The impact of recommendations and warnings on the quality evaluation of health websites: An online experiment

Diviani, N.; Meppelink, C.S.

DOI
10.1016/j.chb.2017.01.057
Publication date
2017
Document Version
Final published version
Published in
Computers in Human Behavior

Computers in Human Behavior 71 (2017) 122–129

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

The impact of recommendations and warnings on the quality evaluation of health websites: An online experiment

Nicola Diviani [a, b, *], Corine S. Meppelink [c]

[a] University of Lucerne, Switzerland
[b] Swiss Paraplegic Research, Switzerland
[c] Amsterdam School of Communication Research, University of Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 22 July 2016
Received in revised form
27 January 2017
Accepted 27 January 2017
Available online 30 January 2017

Keywords:
Online information
Information evaluation
Evaluation criteria
Warnings
Health literacy

ABSTRACT

With the increase in availability of online health information (OHI), consumers need to be able to properly evaluate the quality of health websites. Although several established evaluation criteria for OHI are available, these are rarely used by consumers. To improve people's ability and motivation to critically evaluate OHI, insights into how these criteria can optimally be communicated are needed. This study aims to investigate whether educational messages recommending the use of quality criteria can improve consumers' ability to evaluate OHI credibility, especially among people with low health literacy. We also test whether these messages can yield a stronger effect when combined with information warning consumers against using non-established criteria. In an online experiment, we randomly assigned 403 participants to one of four conditions and asked them to evaluate two websites of different quality. The conditions consisted of recommendations promoting the use of established evaluation criteria, warnings against using commonly adopted non-established criteria, a combination of the two, and a control group. Participants exposed to messages recommending established criteria evaluated the credibility of the lower quality website as poorer, and were better able to discriminate between high and low quality information compared to those in the control group. A combination of recommendations and warnings also improved people's ability to evaluate, but did not yield a stronger effect than the recommendations alone. Subgroup analysis, however, showed that these effects existed only for those with high health literacy. We conclude that providing consumers with criteria to evaluate OHI might be a viable way to improve people's evaluation skills of online health materials. Further research is needed to identify efficient ways to communicate these criteria to low health literate audiences.

1. Introduction

During the last decades, the amount of online health information (OHI) has dramatically increased and more and more people turn to the Internet to find information about health-related issues (Fox & Duggan, 2013). Despite the widespread availability of health information, large differences in the quality of OHI on different topics exist (Eysenbach, Powell, Kuss, & Sa, 2002; Zhang, Sun, & Xie, 2015). Consequently, the quality of the information that consumers encounter during their web searches has been a reason of concern (Berland et al., 2001; Bernstam et al., 2008; Eysenbach et al., 2002; Hardey, 2001; Stvilia, Mon, & Yi, 2009). OHI- seeking thus poses several major challenges to the users of health information. This particularly applies to those with limited health literacy, who are less likely than people with adequate health literacy to possess the skills required to evaluate of the quality and credibility of the vast amount of often unverified health information that can be found on the Internet (Flanagin & Metzger, 2008; Jiang & Beaudoin, 2016).

To distinguish between reliable and unreliable health information, knowledge about established quality criteria for OHI could be extremely helpful to consumers. Such criteria exist in form of checklists and have shown to be useful tools for web designers, providers of health information, and researchers (for a summary view, see Kim, Eng, Deering, & Maxfield, 1999; Zhang et al., 2015). However, research has also shown that these criteria are hardly used by consumers in their assessment of OHI quality (Eysenbach & Kohler, 2002; Flanagin & Metzger, 2000; Fogg et al., 2003; Rieh,

* Corresponding author. University of Lucerne, Department of Health Sciences & Health Policy, Frohburgstrasse 3, 6002 Lucerne, Switzerland.
E-mail address: nicola.diviani@unilu.ch (N. Diviani).

http://dx.doi.org/10.1016/j.chb.2017.01.057
0747-5632/

2002; Scholz-Crane, 1998). Instead, people often rely on unreliable subjective feelings or heuristics such as the ranking of a website in the search results or its layout (Diviani, van den Putte, Giani, & van Weert, 2015; Diviani, van den Putte, Meppelink, & van Weert, 2016; Feufel & Stahl, 2012; Gauld & Williams, 2009; Mackert, Kahlor, Tyler, & Gustafson, 2009). The explanation for this could be twofold. First, it has been suggested that existing quality guidelines are usually very detailed and complex and therefore difficult for consumers to apply in their regular searches for OHI (Meola, 2004; Metzger, 2007). Educating people about these criteria using an easy-to-use format could therefore be useful to improve consumers' ability to evaluate OHI. Second, many OHI consumers lack awareness of the issue of health information quality, which means that they do not recognize the need to systematically apply evaluation criteria (see, e.g., Diviani et al., 2016). In these situations, a different message strategy is possibly needed, in which people are not only informed about established evaluation criteria, but also warned against using non-established ones.

This study responds to a call for attention to the issue of the quality of online health information and for successful strategies to improve people's ability to evaluate OHI (Berland et al., 2001; Bernstam et al., 2008; Car, Lang, Colledge, Ung, & Majeed, 2011; Eysenbach et al., 2002; Hardey, 2001; Stvilia et al., 2009). If people are better able and more motivated to evaluate OHI credibility and quality, adverse health outcomes such as low participation in screening programs or low adherence to treatments caused by wrong or incomplete information, could be prevented (Cline & Haynes, 2001). As it has been suggested that the evaluation of OHI is mainly problematic among people with low levels of health literacy (Diviani et al., 2015), it could be assumed that the findings of this study are particularly important for this group. Our study addresses the following research question: Do educational messages recommending the use of established evaluation criteria improve consumers' ability to evaluate OHI credibility, especially among people with low health literacy, and can these messages yield stronger effects when combined with information warning consumers against using non-established criteria?

1.1. Established criteria to evaluate credibility

With the rise of the Internet, the amount of information that is available to the public has become virtually unlimited (Viswanath, 2005). However, as the production of online content is not restricted to authorities, this has reasonably raised concerns regarding the credibility and quality of online information (Metzger, 2007). In contrast to traditional public information, online information is usually not filtered by professional gatekeepers, resulting in large amounts of inaccurate information online (Zhang et al., 2015). On the Internet, inaccurate information is not only equally accessible as accurate information, but is also presented in the same format (e.g., a website). Consequently, this could induce a 'leveling effect', which suggests that consumers perceive both accurate and inaccurate information as equally credible (Burbules, 1998). As consumers do not (and are not expected to) possess the expertise to assess the actual accuracy of online information, to accurately differentiate between accurate and inaccurate information, consumers need to pay attention to specific indicators of information credibility and quality that are often beyond the first impression. According to Metzger (2007), careful credibility evaluation of online information incorporates assessment of the following five domains: accuracy, authority, objectivity, currency, and coverage or scope. These indicators of information quality are based on extensive research, and are the ones most commonly included in recommended criteria to evaluate OHI (Zhang et al., 2015). We will therefore refer to these as 'established criteria'.

1.2. Informing OHI consumers about established criteria

Over the years, multiple instruments to assess OHI quality have been developed. In a recent review, Zhang et al. (2015) were able to identify 29 different established instruments (i.e., used in more than one study). In their paper, the authors showed that some of them (e.g., DISCERN, HONcode, and JAMA benchmarks) have been used more often than others in research settings. Although the separate instruments include different specific indicators (sometimes context-dependent), and some also assess additional aspects such as accessibility, aesthetics, or navigability, all of them cover at least to a certain extent the basic credibility criteria outlined by Metzger (2007). To be considered accurate, for instance, information presented on a website should be error-free; to be considered authoritative, a website should present the name of its authors, their credentials, and their contact details; to be considered objective, a website has to be open about its purpose, has to clearly distinguish between editorial and commercial contents, and has to disclose commercial intent and/or conflicts of interest; to be considered current, a website should include a publication date and the time when the information was last updated; last, coverage or scope is assessed by establishing whether information on a website adequately covers all the relevant aspects of a given topic (e.g., presents all available treatment options with pro and contra).

Because of these checklists, we are now able to create websites fulfilling established quality criteria (i.e., high quality health websites), which are in turn perceived as more accurate (see e.g, Allam, Sak, Diviani, & Schulz, 2017), as well as to conduct large scale quality assessments of OHI in different health domains (see, e.g., Grewal & Alagaratnam, 2013 for websites about colorectal cancer or Kaicker, Debono, Dang, Buckley, & Thabane, 2010 for websites about chronic pain). Research on consumer OHI seeking, however, has shown that these criteria are rarely used by consumers to assess OHI quality (Eysenbach & Kohler, 2002; Flanagin & Metzger, 2000; Fogg et al., 2003; Rieh, 2002; Scholz-Crane, 1998). There are at least two plausible reasons for this, both related to the nature of the checklists themselves. First, checklists have originally been developed for online health information providers (e.g., web designers or content managers) or for researchers, and have therefore not been designed to be particularly consumer-friendly. They are usually very detailed, verifying all the relevant aspects of such checklists requires time, and applying them requires advanced web browsing skills. It is therefore not surprising that consumers of OHI do not apply them in their everyday searches for OHI (Meola, 2004; Metzger, 2007). Second, no efforts have been devoted to their dissemination. As a result, as suggested among others by Diviani et al. (2016), many OHI consumers do not question health information quality, which means that they do not even recognize the need to apply evaluation criteria. Our first hypothesis therefore states the following: *Providing easy-to-use information recommending the use of established evaluation criteria improves people's ability to evaluate the credibility of OHI* (*H1a*) *and to distinguish between high and low quality health information* (*H1b*).

1.3. Warnings against using non-established criteria

Metzger (2007) notes that merely informing people about how to evaluate the credibility of online information is probably not sufficient. Motivation to carefully evaluate the credibility of online sources is often lacking, and people rarely check the accuracy of information by using established criteria (Flanagin & Metzger, 2000; Scholz-Crane, 1998). In line with other processing models, Metzger (2007) therefore introduces the 'dual processing model of Web site credibility assessment' which proposes that thorough credibility judgment of online information only takes place when

people are both able and motivated to evaluate. When one or both are lacking, no evaluation takes place or evaluation will be based on heuristics or peripheral cues. Research has shown that consumers of OHI lack awareness about the credibility of OHI. In-depth interviews conducted among 44 participants, for instance, revealed that many people do not question the reliability of the information they encounter (Diviani et al., 2016). To motivate consumers to carefully evaluate online information, Metzger (2007) suggested that providing information about the negative consequences of misinformation online could be an effective strategy in addition to educational messages.

One thing consumers should specifically be warned against in the context of OHI searches is the use of non-established criteria. Research has shown that people often use non-established criteria or heuristics to evaluate the quality of OHI. For instance, consumers of OHI information perceive websites as more credible when they appear on top of Google search results, when they have a clean layout, or when they present information that is consistent with pre-existing ideas. However, none of these attributes can be considered a good indicator of quality (Diviani et al., 2016; Feufel & Stahl, 2012; Gauld & Williams, 2009; Mackert et al., 2009). Therefore, we argue that warning people against using non-established criteria in addition to the educational messages could be an effective strategy. By creating awareness of the risks of using non-established criteria and by recommending evaluation criteria that can be easily applied, this could possibly further improve consumers' evaluation skills. Warnings alone, however, cannot be expected to be useful as educational messages because these are by definition vague and unspecified and do not provide any actionable information about specific informational cues. We therefore hypothesize that: *Compared to presenting recommendations only, educational messages including warnings against using non-established criteria will yield stronger effects on people's ability to evaluate the credibility of OHI* (*H2a*) *and to distinguish between high and low quality health information* (*H2b*).

1.4. Health literacy and OHI evaluation

According to Wathen and Burkell (2002), credibility assessment of online information is the result of two interacting factors: the website itself and the consumer. The way in which a website's credibility is assessed therefore likely depends on consumer factors, such as the need for information, need for cognition, and prior knowledge about the topic. Depending on the context and the goal of the search, people can have a higher or lower need to carefully assess the information they encounter (Metzger, 2007). In the context of health information searches, however, this does not always apply. Qualitative research has shown that people report to use the same evaluation criteria for serious and less serious health issues (Diviani et al., 2016). A personal factor that has shown to be relevant in the context of OHI evaluation is health literacy. Health literacy broadly refers to the "ability to obtain, process, understand, and communicate about health-related information needed to make informed health decisions" (Berkman, Davis, & McCormack, 2010, p. 16). Ability to evaluate online information has been recognized as a crucial health literacy skill, and is therefore expected to be related to one's overall health literacy level (Diviani et al., 2015).

Low levels of health literacy are quite prevalent (Sørensen et al., 2015) and have been shown to be related to a wide range of adverse health outcomes (Berkman, Sheridan, Donahue, Halpern, & Crotty, 2011). It has been argued that one of the pathways linking low health literacy with suboptimal health outcomes could be low health literate people's inability to adequately assess the information they access outside the medical encounter (Jiang & Beaudoin, 2016). Research in the field of OHI-seeking has shown that people with low health literacy levels more often use non-established criteria to evaluate a website's credibility compared to people with high health literacy (Diviani et al., 2016). Moreover, people with lower levels of health literacy more often lack awareness of the issue of OHI credibility. It could therefore be expected that both educational messages and additional warnings are especially effective among this group. We therefore hypothesize that, *among people with low health literacy, providing easy-to-use recommendations on how to use established evaluation criteria particularly improves the ability to evaluate the credibility of OHI (H3a)* and *to distinguish between high and low quality health information (H3b)*. Furthermore, we expect that *adding warnings against using non-established criteria to educational messages is particularly effective for people with low health literacy in improving their ability to evaluate the credibility of OHI (H3c)* and *to distinguish between high and low quality health information (H3d)*.

2. Method

2.1. Participants and procedure

An online survey was conducted using Amazon's Mechanical Turk (MTurk), an online crowd sourcing system. The survey was available for completion from March 24 to 29, 2016 to English-speaking MTurk workers residing in the United States. MTurk respondents who opted to complete the survey were directed to an outside survey site. All respondents were anonymous and received $ 2.00 as a reward for their participation. Every participant could only take part in the survey once. To ensure the validity of the data, participants were required to use a personal computer (desktop or laptop) or a tablet. Information on the type of device used was automatically collected. Additionally, the online questionnaire included an attention check which was programmed to record the time participants spent reading the instructions and browsing the assigned website (Crump, McDonnell, & Gureckis, 2013). The Ethics Review Board of the Amsterdam School of Communication Research (ASCoR) provided ethical approval for this study (identification number 2016-PC-6760) and all participants had to provide informed consent before filling out the questionnaire.

The survey was accessed by 436 individuals. Responses were excluded from the analysis of those who did not complete the questionnaire ($n = 8$, 1.8%), used a smartphone ($n = 4$, 0.9%), failed the attention check ($n = 6$, 1.4%), and spent less than 15 s reading the instructions or less than 30 s browsing the assigned website ($n = 15$, 3.4%). The final dataset thus consisted of 403 complete and usable responses (92.4%). An overview of the characteristics of the study participants is presented in Table 1.

The experiment had a one-factor between-subjects design. The second independent variable, functional health literacy, was measured. The online questionnaire was programmed to randomly assign participants to one of four conditions (i.e., recommendations, warnings, combination, control). Participants in the *control* condition were instructed on how to fill in the questionnaire and redirected to the page with the link to open the website in a new tab of the browser. Participants in the *recommendations* condition were presented with, along with the instructions for completing the questionnaire and before being redirected to the page with the link to open the website, four tips recommending the use of recognized evaluation criteria that are included in established quality guidelines. Participants in the *warnings* condition were presented with four warnings against the use of evaluation criteria not recognized by established quality guidelines but commonly used by consumers. Participants in the *combination* condition were presented with a mix of criteria and warnings: half of the participants received all eight tips, while the other half received a random combination of established criteria and warnings (four tips, two of

Table 1
Mean demographics for participants by experimental group.

	Control ($n = 80$)	Recomm. ($n = 80$)	Warnings ($n = 79$)	Combination ($n = 164$)	p-value[a]
Gender (male), *n (%)*	44 (55.7)	49 (62.0)	42 (53.2)	79 (48.2)	0.228
Age, *M (SD)*	34.9 (10.52)	35.6 (11.40)	32.8 (8.21)	35.2 (11.06)	0.335
Education (college degree or more), *n (%)*	57 (67.5)	48 (60.8)	50 (63.3)	106 (67.5)	0.843
Functional health literacy (NVS), *M (SD)*	5.09 (1.333)	5.15 (1.284)	4.82 (1.738)	5.25 (1.099)	0.134
Internet use (Every day), *n (%)*	75 (93.8)	77 (96.3)	77 (97.5)	158 (96.3)	0.315
Online health information seeking (More than once a month), *n (%)*	54 (67.5)	42 (53.2)	88 (53.7)	49 (61.3)	0.149
Perceived knowledge on type 2 diabetes, *M (SD)*[b]	2.65 (1.092)	2.63 (0.960)	2.61 (0.980)	2.59 (1.015)	0.971
Personal relevance of type 2 diabetes, *M (SD)*[c]	2.31 (1.318)	2.55 (1.231)	2.23 (1.143)	2.37 (1.199)	0.394

Notes:
[a] Between-groups differences were assessed via Chi-square tests (categorical variables) and F-tests (continuous variables).
[b] Five-point Likert scale: The higher the score, the higher the perceived knowledge.
[c] Five-point Likert scale: The higher the score, the higher the personal relevance.

each type). This was done to control for possible effects of the number of tips. All messages were based on actual quotes of OHI consumers collected by the authors in a previous qualitative study (see Appendix 1 for an overview of the conditions). Subsequently, participants were exposed to two websites on type 2 diabetes of different quality. Type 2 diabetes was chosen because it was the most frequently searched condition in the United States in the 12 months preceding the present study according to Google Trends™. Additionally, because of its high prevalence (WHO, 2016), diabetes is potentially relevant for a large segment of the population. The two websites were selected among those appearing on the first two pages of results of a Google search for "type 2 diabetes" on a cookie-free computer, as previous research has shown that most people limit their search to the first pages of Google results (Eysenbach & Kohler, 2002) thus increasing the ecological validity of the experiment. In order to limit possible effects of previous experience with the websites, we excluded well-known health websites such as WebMD. The websites Dlife.com (a "platform to inform, inspire, and connect with millions of diabetes patients, consumers, and caregivers and in the process positively impact engagement and ultimately health outcome", http://www.webcitation.org/6inPgcloI) and Diabeticlivingonline.com (the website of a magazine on living with diabetes, http://www.webcitation.org/6inPLCOSa) were chosen because, despite both being commercial websites and having a similar layout, the first presented a Health on the Net (HON) quality certificate and the second did not. A website without a quality certificate is not necessarily of low quality. However, the higher quality website chosen for the experiment presented more quality markers compared to the low quality one, for instance more information about authorship of its different sections, or a more detailed description of the nature and goals of the website. In the remainder of this article we will refer to the two websites as the 'higher quality' and 'lower quality' one. However, it must be noted that they are just examples of websites on type 2 diabetes and the results regarding both sites cannot be generalized in this respect. We decided to apply the test of our hypotheses to two different websites of different quality to increase the validity of our results and to be able to compute a measure of quality discrimination ability. To control for potential order effects, we programmed the online questionnaire to randomize the sequence in which the two websites were presented to the participants. After having visited each website, participants were asked to answer several questions about its quality and credibility. To make sure that all participants were exposed to the same versions of the websites (i.e., the websites contained the same information during the time of data collection), we compared the information on the respective homepages on the first and on the last day of the experiment. No noticeable change was observed. A second part of the survey included measures of functional health literacy, eHealth literacy, and other potentially relevant variables, as well as socio-demographic questions.

2.2. Measures

2.2.1. Health literacy

Functional health literacy was assessed using the Newest Vital Sign (NVS, Weiss et al., 2005), which requires participants to read an ice cream nutrition label and to answer six questions (e.g.; *If you eat the entire container, how many calories will you eat?*). The final NVS score corresponds to the number of correct responses. The scale showed acceptable internal consistency ($\alpha = 0.70$, $M = 5.11$, $SD = 1.33$). For the analysis, participants who answered correctly to all questions of the NVS were considered as having higher health literacy, whereas all the other respondents were considered having lower health literacy.

2.2.2. Website credibility

In line with previous research (see, e.g., Flanagin & Metzger, 2000), perceived credibility of the two websites was measured by five-item scales asking participants to rate the extent to which they perceived the website to be believable, trustworthy, accurate, complete, and biased (reverse coded) on a seven-point Likert scale. The scale showed acceptable internal consistency for both websites (Low quality website: $\alpha = 0.86$, $M = 5.08$, $SD = 1.16$; High quality website: $\alpha = 0.82$, $M = 5.39$, $SD = 0.99$).

2.2.3. Website quality discrimination

Respondents were asked to rate the perceived quality of the two websites on a seven-point Likert scale (Lower quality website: $M = 5.43$, $SD = 1.39$; Higher quality website: $M = 5.52$, $SD = 1.30$). Discrimination ability was computed by subtracting the quality evaluation of the lower quality website from the quality evaluation of the higher quality website. The resulting scores could range from -6 to $+6$, with higher scores indicating higher discrimination ability to discriminate ($M = 0.32$, $SD = 1.19$).

2.2.4. Control variables

In addition to the traditional demographic variables (gender, age, and education) we measured several variables to be able to control for their potential influence. Data about previous knowledge of the websites (Low quality website: yes = 5.5%; High quality website: yes = 3.0%), relevance of the topic (five-point scale, $M = 2.37$, $SD = 1.22$), perceived knowledge on the topic (five-point scale, $M = 2.61$, $SD = 1.01$), perceived usefulness (five-point scale, $M = 3.87$, $SD = 0.91$) and intention to use the tips in future searches for OHI (five-point scale, $M = 3.91$, $SD = 0.79$), general Internet use (96% every day), and online health information seeking (76.7% at least once a month) were all collected using single-item indicators. eHealth literacy was assessed using the eHealth Literacy Scale (Norman & Skinner, 2006). The scale consists of 8 items asking

participants to rate their ability to complete a series of tasks related to online health information seeking on a 5-point Likert scale. The scale presented good internal consistency and a sum score was computed ($\alpha = 0.89$, $M = 33.34$, $SD = 4.89$).

2.2.5. Statistical analysis

Data analysis was conducted using IBM SPSS 22.0 (SPSS, Inc., Chicago, IL). F-statistics and Chi-square statistics were used to test for differences across experimental conditions on possible covariates. Preliminary analyses showed that participants who were exposed to a combination of recommendations and warnings (either four or eight tips) did not differ in the credibility evaluation of the lower quality website, $t(162) = -0.719$, $p = 0.473$, of the higher quality website, $t(162) = 0.028$, $p = 0.978$, and in their discrimination ability, $t(162) = 0.645$, $p = 0.520$. The two groups were therefore merged for subsequent analyses.

Participants in the four conditions did not differ on gender ($\chi^2 = 4.333$, $p = 0.228$), age, $F(3, 399) = 1.133$, $p = 0.335$, $\eta_p^2 = 0.01$, educational level ($\chi^2 = 0.827$, $p = 0.843$), functional health literacy, $F(3, 399) = 1.872$, $p = 0.134$, $\eta_p^2 = 0.01$, eHealth literacy, $F(3, 399) = 1.118$, $p = 0.303$, $\eta_p^2 = 0.01$, general Internet use ($\chi^2 = 10.543$, $p = 0.315$), online health information seeking ($\chi^2 = 5.326$, $p = 0.149$), perceived knowledge about type 2 diabetes, $F(3, 399) = 0.080$, $p = 0.971$, $\eta_p^2 = 0.00$, personal relevance of type 2 diabetes, $F(3, 399) = 0.998$, $p = 0.394$, $\eta_p^2 = 0.01$, and previous knowledge of the websites (Lower quality website: $\chi^2 = 1.700$, $p = 0.637$; Higher quality website: $\chi^2 = 0.444$, $p = 0.931$). We therefore did not include any covariates in the statistical analysis. To explore the effect of the different messages and health literacy on the credibility evaluation of the two websites and on the participants' ability to correctly discriminate between websites of different quality, a multivariate analysis of variance (MANOVA) was performed with functional health literacy and condition as independent factors and credibility evaluations of the two websites and quality discrimination ability as dependent variables.

3. Results

Our first hypothesis stated that providing consumers with easy-to-use information about evaluation criteria would improve their ability to evaluate the credibility of health websites (*H1a*) and to distinguish between high and low quality health websites (*H1b*). The MANOVA showed an overall significant main effect of the type of message on the independent variables, $F(6, 790) = 2.743$, $p = 0.012$, $\eta_p^2 = 0.020$. Tests of between-subjects effects and subgroup analyses using Bonferroni correction for multiple comparisons showed that the people who were exposed to the *recommendations* ($M_{rec} = 4.88$, $SD = 1.27$) or a *combination* of recommendations and warnings ($M_{com} = 4.92$, $SD = 1.13$) assigned lower credibility to the lower quality website compared to those in the *control* condition ($M_{contr} = 5.47$, $SD = 0.98$), while no significant effect was found for the message that only presented *warnings* ($M_{warn} = 5.21$, $SD = 1.17$), $F(3, 395) = 4.857$, $p = 0.002$, $\eta_p^2 = 0.036$.

Similar results were found with respect to people's ability to discriminate between the higher and lower quality websites: those in the *recommendations* ($M_{rec} = 0.55$, $SD = 1.15$) and *combination* ($M_{com} = 0.43$, $SD = 1.30$) conditions showed higher discrimination ability compared to those in the *control* condition ($M_{contr} = -0.05$, $SD = 0.91$), while no significant effect was found for the *warnings* condition ($M_{warn} = 0.21$, $SD = 1.16$), $F(3, 395) = 4.089$, $p = 0.007$, $\eta_p^2 = 0.030$. No significant effect of the manipulation was found on the credibility evaluation of the higher quality website, $F(3, 395) = 0.107$, $p = 0.956$, $\eta_p^2 = 0.001$. Providing recommendations about evaluation criteria therefore had a positive impact on the credibility evaluation of the low quality website and on the overall quality discrimination ability. As no effect was observed on credibility evaluation of the high quality website, *H1a* and *H1b* were partly supported by our data.

We subsequently hypothesized that adding warnings against using non-established criteria to educational messages recommending evaluation criteria would be more effective than only presenting evaluation criteria in influencing people's ability to evaluate the credibility of OHI (*H2a*) and to distinguish between high and low quality health information (*H2b*). No significant difference was found between the effects of the *recommendations* and the *combination* condition on the outcomes. The two messages have therefore to be considered equally effective, so *H2a* and *H2b* were not supported by our data.

According to our last set of hypotheses, we expected that providing easy-to-use recommendations about established evaluation criteria would particularly improve people with low health literacy's ability to evaluate the credibility of OHI (*H3a*) and to distinguish between high and low quality health information (*H3b*). Furthermore, we expected that adding warnings against the use of non-established criteria to educational messages would be particularly effective for people with limited health literacy's in improving their ability to evaluate the credibility of OHI (*H3c*) and to distinguish between high and low quality health information (*H3d*). Although we found no overall significant effect of the interaction between the type of educational message and health literacy on the outcomes, $F(6, 790) = 1.317$, $p = 0.247$, $\eta_p^2 = 0.010$, subgroup analyses using Bonferroni correction revealed some significant simple effects. Regarding the evaluation of the lower quality website, the difference between the group in the *recommendations* ($p = 0.004$) and *combination* conditions ($p < 0.001$) and the *control* group was significant only among those with higher health literacy. A similar result was found regarding discrimination ability, where the difference between the groups in the *recommendations* ($p = 0.019$) and *combination* conditions ($p = 0.057$) and the *control* group was significant (or marginally significant) only among those with higher health literacy. *H3a, H3b, H3c, and H3d* were therefore not supported. Mean evaluations for the two websites and quality discrimination ability for the four experimental conditions and health literacy groups can be found in Table 2.

4. Discussion

Although the improvement of consumers' ability to evaluate OHI has been one of the main goals of public health for several years, there is still a lack of theory- and evidence-based interventions in this area (Car et al., 2011). The present study aims to enrich the existing evidence base in this context. This was done, first, by investigating whether educational messages recommending existing quality criteria for OHI to consumers could be a viable way to improve their online health information evaluation skills. As it has been suggested that merely informing people about how to evaluate the credibility of online information is probably not sufficient (Metzger, 2007), and that people often use subjective criteria or heuristics to evaluate OHI (Diviani et al., 2016), our study also assessed the effectiveness of additional warnings against the use of evaluation criteria not recognized as valid by established guidelines. This was expected to increase people's motivation to carefully evaluate the credibility of online sources and to yield stronger effects of the educational message. Last, we investigated whether health literacy plays a role in the effectiveness of the different types of messages.

Our study showed that when messages combined recommendations about established evaluation criteria with warnings, these were not more effective than educational messages alone. The added value of warnings, which was expected on basis of the dual

Table 2
Websites evaluation and quality discrimination ability for experimental conditions and health literacy groups.

	Credibility low quality website			Credibility high quality website			Quality discrimination ability		
	Lower HL	Higher HL	Total	Lower HL	Higher HL	Total	Lower HL	Higher HL	Total
Control	5.186 (0.192)^{a}	5.700 (0.170)^{a}	5.443 (0.128)^{a}	5.157 (1.68)^{a}	5.633 (0.149)^{a}	5.395 (0.112)^{a}	−0.029 (0.198)^{a}	−0.067 (0.175)^{a}	−0.048 (0.132)^{a}
Recomm.	4.864 (0.198)^{a}	4.888 (0.166)^{b}	4.876 (0.129)^{b}	5.265 (0.173)^{a}	5.548 (0.145)^{a}	5.407 (0.113)^{a}	0.402 (0.204)^{a}	0.660 (0.171)^{b}	0.531 (0.133)^{b}
Warnings	5.329 (0.192)^{a}	5.114 (0.172)$^{(a)b}$	5.221 (0.129)^{ab}	5.264 (0.168)^{a}	5.545 (0.150)^{a}	5.405 (0.113)^{a}	-0.064 (0.198)^{a}	0.432 (0.177)^{ab}	0.184 (0.133)^{ab}
Combinat.	5.014 (0.135)^{a}	4.841 (0.118)^{b}	4.928 (0.090)^{b}	5.359 (0.118)^{a}	5.331 (0.103)^{a}	5.345 (0.079)^{a}	0.345 (0.139)^{a}	0.489 (0.122)$^{(a)b}$	0.417 (0.092)^{b}

Notes: Values in the table are group means (std. errors); HL = health literacy; Means with different superscripts within a column indicate significant difference ($p < 0.05$); Superscripts in parentheses indicate marginally significant differences ($p < 0.10$); Adjustment for multiple comparisons: Bonferroni.

processing model of Web site credibility assessment (Metzger, 2007), was therefore not supported by our results. One potential explanation could be that our choice to present warnings against using non-established criteria was not effective enough. Although many people frequently use non-established criteria to evaluate OHI, it could be that warnings against the negative consequences of a specific behavior (e.g., acting upon unreliable health information) raise more awareness and are therefore more effective. Another explanation is that we did not measure people's actual level of motivation to evaluate OHI credibility during the experiment. Consequently, it is possible that people were already highly motivated to evaluate the information (e.g., because of demand effects of doing an online survey) and that the warnings could therefore not have an additional effect. Future research is needed to further investigate the mechanisms that underlie the process of credibility evaluation. Specific attention needs to be devoted to the role of consumers' ability and motivation to evaluate OHI and the way in which both factors can be influenced.

Furthermore, our study showed that educational messages - with or without additional warnings - are only effective among people with higher levels of health literacy. This contradicts our hypotheses and is unfortunate, as especially people with low levels of health literacy usually have problems with evaluating OHI credibility (Jiang & Beaudoin, 2016). In the current study, we aimed to optimize the messages by tailoring them to the needs of people with low health literacy levels by formulating them in a clear and simple way (Meppelink, Smit, Buurman, & van Weert, 2015). Apparently, other message strategies are needed to improve this evaluation skill among people with low health literacy. Research has shown that people with limited health literacy find it easier to process complex information when it is complemented with, or presented as, illustrations or animations (Bol, van Weert, de Haes, Loos, & Smets, 2015; Meppelink et al., 2015; Meppelink & Bol, 2015). Future studies should therefore test the effects of different message formats, by adding for instance illustrations or videos to the information.

Our study showed a significant effect of the educational messages on the evaluation of the lower quality but not of the higher quality website. A possible explanation for this finding could be that the higher quality website presented some quality markers, which people are likely to recognize independently from the experimental message they were exposed to. For instance, every piece of information on the website was followed by the name and qualifications of its authors. We know from past research that authorship is among the few established quality markers that are widely known and applied by consumers of OHI (Diviani et al., 2016). In addition, the presence of the HON certificate on the homepage of the high quality website might have played a role. However, it has been shown that very few consumers are aware of the existence of such certifications and check for them systematically (Diviani et al., 2016; Eysenbach & Kohler, 2002).

Our study has some limitations. First, it should be noted that the actual accuracy of the information presented on both websites that were used in this study was not formally assessed. Classification of the websites as high or low in quality was based on the presence (or absence) of established quality criteria. We did this on purpose, because many people lack the detailed medical knowledge that is required to determine whether OHI is completely accurate. Only medical experts in a specific field could do this in theory. For this reason, it is particularly important that people use the correct criteria to help them determine whether the information they encounter is reliable. As a result of this choice, however, it is possible that the content of the low quality website was indeed accurate and true. As people also use their prior knowledge when evaluating OHI, it could be that people used their knowledge about diabetes to evaluate both websites and decided that both websites were equally credible. Second, we recognize that, although it was functional to increase the ecological validity of the experiment, the choice of using existing websites as stimuli for the experiment might have limited the extent to which we were in control of the experiment itself. We checked the homepages of the websites prior to and after the experiment to make sure that the information did not change, but we cannot fully rule out changes beyond the homepage which might have had an impact on the participants' evaluation of the websites. On a related note, we acknowledge that some factors could have impacted the judgment of the websites, such as varying page loading speed or browser-related visualization differences. The validity of our results would therefore highly benefit from replication in a more controlled environment.

5. Conclusion

Based on the results of the study, we are able to conclude that providing consumers with criteria to evaluate OHI could be a viable way to improve their evaluation skills and could therefore contribute to preventing negative health outcomes related to using incomplete or wrong information. Furthermore, our results suggest that warnings against the use of non-established criteria people usually apply has no added value. However, before these results can be implemented in the development of the often called for evidence-based communication interventions to enhance consumers' ability to appraise OHI, further research is needed to identify efficient ways to communicate these criteria to low health literate audiences.

Author note

The present study was funded by a grant awarded to the first author by the Swiss National Science Foundation (P300P1_158408). The funding source had no role in study design, data collection, analysis, interpretation, or writing of the report.

Appendix 1

Educational messages used in the experiment

Recommendations message

TIP #1: **Certification is the key!** Several institutions provide a quality certification to high quality websites. If the website you are visiting does not show any certification, you are probably visiting a low quality health website.
TIP #2: **Always look for the authors!** Good health websites are usually written by identifiable medical professionals or institutions and include references to medical publications.
TIP #3: **Beware of hidden ads!** Ads are not necessarily bad news. If ads and website contents are clearly distinguishable, you are probably on a good health website.
TIP #4: Text is not enough! Good health websites usually explain difficult terms and concepts, for instance with the help of images or videos.

Warnings message

TIP #1: **Don't let the looks fool you!** Good web-designers are not medical professionals, so a serious-looking website is not necessarily a good health website.
TIP #2: **Be ready to be proven wrong!** If the information on a website confirms your idea or your opinion this does not necessarily mean that it is also correct. Conversely, do not dismiss a website only because you do not agree with what it says. .
TIP #3: **Most used does not mean best!** There are several reasons why a health website attracts several users. High quality is not necessarily the main reason.
TIP #4: Don't let your emotions decide! Gut feelings might help you in many situations. When evaluating a health website, however, it is wise to stick to more objective criteria.

Combination message

Included either all eight tips or a random selection of two tips from the *recommendations* message and two tips from the *warnings* message. Preliminary analyses showed that participants exposed to one of the two *combination* messages did not differ in the evaluation of the low and high quality website, and in their discrimination ability. The two groups were therefore merged for all analyses.

References

Allam, A., Sak, G., Diviani, N., & Schulz, P. J. (2017). Do quality markers for health websites affect the perception of vaccination webpages? *Computers in Human Behavior, 67*, 273–281. http://dx.doi.org/10.1016/j.chb.2016.11.003.

Berkman, N. D., Davis, T. C., & McCormack, L. (2010). Health literacy: What is it? *Journal of Health Communication, 15*(S2), 9–19. http://dx.doi.org/10.1080/10810730.2010.499985.

Berkman, N. D., Sheridan, S. L., Donahue, K. E., Halpern, D. J., & Crotty, K. (2011). Low health literacy and health outcomes: An updated systematic review. *Annals of Internal Medicine, 155*(2), 97–107. http://dx.doi.org/10.7326/0003-4819-155-2-201107190-00005.

Berland, G. K., Elliott, M. N., Morales, L. S., Algazy, J. I., Kravitz, R. L., Broder, M. S., … Watkins, K. E. (2001). Health information on the internet: Accessibility, quality, and readability in English and Spanish. *JAMA, 285*(20), 2612–2621. http://dx.doi.org/10.1001/jama.285.20.2612.

Bernstam, E. V., Walji, M. F., Sagaram, S., Sagaram, D., Johnson, C. W., & Meric-Bernstam, F. (2008). Commonly cited website quality criteria are not effective at identifying inaccurate online information about breast cancer. *Cancer, 112*(6), 1206–1213. http://dx.doi.org/10.1002/cncr.23308.

Bol, N., van Weert, J. C., de Haes, H. C., Loos, E. F., & Smets, E. M. (2015). The effect of modality and narration style on recall of online health information: Results from a web-based experiment. *Journal of Medical Internet Research, 17*(4), e104. http://dx.doi.org/10.2196/jmir.4164.

Burbules, N. C. (1998). Rhetorics of the Web: Hyperreading and critical literacy. In I. Snyder (Ed.), *Page to screen: Taking literacy into the electronic era* (pp. 102–122). London: Routledge.

Car, J., Lang, B., Colledge, A., Ung, C., & Majeed, A. (2011). Interventions for enhancing consumers' online health literacy. *Cochrane Database of Systematic Reviews, 6*. http://dx.doi.org/10.1002/14651858.CD007092.pub2.

Cline, R. J., & Haynes, K. M. (2001). Consumer health information seeking on the internet: The state of the art. *Health Education Research, 16*(6), 671–692. http://dx.doi.org/10.1093/her/16.6.671.

Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PloS One, 8*(3), e57410. http://dx.doi.org/10.1371/journal.pone.0057410.

Diviani, N., van den Putte, B., Giani, S., & van Weert, J. C. M. (2015). Low health literacy and evaluation of online health information: A systematic review of the literature. *Journal of Medical Internet Research, 17*(5), e112. http://dx.doi.org/10.2196/jmir.4018.

Diviani, N., van den Putte, B., Meppelink, C. S., & van Weert, J. C. M. (2016). Exploring the role of health literacy in the evaluation of online health information: Insights from a mixed-methods study. *Patient Education and Counseling, 99*(6), 1017–1025. http://dx.doi.org/10.1016/j.pec.2016.01.007.

Eysenbach, G., & Köhler, 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*(7337), 573–577. http://dx.doi.org/10.1136/bmj.324.7337.573.

Eysenbach, G., Powell, J., Kuss, O., & Sa, E. R. (2002). Empirical studies assessing the quality of health information for consumers on the world wide web: A systematic review. *JAMA, 287*(20), 2691–2700. http://dx.doi.org/10.1001/jama.287.20.2691.

Feufel, M. A., & Stahl, S. F. (2012). What do web-use skill differences imply for online health information searches? *Journal of Medical Internet Research, 14*(3), e87. http://dx.doi.org/10.2196/jmir.2051.

Flanagin, A. J., & Metzger, M. J. (2000). Perceptions of Internet information credibility. *Journalism & Mass Communication Quarterly, 77*(3), 515–540. http://dx.doi.org/10.1177/107769900007700304.

Flanagin, A. J., & Metzger, M. J. (2008). Digital media and youth: Unparalleled opportunity and unprecedented responsibility. Digital Media, Youth, and Credibility. In M. J. Metzger, A. J. Flanagin, D. John, & T. Catherine (Eds.), *MacArthur foundation series on digital media and learning* (pp. 5–28). Cambridge, MA: The MIT Press. http://dx.doi.org/10.1162/dmal.9780262562324.005.

Fogg, B. J., Soohoo, C., Danielson, D. R., Marable, L., Stanford, J., & Tauber, E. R. (2003, June). How do users evaluate the credibility of Web sites?: a study with over 2,500 participants. In *Proceedings of the 2003 conference on Designing for user experiences* (pp. 1–15). ACM. http://dx.doi.org/10.1145/997078.997097.

Fox, S., & Duggan, M. (2013). *Health online 2013*. Washington, DC: Pew Internet & American Life Project.

Gauld, R., & Williams, S. (2009). Use of the internet for health information: A study of Australians and New Zealanders. *Informatics for Health and Social Care, 34*(3), 149–158. http://dx.doi.org/10.1080/17538150903102448.

Grewal, P., & Alagaratnam, S. (2013). The quality and readability of colorectal cancer information on the internet. *International Journal of Surgery, 11*(5), 410–413. http://dx.doi.org/10.1016/j.ijsu.2013.03.006.

Hardey, M. (2001). 'E-health': The internet and the transformation of patients into consumers and producers of health knowledge. *Information, Communication & Society, 4*(3), 388–405. http://dx.doi.org/10.1080/713768551.

Jiang, S., & Beaudoin, C. E. (2016). Health literacy and the Internet: An exploratory study on the 2013 HINTS survey. *Computers in Human Behavior, 58*, 240–248. http://dx.doi.org/10.1016/j.chb.2016.01.007.

Kaicker, J., Debono, V. B., Dang, W., Buckley, N., & Thabane, L. (2010). Assessment of the quality and variability of health information on chronic pain websites using the DISCERN instrument. *BMC Medicine, 8*(1), 59. http://dx.doi.org/10.1186/1741-7015-8-59.

Kim, P., Eng, T. R., Deering, M. J., & Maxfield, A. (1999). Published criteria for evaluating health related web sites: Review. *BMJ, 318*(7184), 647–649. http://dx.doi.org/10.1136/bmj.318.7184.647.

Mackert, M., Kahlor, L., Tyler, D., & Gustafson, J. (2009). Designing e-health interventions for low-health-literate culturally diverse parents: Addressing the obesity epidemic. *Telemedicine and e-Health, 15*(7), 672–677. http://dx.doi.org/10.1089/tmj.2009.0012.

Meola, M. (2004). Chucking the checklist: A contextual approach to teaching undergraduates web-site evaluation. *Libraries and the Academy, 4*(3), 331–344. http://dx.doi.org/10.1353/pla.2004.0055.

Meppelink, C. S., & Bol, N. (2015). Exploring the role of health literacy on attention to and recall of text-illustrated health information: An eye-tracking study. *Computers in Human Behavior, 48*, 87–93. http://dx.doi.org/10.1016/j.chb.2015.01.027.

Meppelink, C. S., Smit, E. G., Buurman, B. M., & van Weert, J. C. (2015). Should we be afraid of simple messages? The effects of text difficulty and illustrations in people with low or high health literacy. *Health Communication, 30*(12), 1181–1189. http://dx.doi.org/10.1080/10410236.2015.1037425.

Metzger, M. J. (2007). Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal of the*

American Society for Information Science and Technology, 58(13), 2078–2091. http://dx.doi.org/10.1002/asi.20672.
Norman, C. D., & Skinner, H. A. (2006). eHEALS: The eHealth Literacy Scale. *Journal of Medical Internet Research, 8*(4), e27. http://dx.doi.org/10.2196/jmir.8.4.e27.
Rieh, S. Y. (2002). Judgment of information quality and cognitive authority in the Web. *Journal of the American Society for Information Science and Technology, 53*(2), 145–161. http://dx.doi.org/10.1002/asi.10017.
Scholz-Crane, A. (1998). Evaluating the future: A preliminary study of the process of how undergraduate students evaluate web sources. *Reference Services Review, 26*(3/4), 53–60. http://dx.doi.org/10.1108/00907329810307759.
Sørensen, K., Pelikan, J. M., Röthlin, F., Ganahl, K., Slonska, Z., Doyle, G., … Falcon, M. (2015). Health literacy in Europe: Comparative results of the European health literacy survey (HLS-EU). *European Journal of Public Health, 25*(6), 1053–1058. http://dx.doi.org/10.1093/eurpub/ckv043.
Stvilia, B., Mon, L., & Yi, Y. J. (2009). A model for online consumer health information quality. *Journal of the American Society for Information Science and Technology, 60*(9), 1781–1791. http://dx.doi.org/10.1002/asi.21115.
Viswanath, K. (2005). The communications revolution and cancer control. *Nature Reviews Cancer, 5*(10), 828–835. http://dx.doi.org/10.1038/nrc1718.
Wathen, C. N., & Burkell, J. (2002). Believe it or not: Factors influencing credibility on the Web. *Journal of the American Society for Information Science and Technology, 53*(2), 134–144. http://dx.doi.org/10.1002/asi.10016.
Weiss, B. D., Mays, M. Z., Martz, W., Castro, K. M., DeWalt, D. A., Pignone, M. P., … Hale, F. A. (2005). Quick assessment of literacy in primary care: The newest vital sign. *The Annals of Family Medicine, 3*(6), 514–522. http://dx.doi.org/10.1370/afm.405.
WHO. (2016). *Global report on diabetes*. Geneva: World Health Organization.
Zhang, Y., Sun, Y., & Xie, B. (2015). Quality of health information for consumers on the web: A systematic review of indicators, criteria, tools, and evaluation results. *Journal of the Association for Information Science and Technology, 66*(10), 2071–2084. http://dx.doi.org/10.1002/asi.23311.