Evaluating the Quality of Health Information in a Changing Digital Ecosystem

Abstract

**Background:** Critical evaluation of online health information has always been central to consumer health informatics. However, with the emergence of new Web media platforms and the ubiquity of social media, the issue has taken on a new dimension and urgency. At the same time, many established existing information quality evaluation guidelines address information characteristics other than the content (e.g., authority, currency), target information creators rather than users as their main audience, or do not address information presented via novel Web technologies.

**Objective:** The objectives of this formative study are to 1) develop a methodological approach for analyzing health-related webpages and 2) apply it to a set of relevant webpages.

**Methods:** The study analyzed twenty-four type 2 diabetes pages which were the results of a Google search with the keywords “diabetes, reversal, natural.” The coding scheme, developed via a combination of theory-driven and data-driven approaches, includes five categories from existing guidelines (resource type, information authority, validity of background information sources, objectivity, currency) and seven novel categories (treatment/reversal method, promises and certainty, criticisms of establishment, emotional appeal, vocabulary, rhetoric and presentation, use of science in argumentation). The coding involves both categorical judgement and in-depth narrative characterization. Upon establishing satisfactory level of agreement on the narrative coding, the team coded the complete dataset of twenty-four pages.

**Results:** The results set included “traditional” static pages, videos, and digitized versions of printed newspapers or magazine articles. Treatments proposed by the pages included a mixture of conventional evidence-based treatments (e.g., healthy balanced diet, exercise) and unconventional treatments (e.g., dietary supplements, optimizing gut flora). Most pages either promised or strongly implied high likelihood of complete recovery. Pages varied greatly with respect to the authors’ stated background and credentials, as well as the information sources they referenced or mentioned. The majority included criticisms of the traditional healthcare establishment. Many sold commercial products ranging from dietary supplements to books. The pages frequently used colloquial language. A significant number included emotional personal anecdotes, made positive mentions of the word “cure,” and included references to nature as a positive healing force. Most pages presented some biological explanations of their proposed treatments. Some of the explanations involved the level of complexity well beyond the level of an educated lay person.

**Conclusions:** Both traditional and data-driven categories of codes used in this work yielded insights about the resources and highlighted challenges faced by their users. This exploratory study underscores the challenges of consumer health information.
seeking and the importance of developing support tools that would help users seek, evaluate, and analyze information in the changing digital ecosystem.

**Keywords:** eHealth, eHealth literacy, diabetes, consumer health information, health literacy, information evaluation, information quality

**Background**

**Introduction**
Evaluating the quality of online consumer health information has been a central issue in consumer health informatics for many years. However, with the emergence of new Web media platforms and the ubiquity of social media, the need for critical evaluation has taken on a new dimension and urgency. Individuals living with life-threatening and chronic diseases search the internet for treatment alternatives. Many such searches lead to sites containing non-evidence-based advice with targeted marketing and clickbait headlines.

Type 2 diabetes is a chronic disease that affects an increasingly high proportion of Americans who have access to a plethora of online resources of highly variable quality. The objectives of this formative study are to develop a methodological approach for analyzing health-related webpages likely to be viewed by individuals with type 2 diabetes and apply it to a set of relevant pages.

**Quality of Online Health Information**
For health information seekers, the World Wide Web can be a source both for valuable information and misinformation. In the early 1990s, Gorden Guyatt was credited with coining the term, evidence-based medicine (EBM), reflecting the hierarchy of scientific evidence employed in the development of clinical advice [1]. In 1997, Medline indexed the first study evaluating consumer health webpages [2]. Published in BMJ, the study reviewed 41 pages with advice on managing children’s fever at home, concluding with alarm that “only a few web sites provided complete and accurate information for this common and widely discussed condition” (page 1875).

Today, the problem persists, and the rapidly growing popularity of social media is making the problem of identifying quality information more pressing. For example, although Facebook has recently made an effort to reduce unsolicited commercial content that appears as news headlines, these are not typically vetted, checked for accuracy or monitored in any way [3, 4]. Many sites prey on vulnerable populations who may be receptive to promises of a quick and easy cure or an alternative to medical establishment recommendations. A recent study found that the majority of the most-shared articles on Facebook in 2016 including the word “cancer” in the headline contained claims discredited by health authorities [5].
Health-related misinformation may be considered on a continuum that ranges from deliberately deceitful with the intent to promote specious products to sites that may more benignly endorse a product or claims that lack scientific credibility. Other sites may advocate for or against a particular health practice. An example of overt fraud is the Web-perpetuated “one-weird trick” meme. These advertisements made promises tailored to specific problems such as reducing belly fat, improving aging skin and boosting testosterone levels [6]. These schemes were immensely successful, capturing over 4 million visitors and bilking health consumers out of nearly half a billion dollars [7]. The epidemic of fake or controversial health news presents formidable challenges as well as interesting research opportunities to the consumer health informatics community.

**Type 2 Diabetes as a Source of Consumer-Targeting Websites**

With its ubiquitous and growing prevalence, diabetes is the focus of many consumer-targeting health information websites. A common theme is “reversal” of diabetes. According to the American Diabetes Association (ADA), diabetes affects 25 million Americans, over 8% of our population. Another 7 million may be undiagnosed, while nearly 80 million adults have impaired glucose tolerance, also known as “prediabetes.” Over 95% of diabetes mellitus is type 2, a complex, polygenic disease characterized by problems with insulin secretion and insulin action, eventually resulting in an elevated blood glucose (hyperglycemia). Over time, diabetes can damage small blood vessels and nerves, causing serious problems in the eyes, heart, brain, kidneys, skin and feet, ranking in the top ten “killer” diseases in the US [8].

Type 2 diabetes is a livelong disease that starts with silent metabolic changes that precede symptoms and frank hyperglycemia by 7-10 years. Classic symptoms include increased thirst, frequent urination, hunger, fatigue, and blurred vision. Treatment includes lifestyle measures (diet, exercise and stress reduction) and medications, including insulin therapy. Type 2 diabetes can be prevented, postponed, and placed into remission by lifestyle measures and therapy. However, except in rare and extreme circumstances, it cannot be cured or reversed. Medical literature reference to diabetes “reversal” are sparse and limited to those on very low carbohydrate diets, post-bariatric surgery, or in experimental animal models. Therefore, frank “reversal” of the complex metabolic derangements of type 2 diabetes is uncommon and quite difficult to accomplish. The lay use of the term “diabetes reversal” observed in the web sites of interest, therefore, does not signal true medical reversals; instead, these are descriptions of are diseases in remission characterized by reduction or discontinuation of medication.

**Barriers to Effective Functioning in the World of Digital eHealth**

The danger of inaccurate health information is heightened by the public’s potential vulnerability to it. Vulnerability factors vary, from a desperate desire for cure and wellbeing and dissatisfaction with traditional healthcare, to limited health literacy. Health literacy has been defined as “the degree to which individuals have the
capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions” (p. 2) [9]. The magnitude of the problem, affecting more than 90 million Americans, and its impact on healthcare and public health has been well established. Health literacy-related knowledge and skills are particularly deficient among vulnerable populations such as the elderly [10], disadvantaged youth [11], or people with lower levels of education. When individuals with lower health literacy conduct web searches, they rely on criteria that do not correspond to the commonly cited quality guidelines when evaluating online health information [12]. For example, Mackert et al [13] showed that individuals low in health literacy used “position in search results, quality of pictures, celebrity endorsement, and website authorship as criteria to evaluate online health information”.

eHealth literacy has emerged as both a distinct construct and area of research into competencies needed for successful functioning in the world of digital health information [14]. Norman and Skinner introduced an influential eHealth literacy model which comprised the following six types of literacy:

1) **Computer Literacy**: the skills to use computers productively;
2) **Information Literacy**: the skills to articulate information needs, to locate, evaluate, use information, and to apply information to create and communicate knowledge;
3) **Media Literacy**: the ability to select, interpret, evaluate, contextualize, and create meaning from resources presented in a variety of visual or audio forms [15];
4) **Conventional Literacy and Numeracy**: reading comprehension and quantitative skills for interpreting information artifacts such as graphs, scales, and forms);
5) **Science Literacy**: familiarity with basic biological concepts and the scientific method as well as the ability to understand, evaluate, and interpret health research findings using appropriate scientific reasoning; and
6) **Health Literacy**: the acquisition, evaluation, and appropriate application of relevant health information as described previously.

Chan and Kaufman extended and applied this model to analyze how individuals use Web resources to answer questions across different health topics [16-18]. They identified notable barriers pertaining to basic literacy. However, the most frequently encountered barriers related to information literacy. These resulted in failures to identify relevant links and cues on websites, failing to identify relevant information and difficulty in recognizing what constituted an answer to an information-seeking question. Chan and Kaufman also found barriers related to evaluating the trustworthiness and credibility of health information [11, 19, 20]. On the other hand, the skills associated with media literacy, especially in the context of new media / social media is not well understood [21-23]. Science literacy as it pertains to matters of health, has been an area of inquiry in educational research for some time [24]. However, it has not been given much attention in consumer health research.

**Quality Information Evaluation Guidelines**
Concern over health information quality online has been present from the dawn of the World Wide Web era. The Health on the Net Foundation’s (HON) Code of Conduct, launched in 1996, offered a set of best practice guidelines for website maintainers to follow [25]. To evaluate content itself, the DISCERN instrument was developed in 1996 and 1997 as a joint collaboration between the National Health Service and the British Library. DISCERN was a product of stakeholders chosen from across healthcare: generalist and specialist physicians, but also librarians and health communications specialists, self-help patient group representatives, medical publishers and journalists, and health services researchers [26].

DISCERN’s creators designed it to support websites’ evaluation by health information providers, serve as a checklist for content creators and a training tool for healthcare professionals, and most important, as a decision support for consumers who want to know more about a treatment they are using [27, 28]. While DISCERN was originally designed to target patient-facing leaflets, it can be used to evaluate any text-based information pertaining to treatment. This freely available instrument measures 16 items pertaining to markers of information quality (e.g., reliability, relevance, balance, description of a treatment’s risks and benefits. Today this Web 1.0 tool remains in use, with results reported in over 150 published studies suggesting that DISCERN rankings are similar regardless of whether they are done by patients / consumers or healthcare professionals.

**Study Objectives**
This study aims to characterize health information sources in our ever-expanding digital ecosystem, including non-evidence-based pages presenting information about type 2 diabetes reversal. Specific objectives involve: 1) reviewing top results pages in response to query about natural reversal of type 2 diabetes, 2) developing a methodological approach for capturing their essential content and informational characteristics, 3) testing the approach with the above set of pages.

**Methods**

**Page Selection**
We started with reviewing a collection of non-EBM type 2 diabetes health advice webpages that we have accumulated over the years of exploratory interest in the topic, noting frequent references to “reversal” and “natural” remedies. We then performed a Google search using the keywords “diabetes, reversal, natural” and collected the first three pages of results (31 pages). After reviewing these links, we excluded those that did not pertain to type 2 diabetes, that focused on animals, or that required creating a password-protected login. This resulted in 23 pages. The 24th page, hyperlinked from one of the search results, came from our initial review cluster.

**Coding Scheme Development and the Final Scheme**
The coding scheme was developed using a combination of theory-driven and data-driven approaches [29]. First, the authors reviewed existing information evaluation criteria from various information science instruments, including the DISCERN instrument described above. Next, the authors conducted several rounds of review of three pages from the study set, noting and discussing perceived relevant characteristics. This resulted in the final coding scheme that included both established categories (that is, represented in existing guidelines) and some novel categories, as described in Table 1.

Our aim was to conduct detailed descriptive analysis that frames the user’s information-seeking experience. While most questions are phrased to require binary (yes / no) responses, coding also involved writing short narrative responses (e.g., “although citations are not provided, studies are described with partial information that would enables studies to be found eventually”). Three team members coded the data. To establish inter-coder agreement, thirteen pages were reviewed by two coders (in different permutations), which was followed by iterative in-depth team review and discussion of the coding. While the narrative data were not amenable to inferential statistical analysis, the team found the level of agreement satisfactory; disagreements were resolved via discussion and the narratives were merged. The remaining pages were each reviewed by one coder. With the exception of the Information Authority and Objectivity coding that involved reviewing the parent site’s “about” page, all the coding was done based on the information within the page only. The pages were coded during July – November of 2017, as they appeared at the time.

Table 1. Final coding scheme

<table>
<thead>
<tr>
<th>Source</th>
<th>Category</th>
<th>Subcategories</th>
</tr>
</thead>
<tbody>
<tr>
<td>EG</td>
<td>Resource type</td>
<td>Digitized content (simultaneously published in traditional mass media)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Static webpages</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Web 2.0 content (Wikipedia; blogs; support groups; online communities; listservs; social networking sites; RSS feeds; YouTube)</td>
</tr>
<tr>
<td>DD</td>
<td>Treatment/reversal method</td>
<td>What is the proposed diabetes treatment / reversal method?</td>
</tr>
<tr>
<td>DD</td>
<td>Promises and certainty</td>
<td>Does the page make a claim of having a solution (approach or product) producing results that are:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Quick</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Painless / noninvasive / implemented via a simple procedure or with simple ingredients</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Relatively inexpensive</td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td><strong>Is there a promise of complete recovery for a condition that is known to be chronic / incurable?</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>EG</strong></td>
<td>Information authority</td>
<td></td>
</tr>
<tr>
<td><strong>Validity of background information sources</strong></td>
<td>Is there content that needs to be cited, but isn't? Are the sources for factual information clearly listed / cited so they can be verified in another source? (Subsumes: are authors of testimonials and stated support verifiable?) Is the information from sources known to be reliable? Do citations / references actually support the information presented on the page? Are there endorsements by celebrity non-experts? Who is referred to as “Dr” or “physician”? Is there mention of “secret recipe” (“virtually unknown method”) known only to the site’s owners/promoters? Is there a disclaimer on the page (what does it state)?</td>
<td></td>
</tr>
<tr>
<td><strong>EG</strong></td>
<td>Objectivity</td>
<td></td>
</tr>
<tr>
<td><strong>Currency</strong></td>
<td>Are there dates on the page to indicate when the page was written, when the page was first placed on the Web, or when the page was last revised?</td>
<td></td>
</tr>
<tr>
<td><strong>DD</strong></td>
<td>Criticisms of establishment</td>
<td></td>
</tr>
<tr>
<td><strong>Are there implication or statement of conspiracy or purposeful misleading on the part of:</strong> Pharmaceutical companies? Doctors / conventional healthcare providers? Government agencies? Are there suggestions of media bias in covering relevant health issues? Are there implication or statements that the reader’s / viewer’s doctor is incompetent? Are there criticisms of biomedical research supporting “the establishments’ guidelines” (e.g., methodology, research focus on medications, etc.)?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
due to funding, etc.).

| DD     | Emotional appeal | Does the page contain emotional testimonies / personal anecdotes?  
|        |                  | Does the page contain disturbing photos/ images of healthcare professionals and procedures? |
| DD     | Vocabulary       | Does the page refer to “cure” or other words that are unlikely to be used in evidence-based medical literature (e.g., “proven”)? |
| DD     | Rhetoric and presentation | Are there cliff-hangers in the content? For example, “In the next few minutes, I’m going to share with you the little-known natural remedy that will help you leave pills and needles behind forever…”  
|        |                  | Is there an appeal to buy something right away?  
|        |                  | Is the language very colloquial?  
|        |                  | Is there a long speech that culminates in a request for money? |
| DD     | Use of science in argumentation | Are biological mechanisms of diseases and treatments presented?  
|        |                  | Are there claims that the coder perceives as: Exaggerated? False? unverifiable? Mention controversial / not quite scientific concepts? Otherwise problematic?  
|        |                  | Is there a contrast between claims about the complexity / uncertainty of the condition / treatment and the simplicity and certainty of the proposed solution? (statements that are too good to be true?) |

a EG – based on existing guidelines.  
b DD – data-driven.  

Results  

Types of Online Information Resources  
Twelve of the 24 pages were “traditional” static pages, eight were web 2.0 sources or static pages with video components. Four pages were digitized versions of printed newspaper or magazine articles. Three of these were patient testimonials.  

Diabetes Treatment/Reversal Methods  
The number of remedies proposed by each page ranged from 1 to 6, with the mean of 2.3 (Table 2). The two most commonly mentioned remedies involved 1) adhering to some general nutritional guidelines / healthy eating and 2) taking dietary supplements. General nutritional guidelines typically mentioned avoiding refined sugars and grains and eating more fiber and healthy fats. One page stated that a vegan diet was essential, describing an uncited study linking diabetes risk with consuming animal products. Dietary supplements recommended by the pages included a range of herbs, vitamins and minerals, with several mentions of
cinnamon, turmeric, and chromium picolinate. These recommendations conflict with the American Diabetes Association statement that “research has not been able to prove that dietary or herbal supplements (including omega-3 supplements, cinnamon, and other herbs) help to manage diabetes” [30].

Other common methods, mentioned by a quarter to a third of the pages, involved 1) exercise, 2) special nutritional protocols / specific “superfoods”, and 3) caloric restriction / intermittent fasting. Unlike general healthy eating guidelines, “superfoods” recommendations focused on specific “healing properties” of a particular food (e.g., grapefruit). A special nutritional protocol advertised by one of the pages promised “to kill the microbes and parasites (e.g. pancreatic flukes) identified by the consultation.” The undescribed protocol, apparently available from clinics promoted by the page, had to be followed by “electromedicine” said to “use gentle electrical waves to do the things necessary to rebuild the immune system.”

Table 2. Numbers of pages (of 24) proposing specific remedies

<table>
<thead>
<tr>
<th>Remedy</th>
<th>Pages mentioning the remedy</th>
</tr>
</thead>
<tbody>
<tr>
<td>General nutrition guidelines / healthy eating</td>
<td>13</td>
</tr>
<tr>
<td>Supplements</td>
<td>13</td>
</tr>
<tr>
<td>Exercise</td>
<td>8</td>
</tr>
<tr>
<td>Special nutritional protocol; specific “superfoods”</td>
<td>7</td>
</tr>
<tr>
<td>Caloric reduction / intermittent fasting</td>
<td>6</td>
</tr>
<tr>
<td>Stress reduction</td>
<td>3</td>
</tr>
<tr>
<td>Improved sleep</td>
<td>2</td>
</tr>
<tr>
<td>Electromedicine</td>
<td>1</td>
</tr>
<tr>
<td>Optimizing gut flora</td>
<td>1</td>
</tr>
<tr>
<td>Weight loss</td>
<td>1</td>
</tr>
</tbody>
</table>

Promises and Certainty

The specific promises made by the pages varied greatly. Fourteen of 24 either promised complete recovery or strongly implied that it was highly likely. In doing so, they often referred to “reversing” diabetes (e.g., “type 2 diabetes is almost always reversible and this is almost ridiculously easy to prove”). Some expressed very high level of certainty: “If you follow our recommendations to the letter we guarantee that you will eventually be able to throw your medication away and never need it again!” Only one article, authored by a registered dietitian, discussed what it meant by “diabetes reversal,” explaining how “remission” is a more accurate term than “cure.”
Pages often promoted their approaches to “reversing” diabetes as quick, easy, and low-cost. For example, eleven claimed to have a solution guaranteed to work within a specified period of time, from eleven days to three months. While some articles described difficult, extremely low-calorie regimens, ten touted the ease of reversing diabetes (e.g., “ridiculously simple”). Finally, nine stressed that the treatments they proposed were inexpensive (“so inexpensive it might as well be free”).

**Information Authority**

In 21 out of 24 cases, the content had an identifiable author (an individual or an organization). In 16 out of 21, existence and legitimacy (accuracy of self-identification) of the author was supported by a detailed on-site biography and external web presence (e.g., profiles in LinkedIn and online directories, business listings). Eleven pages had a sponsoring or hosting organization that was separate from the author (e.g., a newspaper / magazine, an association, a public television channel).

Eight of the 24 pages had their content authored or verified by someone described as a credentialed physician with an MD or DO (Doctor of Osteopathy) degree. Three more were authored by individuals who self-identified as naturopaths. Of the remaining authors, three were journalists, two were patients offering their testimonies, one a registered dietitian, one a self-described “clinical nutritionist” (certification not stated), and one an “ex-pharmaceutical chemist”. Additionally, four authors were identified as health coaches or health experts without listed medical credentials. For example, one, a former pro athlete, self-described as “one of the most trusted health and fitness experts.” Another self-identified as a health coach and a popular health and lifestyle reporter. Yet another was described as “a catalyst voice” for alternative treatments and a founder of an independent health research foundation. Finally, two of the stated authors emphasized their lay relationship to the content, referring to themselves as average folk or concerned parents.

**Validity of Background Information Sources**

All pages were in the dot-com domain. Eighteen included a link to a page describing the page’s or the sponsoring organization’s goal. These goals varied in specificity, but typically had to do with information provision. Only one page called for legislative advocacy and encouraged readers to take actions such petitioning the Department of Veteran’s Affairs to “employ licensed naturopathic physicians.”

Ten pages sold products ranging from alternative treatments and supplements to books and films. One had a paid access section. Some others did not sell products directly, but contained links to the author’s books for sale or fee-for-service practice. Thirteen pages contained advertisements, which, in three cases, were not clearly differentiated from the page’s content.

**Currency**
In 20 out of 24 cases the page included a date indicating when the page was written, the information was first placed on it, or the page was copyrighted.

Criticisms of the Establishment
Fourteen of the 24 sources made some critical remarks about the pharmaceutical and healthcare establishment. Of these, thirteen suggested malevolent intent / conspiracy on the part of pharmaceutical companies (7), doctors (5), biomedical research (3), media (1) or some other general or specific entity (6). Pharmaceutical companies received the greatest amount of criticism. For example, one page stated that “the pharmaceutical industry is a gigantic machine which has to sustain itself” and asked, “why would these companies be at all interested in truly reversing diabetes? How would that benefit them financially?” Another wrote, “Most big pharma companies don’t know squat about how to reverse your diabetes.”

Two pages, one with a disclaimer, made explicit claims that diabetes medications recommended by health professionals are extremely dangerous and should be avoided (e.g., “Avoid these 3 doctor recommended treatments: oral medication, insulin therapy, [and] other injectables.”) One page suggests that the American Diabetes Association dietary guidelines “seem to serve the medical practitioners more than the patients.” Another criticism warned readers that they’ve been “lied to for years.” In additions to claims of malevolence, five pages suggest medical doctors’ incompetence in diabetes management due to their lack of familiarity with nutrition and herbal medicine (e.g., “doctors get little if any formal nutrition training in medical school”).

Emotional Appeal, Vocabulary, Rhetoric and Presentation
Thirteen pages included language that was judged by the coders as “very colloquial,” appearing to aim for establishing commonalities and rapport with the reader or viewer. For example, one page stated, “First I want you to know that here at [Company Name] we’re really a lot like you. Average people who just so happened to be committed to helping people.” Several pages included emotional personal anecdotes, such as stories of family members suffering from diabetes complications. Because “natural” was part of the query used to identified pages for the study, it is not surprising that positive mentions of “natural remedies” and “natural treatment” were common. Fourteen pages included these terms; eight made positive references to “cure”. Pages referred to nature and “Mother Nature” as a wise positive “healing” force, regaling the reader with “natural remedies,” “natural healing,” “natural processes,” and “natural cures.” Three pages (videos) included the infomercial approach of a long speech culminating in a direct request for making an immediate purchase.

Use of Science in Argument – Explanatory Mechanisms Behind Treatments
Twenty-one of 24 pages presented some biological explanation of their proposed treatment mechanism(s), with most pages including more than one. Often, pages combined mechanisms and methods widely accepted in standard care (e.g., described by the American Diabetes Association) with more uncertain and
controversial ones. Depth of explanations ranged widely, from simple statements that a specific method (e.g., a supplement) “improves sugar metabolism” to detailed explanations of intra-cellular molecular mechanisms, in our view, well beyond the level of comprehension of an educated lay person (e.g., “Acetic acid protects the liver by increasing tolerance of lipogenesis and fatty acid synthesis responsible for improving cholesterol levels.”).

The following explanatory mechanisms were particularly prominent:

- **“Unclogging” liver and pancreas for normal insulin production.** A number of pages recommending low calorie diets or intermittent fasting explained that this method “unclogs” fat from liver and pancreas, thus restoring them to normal functioning essential for insulin production and glycemic control.
- **Reducing blood glucose and improving glucose metabolism (without mentioning insulin).** Many pages explained their treatment methods (e.g., specific foods, supplements) by stating that these methods “reduce[d] blood glucose,” “improve[d] glucose tolerance factor,” or “help[ed] metabolize glucose” without mentioning insulin. Three pages that promoted exercise mentioned that it builds muscle that burns more glucose. In addition, a number of pages included a biological explanation of how foods high in sugar / refined carbohydrates created spikes in blood glucose (sugar) levels and needed to be avoided. These pages also often explained that foods high in fiber were beneficial because they slowed down glucose absorption.
- **Improving insulin secretion and insulin sensitivity.** A number of pages stated that foods or supplements implicated in their treatment methods influenced production or secretion of insulin, improving insulin resistance / sensitivity, and, in some cases, “mimicked insulin.”
- Seven pages related diabetes to inflammation, stating that food or supplements described by them fought it or increased “good bacteria” in the intestinal lining.
- Several pages related diabetes to weakened immune system, “weakened organs” (in particular, liver and pancreas), and “weakened cells”. They proposed that their methods “strengthened” cells and organs. Mechanisms ranged from supplements that “help strengthen the cellular signal” to “electrical waves” that “kill parasites and microbes that weaken organs”.

**Discussion**

Web pages about natural treatment of diabetes analyzed in this study proposed a number of reversal methods and differed greatly in terms of their alignment with accepted standard of care recommendations, promises, levels of certainty, authors’ background, transparency of sources, rhetoric, style, and attitude towards
pharmaceutical and medical establishment. This exploratory study only looked at a small sample of pages pertaining to a single topic, using a largely qualitative methodological approach. Yet, it underscores the challenges of consumer health information seeking and the importance of developing support tools that would help users seek, evaluate, and analyze information in the changing digital ecosystem. Future work within this research program will focus on extending the approach to a number of domains, developing more robust evaluation criteria, and exploring computational approaches to pages’ analysis.

Is the Consumer Health Web Universe a Dangerous Place?
Consequences of following recommendations promoted by the sources analyzed in this study are likely to vary. The core lifestyle modifications recommended by many pages, namely exercise, weight loss, stress reduction, and a healthy balanced diet are evidence-based components of conventional type 2 diabetes regimes. Very low-calorie diets and intermittent fasting may be beneficial, but due to the known risks (including death) they require medical supervision, especially for those with diabetes requiring medication. Supplements (including cinnamon), superfoods and optimization of gut flora, promoted by many pages, have a low health risk, but little to no proven benefit, and can be financially draining. However, as illustrated in Table B, most pages re-package the core lifestyle recommendations and add-on “essential” product purchases, muddying the water. The stakes for harm are even higher when pages promising to reverse diabetes undermine the use of medications, including recommending unsupervised medication discontinuation, and promote an anti-science attitude.

Fit with Existing and Added Health Information Evaluation Criteria
One of the objectives of this work was to develop a methodological approach for analyzing the digital health information sources in the era of online videos and social media. The study suggests that both traditional and data-driven categories of codes (see Table 1) yielded insights about the resources and highlighted challenges faced by their users.

Existing Evaluation Criteria Categories

Information authority. Assessing authority of information authors and sponsors was straightforward, except for the case of determining the ownership of a Youtube channel. In the majority of cases, pages analyzed in this study had clear authorship indicators, with the authors having sufficient web presence to lend credibility to their stated identity and credentials. While a sizable minority of pages (1/3) was authored or verified by credentialed physicians, the majority were created by non-credentialed individuals.

Validity of background information sources. pages in this sample frequently described or mentioned scientific studies without providing references that would allow their unambiguous identification. They also typically did not reference their biological explanations or statements about treatments that were not aligned with ADA guidelines, developed upon an extensive review of the scientific evidence from
peer-reviewed sources [31]. However, the same pages typically included citations of some external sources with background information about type 2 diabetes. These were often high-quality authoritative sources.

While assessing validity of cited information sources was straightforward, the expression of this criteria varied for different publication formats, because of their differing conventions. For example, bibliography-style references are difficult to present in videos and uncommon in newspaper articles, where a detailed description of a background study is a more likely quality indicator. Still, the criterion of validity remains highly relevant, and the lack of credible citations or pointers, or a mismatch between citations and their purported claims, raises concerns about information quality.

**Objectivity.** Across the range of information source types, this criterion was unambiguous. Page’s or sponsor’s goals, typically stated on the site’s About page, as well as the information about sales of relevant products and services, provided information helpful for judging objectivity. Selling services and products such as supplements by default indicated their endorsement.

**Information currency.** Currency turned out to be a challenging criterion because of the range of events that could be time-stamped on the pages. Although most of the pages had a time stamp, these were more likely to be the dates of the page’s copyright than of information authorship.

**Novel Evaluation Criteria Categories**

**Treatment / reversal method and use of science in argumentation.** These categories are discussed together because scientific argumentation usually explained the treatment methods. For this sample, the *Use of Science in Argumentation* code proved valuable for elucidating the challenges facing health information seekers. The pages typically provided some biological information, claiming effects of substances or procedures on insulin production, glucose metabolism, and cells / organs / microorganisms. Often, these explanations blended widely accepted biological mechanisms with controversial ones. While many pages limited their biology to simple causal statements such as “cinnamon improves insulin sensitivity,” others employed complexity well beyond the level of comprehension of an educated lay person. Many of the non-ADA-aligned treatment methods and biological mechanisms mentioned on the pages had corresponding coverage in peer-reviewed science literature, albeit scientific literature described them as more controversial, less certain, and limited to a narrow range of application (e.g., demonstrated effects limited to animal studies or lacking adequate controls). Applying the *Use of Science in Argumentation* code underscores the formidable challenge of supporting lay assessment of plausibility of online health information. It also suggests that the science literacy component of eHealth literacy deserves greater attention.
**Promises.** A promise of recovery from a chronic disease, with a high level of certainty and implied treatment simplicity, was a useful indicator of concern about a page’s quality. Promises was especially troubling when presented as guaranteed within a specific time-frame.

**Criticisms of establishment.** As a criterion, this one is easy to apply, revealing and provides disconcerting information. The importance of the code is illustrated by the disturbingly number of pages making critical comments in our sample (14 of 24). Criticisms of healthcare-related establishment are disconcerting because they attempt to discredit primary sources of evidence-based care, positioning non-evidence-based methods as primary, rather than complementary. In making critical claims, pages often had to tread the line between denouncing some official sources and implying respect for science and evidence and support from some studies. The public’s response to the representation of science, doctors, and pharmaceutical in the digital ecosystem merits further examination.

**Emotional appeal, Vocabulary, Rhetoric and presentation.** General level of colloquialism and informality and the use of personal / emotional anecdotes seemed to be most related to document type, with more informal language used in videos and newspaper articles. Specific high-certainty words such as “cure” and “guaranteed,” on the other hand, are potential quality signifiers that merit further research.

**Science Literacy as a Dimension of eHealth Literacy Revisited**

As mentioned earlier, this study underscores the importance of science literacy as a component of eHealth literacy. The role of science knowledge in daily life has long been debated in the fields of public health and science education. This study illustrates that science literacy, although important, should not be equated with content knowledge. It is often unrealistic to expect lay people to have biomedical knowledge necessary to analyze the argument behind controversial treatment methods. Although some claims may be refuted by high school biology (e.g., diabetes is a disease of “weak organs”), many remain difficult to evaluate even after a thorough analysis of a Pubmed search. In such cases, the relevant aspects of science literacy are not specific content knowledge, but understanding the nature of science and scientific evidence, uncertainty, and the process of biomedical discovery. Such knowledge is likely to trigger skepticism about overgeneralizations, oversimplifications, and exaggerations inherent in many consumer-targeting pages that promise quick and easy fixes of complex health problems. Complexity of science literacy also underscores the importance of promoting traditional information evaluation criteria, such as source authority and objectivity.

**Implications for Research and Practice: Ways to Support Researchers and Consumers**

The methodological approach, described in this study, is a coding scheme designed expressly for research purposes, rather than as a tool for evaluating webpages. Much more work, aimed at expanding, fine-tuning, simplifying, and validating the criteria is needed before this approach can produce a numerical score that could be used to
assess a webpage. Such a tool could be very valuable to both researchers and everyday information seekers. It would be particularly beneficial if developed and validated for a wide range of online information sources, including blogs, message boards, videos and other social media platforms.

In addition to providing an evaluation guide, medical and informatics organizations may investigate developing resources that address common controversial claims. The establishment often ignores non-evidence-based treatment recommendations, despite their visibility in the public domain. For example, the consumer portion of the ADA site dedicates very little space to a discussion about cinnamon. A thorough respectful explanation of why the use of cinnamon should be treated with caution may be a more effective way to help consumers. Computation-based informatics tools may also play a role in helping users evaluate webpages. This study suggests that certain terms and phrases, particularly those indicating high confidence and rejection of traditional medicine, may be alert markers. Research and development into automated language-based categorization may be useful. Finally, this study suggests the importance of science education for the development of science literacy and the potential synergy between classroom science and health informatics.

Acknowledgements
AK, CAS and DRK participated in conceptualizing the study, developing the coding scheme, coding the data, and writing the manuscript. AM participated in conceptualizing the study, reviewing the coding scheme, and writing the manuscript. The work was in part supported by the intramural research program of the National Library of Medicine, NIH.

Conflicts of Interest
None declared.

References
7. Clark T. Jesse Willms, the Dark Lord of the Internet. The Atlantic online 2014; January / February.


