that the quality of information varies with the search term used, that is the more sophisticated the search term, the higher the IQ (Dy et al., 2012;Fabricant et al., 2013). It should be noted, however, that we were more interested in setting up a methodology to study the correlations between classes of websites, type of recommendation and the JAMA criteria as a trustworthiness index, and we did not assign a particular importance to the search terms used.

We selected the search engine Google because it is the most popular search engine , and, as a reference, one specialized health portal (the US Government Medline Plus). We analyzed the first 200 hits and classified the type of websites according to their affiliation (i.e. if they were commercial, professional, health portal, journalism, patient group, non-profit or other). We also analyzed the text of the websites and classified the treatments/interventions described (whether approved drugs, alternative medicine, lifestyle or nutritional advice etc.). We also used an established measure of the health IQ of websites (the JAMA criteria).

Methods

We searched for websites containing information relating to management of migraine. The search was conducted between December 2013 and January 2014, using google.com (US) clearing cookies and using the “private browsing” of Mozilla Firefox browser to prevent, as much as possible, personalised results due to previous browsing history. We are aware that the search results may still be influenced by the detection of the geographical location via the IP address.

Google search found 6,040,000 hits and we considered the first 200. Of those, 198 were accessible (thus excluding non-functioning links; denied direct access through password requirement, payment or subscription; not written in English) and were used in the subsequent analysis. We did not take into account the websites marked "advertisement" that Google shows before the actual search results.

MedlinePlus returned 74 hits that were all accessible and included in the analysis. The flow chart describing how the data were collected and processed is shown in Figure 1. Website links were transferred into a spreadsheet from Google using the SEOquake tool. For MedlinePlus, links were manually transferred into a spreadsheet.

Websites were classified in accordance to their affiliation as being professional, patient group, commercial, health portal, non profit, journalism or others, as summarized in Table 1. Two researchers independently classified the websites and their findings were compared. For those cases where there was no agreement the website was revisited by both researchers and a consensus achieved by discussion.

There were 10 disagreement on 196 web sites (95% agreement). Calculating inter-rater reliability for two coders using Recal2 (Freelon, 2010), showed that the highest disagreement was on “Other” websites (Cohen's kappa coefficient, -0.01; 6 disagreements) followed by “Professional” websites (Cohen's kappa coefficient, 0.76; 3 disagreements), and “Patients Groups” (Cohen's kappa coefficient, 0.86; 3 disagreements). All other classes of websites (Commercial, Health Portal, Journalism, No-profit) showed higher agreement rates, with Cohen's kappa coefficient > 0.93. In the case of Health Portals, there were 5 disagreements but, because of the large number of websites in this class, Cohen's kappa coefficient was 0.95. There was no disagreement on the type of intervention described as that was less subjective as was simply based on an intervention being mentioned in the text.

Websites were then classified according to the different type of intervention they indicated as described in Table 2. One website could mention more than one intervention. If a website gave a link to a different page of the same website, this information was also analyzed. However if the link was given to an external site, this information was dismissed. If a website mentioned an intervention only to state that it is not effective, then the information was dismissed. Therefore, a website was tagged for a type of intervention if either it mentioned the intervention a being an effective one, or listed it as one of the possible interventions without further comments.

We then gave each website a score according to the JAMA criteria (Silberg et al., 1997;Meriç et al., 2002;Barker et al., 2010). For this purpose, websites were analyzed for the following information: 1) authorship (identification of

authors/contributors); 2) attribution (references listing sources of information); 3) disclosure (of ownership, advertising, conflict of interests); and 4) indication of date content was posted or updated. For each of these four criteria, we assigned a score of 1 if the information was present, or 0 if absent or unclear. If the information was not available on the initial website information, then the three click rule was used. The three-click rule is an unofficial website navigation rule that suggests information should be accessible within three clicks (Zeldman, 2001). In previous studies, a website scoring a mean JAMA score of 3 or above has been suggested to be of high quality (Meric et al., 2002;Ghezzi et al., 2014).

The Kruskal-Wallis test was used for multiple comparisons of nonparametric variables, followed by Dunn's test, using GraphPad Prism software (GraphPad Prism Software Inc., La Jolla, USA). The Mann-Whitney test was used where there were 2 independent groups. When indicated, contingency tables were analyzed using a one-tailed Chi-square test for non-parametric data. Hierarchical cluster analysis was performed using the Genesis software (<http://genome.tugraz.at/genesisclient.html>) (Version 1.7.6 for Mac OSX).

Because websites URLs are not permanent, to ensure that the reader will be able to see examples of the search results, the top 10 URLs returned in each of the 5 Google SERPs listed in Supplementary Materials (Table S1 and Table S2) were archived. For this purpose we used WebCite®, an on-demand archiving system for webreferences and the archived URL is provided next to the original URL. Two webpages, all from the domain www.health.com were indicated as not available (n/a) as they could not

be archived, presumably because either the site in question refuses connections by crawling robots, or is inaccessible from the WebCite network.

Results

Distribution of websites

The 198 websites returned by Google were analyzed by their affiliation and the type of intervention they describe. The raw data, with the list of websites URLs and how they were classified in terms of class of websites, type of intervention and JAMA scores is provided in Table S1 in Supplementary Material. As shown in Table 3, commercial websites and health portals made up for 75% of the total, in contrast with the 74 returned by MedlinePlus, where professional websites accounted for nearly 90% of the hits returned.

Table 4 shows the distribution of websites according to the type of intervention described. Of the 198 websites returned by Google, 184 could be assigned to an “intervention” group. However, of the 74 websites returned by MedlinePlus, 70% did not mention an intervention and could not be considered for this type of analysis. The largest proportion of websites returned by Google dealt with alternative medicine (55%) followed by procedures and devices (49%). The total for Google adds up to more than 100% because, many websites mentioned more than one type of intervention.

Cluster analysis of website patterns

We then decided to analyze whether groups of websites could be identified based on the type of interventions they describe and taking into account the fact that most of them will describe more than one type of intervention. The type of data visualization used above does not allow a detailed analysis of co-occurrence of intervention mentioned. Therefore, we have used a graphic representation where each website is assigned a value of 1 for each type of intervention mentioned. Each website is listed in one row and the type of intervention in a column. The value 1 is then represented in red and the visual representation is shown in Figure 2A, where one can appreciate that some websites mention more than one intervention and some only one.

We wanted to see, using this type of visual representation, whether websites mentioned or excluded specific interventions and whether this correlated with the class of website. We thus first performed a hierarchical cluster analysis of the websites according to the intervention they mention, and the results are presented in Figure 2B. In the left part of the figure it can be seen that we can identify clusters of websites that describe only procedure and devices (cluster 1), only approved drugs (cluster 2), only lifestyle (cluster 3), only alternative medicine (cluster 6), and websites that describe all possible interventions (cluster 7) except for the experimental drugs that were mentioned in very few websites. While there were several websites (cluster 8) that described multiple interventions excluding approved drugs, only one website described multiple interventions excluding alternative medicine.

This type of visualization allowed us to then analyze the composition, in terms of the different classes of websites, of these clusters, and the results are shown in the left side of Figure 2B.

The commercial websites preferentially mention only one intervention. In fact, while commercial websites are 28% of the total, they are 75% of those mentioning only nutrition, 73% of those mentioning only alternative medicine, 64% of those mentioning only procedures and devices. However, there are no commercial websites among those mentioning all possible indications (cluster 7), and only 15% in those mentioning all interventions excluding approved drugs (cluster 8). Health portals, which are 48% of the total, were overrepresented among the websites reporting several treatment options (clusters 7 and 8). Journalism websites, which represent 11% of the total, are prevalent among those that only mention “lifestyle” interventions (cluster 3, 63%) and never mention all the possible interventions (cluster 7). It is difficult to comment on the other classes of websites due to their small number. Likewise, a cluster analysis of the MedlinePlus did not provide much information due to the small number of websites, of those returned, that describe an intervention (not shown).

The fact that commercial websites preferentially describe only one type of intervention is also evident by calculating the mean number of different types of intervention described by the different classes of websites (Table 5).

It can be seen that the average number of treatments described by the website is 1 in the commercial websites, followed by journalistic sources with an average of 2, health portals, non-profit and professional websites with an average of 3, while

patient group websites are those describing a wider picture of treatments (an average of 4). Commercial websites describe significantly less treatment options than health portals or patient groups.

Distribution of classes of websites and types of intervention in the ranking of Google search: analysis of the first 10 websites returned.

We wanted to investigate whether the top 10 websites returned by Google were following a different pattern in terms of class of website or type of intervention described. As shown in Table 6, there were significantly less commercial websites (0/10) in the top 10 than in all 198 websites (60/198). No other significant difference was observed for any of the other classes of websites. Thus, commercial websites are underrepresented in the top ten results. Table 6 also shows the occurrence of the different types of intervention in the top 10 websites and in the total number of websites. The only significant difference found was that approved drugs are overrepresented in the 10 top ranking websites.

Analysis of trustworthiness (JAMA scores) of the different classes of websites

In this part of the study, we wanted to assess whether different classes of websites or websites describing different types of interventions differed for their JAMA score, as a recognized measure of website quality/trustworthiness.

Overall, there was a small but significant difference ($P < 0.02$ by Student's t-test) in the average JAMA score of websites returned from Google (1.9 ± 0.8 ; 27% having a score ≥ 3) or MedlinePlus (2.1 ± 0.6 ; 19% having a score ≥ 3). The number of websites

with a JAMA score ≥ 3 was slightly higher with Google (Google, 55/198, 27%; MedlinePlus, 14/74, 19%) but the difference was not statistically significant (Chi-square without Yates correction gave a one-tailed P value of 0.0675). The distribution of JAMA scores was different, and while in MedlinePlus 72% of the websites scored 2 and 19% scored 3, Google had a broader distribution, with 39%, 33% and 27%, scoring 1, 2 and 3 respectively).

We then analyzed the JAMA score of websites returned from Google according to their class and the type of intervention, and the results are reported in Table 7. It can be seen that there are significant differences in the mean JAMA score of different classes of websites, with journalistic websites scoring the highest and commercial websites the lowest. This is also evident if we look at the percentage of websites that have a JAMA score of ≥ 3 , where journalism websites score the highest, followed by health portals, and commercial websites the lowest. It is difficult to comment on classes of websites with 10 or less websites.

When we look at the JAMA score of websites classified by types of intervention, approved drugs score slightly higher but the differences are small and not statistically significant. Again, it is difficult to comment on the high score of websites that mention research drugs, as there are only 6.

Because MedlinePlus returned only two types of websites, and few of them indicated an intervention, this analysis has not been performed on those websites.

To identify the reasons for the different JAMA scores in the different classes of website described in Table 7, we have disaggregated the JAMA score in its 4

components and have analyzed them in the different classes of websites (Table 8). From this is clear that most websites met the "disclosure" criteria, and very few the "attribution". The two criteria for which commercial websites and no profit ones differed the most from the ones having a higher JAMA score (health portals and journalism) were "authorship" and "currency".

We then analyzed the distribution of intervention groups by class of websites. As shown in Figure 3, it is clear that the pattern of classes of websites is similar across all the types on intervention mentioned, i.e. there is not a type of intervention that is preferentially described by a specific class of websites. Possibly there is a preference toward alternative medicine and procedure/devices in commercial websites, and for nutritional interventions in health portals. Of note, the pattern cannot be compared with that given in Table 3 because website can mention multiple types of interventions and an average value, obtained by adding up the values for all the interventions is shown for comparison.

Discussion

The present study extends our previous approach of assessing the quality of the health information of websites returned by search engines using a medical search term. In the previous work we analyzed different classes of websites using instruments (HON code, JAMA score) that measure the trustworthiness of a website rather than the quality of the information provided. We have analyzed the websites returned by a Google search in terms of the type of intervention indicated in their content. Of course, we cannot make conclusion from our findings that could be

extended to any search query. Also, results returned by Google will change with time, location and search history. Nevertheless, the 200 websites returned represent a good sample of migraine-related websites so that some conclusion can be drawn on the usefulness of the methodology described here.

The approach used here is unbiased, and does not provide an absolute indicator of “health IQ”. However, seen from the perspective of EBM one should consider a website of higher quality if it points to a drug approved by a regulatory agency, and that has gone through EBM criteria for its approval, rather than to, for instance, crystal healing.

In this respect, it is important to note that, as shown in Table 7, a trustworthiness score such as the JAMA score is not predictive of whether a website will promote the use of alternative medicine rather than of an approved drug, confirming that the type of intervention and the criteria used in the JAMA score are assessing different dimensions of IQ. This confirms conclusions made from a larger study on over 300 web pages providing information on breast cancer, where the JAMA score was found not to predict whether the information provided was scientifically accurate (Bernstam et al., 2008).

Commercial websites, no profit organizations and patient groups have a lower JAMA score lower than health portals, professional or journalism websites. This is exactly the same pattern we previously reported in a study on health information on diabetic neuropathy, independently on the search engine used, whether it was Google, Yahoo, Bing or Ask (Chumber et al., 2014).

One recommendation that can be made from this observation is that owners/publishers of commercial, patient group and no-profit websites try to improve their trustworthiness, which could often be achieved by providing simple information such as the author of the text, and date of last update.

Another pattern that is evidenced in the present study (particularly in the cluster analysis and in Table 5) is that commercial websites will often describe only one type of intervention. This comes as no surprise and it was probably expected that a website www.tylenol.com describes only acetaminophen; it is probably a very good source of information on how to use this drug. However, for a patient searching the Internet on how to cure their disease it is important to have websites that mention different options, possibly with a critical analysis, or at least a description, of their respective benefits, risks, and the scientific evidence for their efficacy. This might identify criteria to be considered in evaluating health websites, inclusivity as opposed to exclusiveness. Clearly, including various intervention options may be desirable or not depending on the purpose of a website, but it may be helpful to direct patients toward health portals if they want to have a general idea of the different options and to commercial websites if they seek information on one specific type of intervention.

An important question in studies of this kind is whether it is relevant to analyze 200 websites returned by Google. In fact, many studies have shown that the user directs most of the attention to the first items in the search list (Cutrell and Guan, 2007) and a study in patients searching health-related information found that they will usually look at the first 10 websites returned by the search engine (Eysenbach and Kohler,

2002). The difference in the pattern observed for all 200 websites and the top 10 is most interesting. Contrary to common belief that for-profit search engines, like Google, will preferentially return commercial websites, we found, as shown in Table 6, that these were in fact underrepresented, to a statistically significant extent, in the top 10 results. Of course, this may depend on the search terms, or the disease condition we search for. We have re-analysed the raw data from our previous study where we searched the term "diabetic pain" using Google and analysed 200 websites returned (Chumber et al., 2014), and found that also in that dataset, commercial websites are underrepresented in the first 10 hits. In fact, in that study, the total number of commercial websites was 42 but they were not present in the first 10 hits. Also in that case, commercial websites were significantly underrepresented. To assess how much this depends on the terms used in the query, we performed, during the revision of this manuscript, 4 different searches on "migraine cure", "migraine medicine", "migraine treatment" and "migraine therapy" and looked at the top 10 websites returned by Google. As shown in Table 2 in Supplementary Material, there were no websites classifiable as "Commercial" in any of these SERPs. Clearly we cannot generalize and the distribution of websites returned will vary with the terms used in the queries. It would be interesting to characterize the pattern of websites returned by different search queries, possibly analysing also existing, published data. We re-analysed the data reported in a IQ study on kidney transplant for which a list of 94 websites was reported. In that study, commercial websites were 25% of the total, but none was in the top 10 returned by Google (Hanif et al., 2007).

Also, it is often suspected that the Internet would generally point the patient toward alternative medicine approaches rather than medicinal products whose approval is based on EBM criteria. However, the data reported here show that, although among the total 200 websites analysed alternative medicine interventions are mentioned more frequently than approved drugs (55% vs. 35%), this is not true in the first 10 items returned by Google, where approved drugs are significantly over-represented.

Of course there are other features of the websites that should be taken into account. The fact that a website is returned in a Google search does not mean that it will be read. That depends on several issues including attractiveness, readability and various aspects of the website's design and content, that would require, in addition to the readability test mentioned above, using eye-tracking analysis (Granka et al., 2004).

More importantly, we need to stress once again that these results were obtained with a specific query. Anyone who had a chance of searching the Internet for health information knows how often low-quality websites pointing the layperson towards treatments, or strategies for prevention, that are not evidence-based, are found in the top results. Ideally, performing a large number of searches on different health-related queries would point at where misinformation, or disinformation, make it to the top of the search. Clearly this is a larger project that would require machine processing rather than individual scoring. The use of the type of intervention described here is something that could be implemented by defining list of keywords that, using automated machine learning. This study shows that analysing websites for the type of intervention they describe could provide other dimensions of health IQ in addition to those described by the JAMA score. In fact, at least within the

sample of 200 websites retrieved here, the JAMA score did not differentiate between websites describing approved drugs from those describing alternative medicine approaches.

Performing a cluster analysis of websites, based on their classification and the intervention described, can identify patterns of websites pointing patients towards one or more treatments. In the specific example of health-related query studied here, we were surprised that commercial websites ranked low in the search list. In terms of the type of intervention recommended, complementary/alternative medicine occurred more often than approved, EBM-based drugs, but the latter were more frequent in the top ten results, possibly due to the intrinsic higher IQ features of these websites.

It would be important to apply this type of analysis over a wide range of search terms and disease conditions and health topics in order to identify areas that are more at risk of directing patients toward non-EBM types of intervention, and thus potentially pose problems to public healthcare systems and health insurance companies. This would indicate where policy makers and professional organizations should concentrate their effort to inform patients on the benefits, risk and scientific basis of existing therapies.

Limitations

The results of the study are specific for one specific query on one health topic. In particular, the observation that commercial websites are under-represented in the first 10 results, that health portals are the most frequent class of websites and

alternative medicine the most frequently described type of intervention are specific to this topic; searches in other health domains may give completely different results. Additional studies on different health topics will be needed to assess the usefulness of the proposed methodology (classification by type of intervention and cluster analysis). Finally, the type of classification we have used (both in terms of website classes and types of interventions) may need to be tailored to the specific research questions.

Abbreviations

EBM, evidence-based medicine; IQ, information quality; JAMA, Journal of the American Medical Association

Competing Interests

The authors have no competing interests to declare.

Authors' contributions

MY and PG both analyzed the data and wrote the paper.

REFERENCES

- Alexa the web information company. *The top 500 sites on the web*. [Online]. Available: <http://www.alexa.com/topsites> [Accessed Nov 3rd 2012].
- Health on the net foundation. *The HON code of conduct for medical and health web sites (HONcode)* [Online]. Available: <http://www.hon.ch/HONcode/Conduct.html> [Accessed Feb 10th 2013].
- Barker, S., Charlton, N.P., and Holstege, C.P. (2010). Accuracy of internet recommendations for prehospital care of venomous snake bites. *Wilderness Environ Med* 21, 298-302. doi: 10.1016/j.wem.2010.08.016.
- Berland, G.K., Elliott, M.N., Morales, L.S., Algazy, J.I., Kravitz, R.L., Broder, M.S., Kanouse, D.E., Munoz, J.A., Puyol, J.A., Lara, M., Watkins, K.E., Yang, H., and Mcglynn, E.A. (2001). Health information on the Internet: accessibility, quality, and readability in English and Spanish. *JAMA* 285, 2612-2621.
- Bernstam, E.V., Walji, M.F., Sagaram, S., Sagaram, D., Johnson, C.W., and Meric-Bernstam, F. (2008). Commonly cited website quality criteria are not effective at identifying inaccurate online information about breast cancer. *Cancer* 112, 1206-1213.
- Brosnan, S., Barron, E., and Sahm, L.J. (2012). Health literacy and the clozapine patient. *Perspect Public Health* 132, 39-42.
- Charnock, D., Shepperd, S., Needham, G., and Gann, R. (1999). DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. *J Epidemiol Community Health* 53, 105-111.
- Chumber, S., Huber, J., and Ghezzi, P. (2014). A methodology to analyze the quality of health information on the internet: the example of diabetic neuropathy. *The Diabetes Educator*.
- Cutrell, E., and Guan, Z. (Year). "What are you looking for?: an eye-tracking study of information usage in web search", in: *Proceedings of the SIGCHI conference on Human factors in computing systems*: ACM), 407-416.
- Dy, C.J., Taylor, S.A., Patel, R.M., Kitay, A., Roberts, T.R., and Daluiski, A. (2012). The effect of search term on the quality and accuracy of online information regarding distal radius fractures. *J Hand Surg Am* 37, 1881-1887. doi: 10.1016/j.jhsa.2012.05.021.
- Eysenbach, G., and Kohler, C. (2002). 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* 324, 573-577.
- Fabricant, P.D., Dy, C.J., Patel, R.M., Blanco, J.S., and Doyle, S.M. (2013). Internet search term affects the quality and accuracy of online information about developmental hip dysplasia. *J Pediatr Orthop* 33, 361-365. doi: 10.1097/BPO.0b013e31827d0dd2.
- Fox, S., and Duggan, M. (2013). Health online 2013. *Health*.
- Freelon, D.G. (2010). ReCal: Intercoder reliability calculation as a web service. *International Journal of Internet Science* 5, 20-33.

- Ghezzi, P., Chumber, S., and Brabazon, T. (2014). "Educating Medical Students to Evaluate the Quality of Health Information on the Web," in *The Philosophy of Information Quality*. Springer), 183-199.
- Granka, L.A., Joachims, T., and Gay, G. (2004). "Eye-tracking analysis of user behavior in WWW search", in: *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*. (Sheffield, United Kingdom: ACM).
- Hanif, F., Abayasekara, K., Willcocks, L., Jolly, E.C., Jamieson, N.V., Praseedom, R.K., Goodacre, J.A., Read, J.C., Chaudhry, A., and Gibbs, P. (2007). The quality of information about kidney transplantation on the World Wide Web. *Clinical transplantation* 21, 371-376.
- Kaicker, J., Debono, V.B., Dang, W., Buckley, N., and Thabane, L. (2010). Assessment of the quality and variability of health information on chronic pain websites using the DISCERN instrument. *BMC Med* 8, 59. doi: 10.1186/1741-7015-8-59.
- Kaphingst, K.A., Zanfini, C.J., and Emmons, K.M. (2006). Accessibility of web sites containing colorectal cancer information to adults with limited literacy (United States). *Cancer Causes Control* 17, 147-151. doi: 10.1007/s10552-005-5116-3.
- Lee, Y.W., Strong, D.M., Kahn, B.K., and Wang, R.Y. (2002). AIMQ: a methodology for information quality assessment. *Information & management* 40, 133-146.
- Liang, B.A., and Mackey, T. (2009). Searching for safety: addressing search engine, website, and provider accountability for illicit online drug sales. *American Journal of Law and Medicine*, Forthcoming.
- Martins, E.N., and Morse, L.S. (2005). Evaluation of internet websites about retinopathy of prematurity patient education. *Br J Ophthalmol* 89, 565-568. doi: 10.1136/bjo.2004.055111.
- Meric, F., Bernstam, E.V., Mirza, N.Q., Hunt, K.K., Ames, F.C., Ross, M.I., Kuerer, H.M., Pollock, R.E., Musen, M.A., and Singletary, S.E. (2002). Breast cancer on the world wide web: cross sectional survey of quality of information and popularity of websites. *BMJ* 324, 577-581.
- Peterlin, B.L., Gambini-Suarez, E., Lidicker, J., and Levin, M. (2008). An analysis of cluster headache information provided on internet websites. *Headache* 48, 378-384. doi: 10.1111/j.1526-4610.2007.00951.x.
- Silberg, W.M., Lundberg, G.D., and Musacchio, R.A. (1997). Assessing, controlling, and assuring the quality of medical information on the Internet: Caveant lector et viewer--Let the reader and viewer beware. *JAMA* 277, 1244-1245.
- Thompson, A.E., and Graydon, S.L. (2009). Patient-oriented methotrexate information sites on the Internet: a review of completeness, accuracy, format, reliability, credibility, and readability. *J Rheumatol* 36, 41-49. doi: 10.3899/jrheum.080430.
- Wang, R.Y., and Strong, D.M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of management information systems*, 5-33.
- Zeldman, J. (2001). *Taking Your Talent to the Web: A Guide for the Transitioning Designer*. New Riders Publishing }.

Table 1. Classes of websites

Affiliation	Description	Examples
<i>Professional (P)</i>	Website created by a person or organization with professional knowledge of the information e.g. government, institutions, libraries, universities, publishers, online scientific journals, and other educational institutions	nhs.uk umn.edu ninds.nih.gov
<i>Commercial (C)</i>	Websites that buy, sell, or provides a service for a fee e.g. profit organizations	lipigesic.com/ <u>migracap.co.uk/</u> walgreens.com/
<i>Health Portal/ Blog (HP)</i>	Web site or search engine with health information on a variety of health topics e.g. health blogs	joybauer.com/ mayoclinic.com/ healthline.com/
<i>Patient Group (PG)</i>	Websites targeted at patients or created by patients e.g. patient blogs, patient forums, chat rooms and support groups	curetogether.com curezone.org/forums myhomeremedies.com
<i>Journalism</i>	Websites primarily broadcasting news online, providing information relating to health	foxnews.com bbc.co.uk philly.com/

	topics.	
Non-profit (NP)	Websites providing information for educational or charitable reasons with no financial beneficiaries. e.g. charitable organizations	wikipedia.org migrainetrust.org migraine.ie
Other(O)	Websites which do not fit into any of the other affiliations. Includes social networking sites.	facebook.com twitter.com

Table 2. Intervention groups

Intervention	Description
Approved Drug	Pharmacological therapy validated for a therapeutic use by the FDA or the British National Formulary.
Alternative Medicine	Therapies that are not based on scientific evidence. These include homeopathy, herbalism, naturopathy, and crystal healing.
Food	Recommendation of food for management of migraine. These include coffee, lavender tea, ginger and honey
Procedure & Devices	Recommendation of any procedure or use of a device for management of migraine. These include surgery,

	biofeedback and migraine cap.
Lifestyle & Triggers	Altering lifestyle factors and/or avoidance of triggers of migraine (e.g. recommending regular sleeping, or avoiding alcohol).
Research drug	Pharmacological therapy which is still in research stages and not yet approved for the use of migraine (e.g. lidocaine, calcitonin gene-related peptide receptor antagonists)
No information given	Gave no information on how to manage migraine.

Table 3. Distribution of websites generated from search in Google or MedlinePlus according to their affiliation

Affiliation	Google (%)	MedlinePlus (%)
C (Commercial)	31	
HP (Health portal)	44	1
J (Journalism)	11	
NP (Non-profit)	4	
P (Professional)	3	89
PG (Patient Group)	6	10
O (Other)	1	

Data represent the percentage of websites in each affiliation. Total number of websites was 198 for Google and 74 for MedlinePlus.

Table 4. Distribution of websites generated from search in Google or MedlinePlus according to the intervention indicated

Intervention	Google (%)	MedlinePlus (%)
Approved Drug	35	23
Alternative Medicine	55	8
Food	41	3
Procedures & Devices	49	7
Lifestyle & Triggers	42	11
Research Drugs	3	3
No information given	5	70

Data represent the percentage of websites mentioning a type of intervention. Total number of websites was 198 for Google and 74 for MedlinePlus.

Table 5. Average number of intervention types described by the different classes of websites.

Website class	No. of treatments indicated	n
Commercial	1 ± 1 ^{ab}	60
Health Portals	3 ± 1.4 ^a	87
Journalism	2 ± 1.4	22
No Profit	3 ± 1.3	8
Professional	3 ± 2.1	6
Patient Groups	4 ± 1.9 ^b	11

Data are mean ± SD. N is the number of websites in each class. Values bearing the same symbol are significantly different from each other: ^a P<0.0001; ^b P<0.005.

Kruskal-Wallis test was used for multiple comparisons of nonparametric variables, followed by Dunn's test.

Table 6. Distribution of classes of websites and types of intervention in the ranking of Google search

Website class	Number in top 10	Number in total	Type of Intervention	Number in top 10	Number in total
Commercial	0 ^a	60 ^a	Approved Drug	8 ^b	70 ^b
Health Portals	5	89	Alternative Medicine	5	109
Journalism	2	22	Food	8	81
No Profit	1	8	Procedures & Devices	5	98
Professional	1	6	Lifestyle & Triggers	8	84
Patient Groups	1	12	Research Drugs	2	6

Occurrence of classes of websites and type of intervention they describe in the first 10 hits and in the total of 198 websites.

Values bearing the same letter are significantly different from each other by Chi-

square test: ^a P<0.05; ^b P<0.05

Table 7. Mean JAMA score of websites by class and type of intervention

Website class	Mean JAMA score	% >3	Type of Intervention	Mean JAMA score	% >3
Commercial	1.4±0.7ab (60)	10%	Approved Drug	2.09±0.8 (70)	31%
Health Portals	2.1±0.9ae (87)	38%	Alternative Medicine	1.84±0.84 (109)	24%
Journalism	2.6±0.5bcd (22)	64%	Food	1.96±0.83 (21)	26%
No Profit	1.1±0.6ce (8)	0	Procedures & Devices	1.82±0.82 (22)	22%
Professional	2.3±0.5 (6)	33%	Lifestyle & Triggers	1.99±0.87 (26)	31%
Patient Groups	1.5±0.5d (11)	0	Research Drugs	2.5±0.55 (6)	50%

Data are the mean JAMA score ± SD; the number of websites in each group is indicated in parentheses. Values bearing the same letter are significantly different from each other: ^a P<0.0001; ^b P<0.0001; ^c P<0.0005; ^d P<0.01; ^e P<0.05. Kruskal-Wallis test was used for multiple comparisons of nonparametric variables, followed by Dunn's test.

Table 8. JAMA score components in the different classes of websites

Website class	Authorship	Attribution	Disclosure	Currency
Commercial	17	5	100	22
Health Portals	48	7	100	56
Journalism	64	5	100	95
No Profit	0	13	88	13
Professional	33	17	100	83
Patient Groups	9	0	100	45

Data indicate the percentage of websites in each class that met the specific criteria.

Figure Legends

Figure 1. Flow chart of data collection and analysis

Figure 2. Cluster analysis of the websites returned by Google. A. Type of intervention mentioned by the 198 websites (in alphabetical order by website URL). B. Hierarchical cluster analysis on websites from panel A, clustered by type of intervention. The table on the right shows the composition by class of websites of the 8 cluster identified in this figure.

Figure 3. Composition of websites mentioning different types of intervention. Data are expressed as a percentage of the total number of websites mentioning a type of intervention. Data labels are shown above each bar. Research drugs as type of intervention are not included because of the small number of websites (n=4).

Figure 1.TIFF

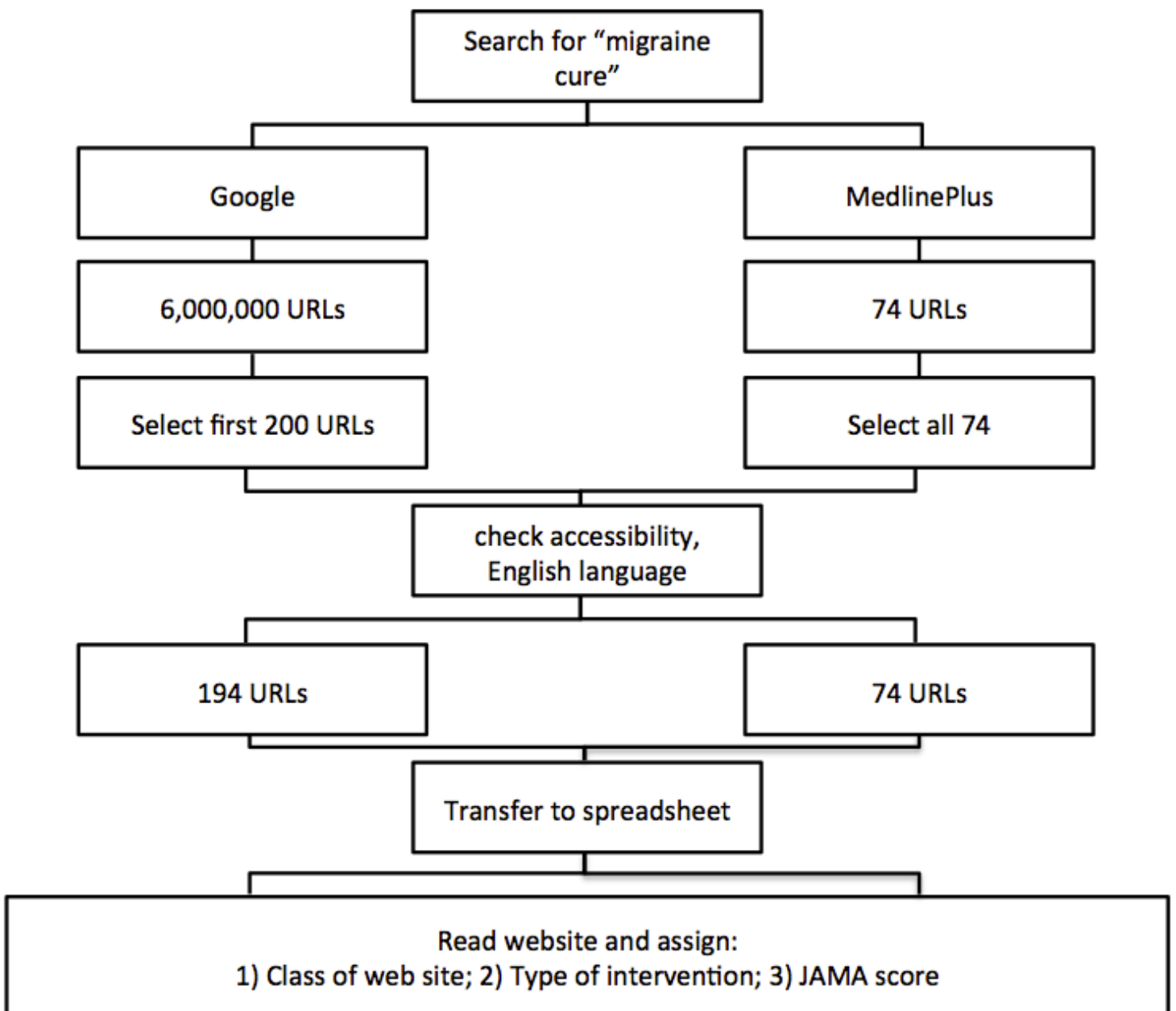
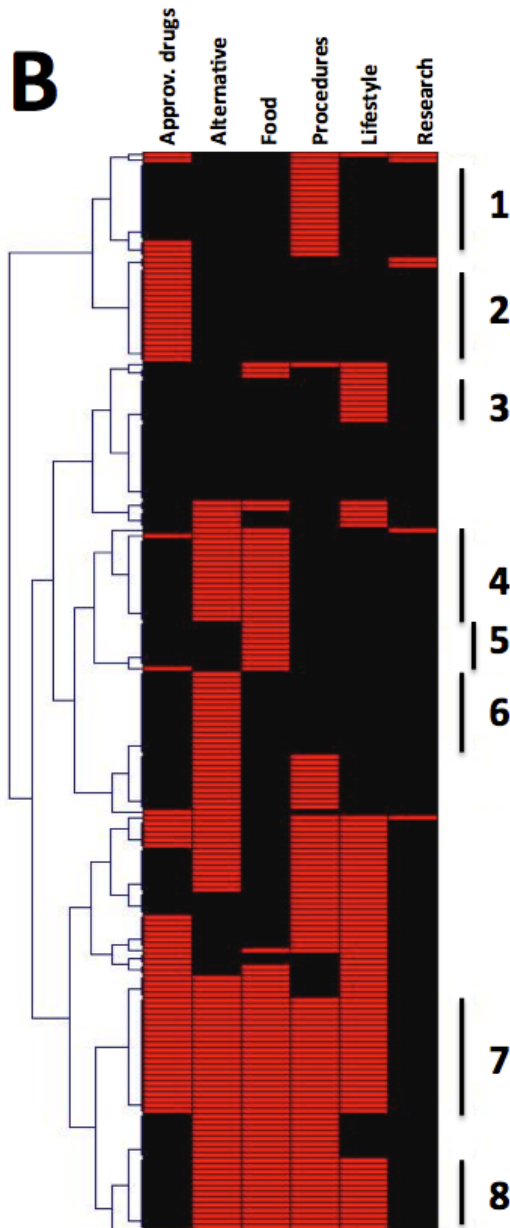
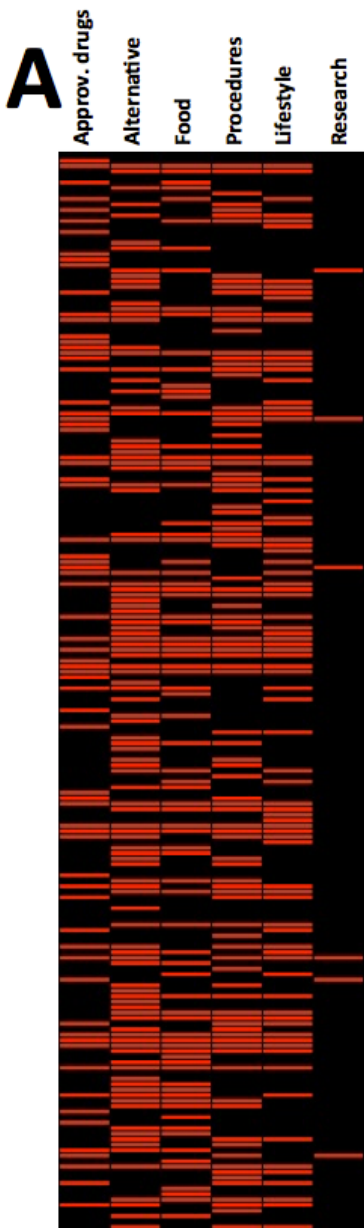


Figure 2.TIFF



	C	HP	J	NP	P	PG
% Total	28	48	11	4	3	6
	64	29	7			
	53	35	6		6	
	13	25	63			
% of websites in each cluster						
	80	13	7			
	75	13				13
	73	13	13			
		57	10		5	29
	15	62		23		

Figure 3.TIFF

