

## The FA4CT Algorithm: A New Model and Tool for Consumers to Assess and Filter Health Information on the Internet

Gunther Eysenbach <sup>a,b</sup>, Maria Thomson <sup>c</sup>

<sup>a</sup> Department. Of Health Policy, Management, and Evaluation, University of Toronto, Canada

<sup>b</sup> Centre for Global eHealth Innovation, University Health Network, Toronto General Hospital, Toronto, Canada

<sup>c</sup> Department of Health Studies and Gerontology, University of Waterloo, Canada

### Abstract

*Background: eHealth-literate consumers, consumers able to navigate and filter credible information on the Internet, are an important cornerstone of sustainable health systems in the 21<sup>st</sup> century. Various checklists and tools for consumers to assess the quality of health information on the Internet have been proposed, but most fail to take into account the unique properties of a networked digital environment. Method: A new educational model and tool for assessing information on the Internet has been designed and pilot tested with consumers. The new proposed model replaces the “traditional” static questionnaire/checklist/rating approach with a dynamic, process-oriented approach, which emphasizes three steps consumers should follow when navigating the Internet. FA4CT (or FACCCCT) is an acronym for these three steps: 1) Find Answers and Compare [information from different sources], 2) Check Credibility [of sources, if conflicting information is provided], 3) Check Trustworthiness (Reputation) [of sources, if conflicting information is provided]. In contrast to existing tools, the unit of evaluation is a “fact” (i.e. a health claim), rather than a webpage or website. Results: Formative evaluations and user testing suggest that the FA4CT model is a reliable, valid, and usable approach for consumers. Conclusion: The algorithm can be taught and used in educational interventions (“Internet schools” for consumers), but can also be a foundation for more sophisticated tools or portals, which automate the evaluation according to the FA4CT algorithm.*

### Keywords:

internet, consumer health informatics, information quality, information retrieval, education

### Introduction

Searching for health information online is often said to be “one of the most popular activities on the Internet”. Such sweeping (and only partially accurate) claims are mostly based on survey data, such as the Pew Internet Report, where people are questioned whether they have “ever looked online for” a certain category of information such as health, entertainment, or shopping. The Pew Internet

Report 2003 [1] found that “fully 80% of adult Internet users, or about 93 million Americans, have searched for at least one of 16 major health topics online” and goes on concluding that “this makes the act of looking for health or medical information one of the most popular activities online, after email (93%) and researching a product or service before buying it (83%).”

*In reality, the question “have you ever used the Internet for y” does not necessarily translate into the prevalence of day-to-day activities. To gauge these, one has to directly observe Web traffic or monitor what people are searching for. Several independent studies using these more “direct” methods to gauge online activities by tapping into the datasets from various search engines, have concluded that the actual volume of health-related searches on the Internet as a proportion of all searches conducted each day is “only” around 5% [2-5], with other areas such as entertainment, shopping, porn, research, places or business being much more popular.*

In summary, survey and search data combined suggest that searching for health information is a popular, but relatively infrequent activity for most people (chronically ill people being a notable exception).

This usage pattern of health information has implications: While people may know where to go for reliable news, weather information, movie reviews, shopping, and business information, medical questions arise infrequently enough so that people not necessarily have a trusted brand names or portal in their mind. While people may be savvy and experienced enough to evaluate the credibility of a general news website or an ecommerce site, they may have insufficient experience and expertise with health websites. Consumers need to be “eHealth literate” in order to succeed in finding and filtering information. “eHealth literacy” [6] consists of six literacy types (traditional, information, media, health, computer, and scientific literacy) which combined form the foundational skills required by consumers’ to engage with electronic health information.

Several attempts have been made to create tools which can be used to educate consumers or which may assist consumers in identifying “credible” information. Most (if not all) previous tools are checklist-like instruments, designed to evaluate information on a webpage or website level.

**Step 1:**  
Enter search terms reflecting the question into Google.  
Find answers on multiple sites and compare results.

Answer A  
URL: http://.....  
Source A:  
.....  
Author A:.....  
Organization A:.....  
Quote:.....  
.....

Answer B  
URL: http://.....  
Source B:  
.....  
Author B:.....  
Organization B:.....  
Quote:.....  
.....

Answer C  
URL: http://.....  
Source B:  
.....  
Author B:.....  
Organization B:.....  
Quote:.....  
.....

Compare answers - bottom line:  
☐ no consensus  
☐ consensus answer: .....

**Step 2:**  
check how CREDIBLE the documents are

Document A

Current : n (-1) (0) y (+1)  
References : n (-1) (0) y (+1)  
Explicit purpose : n (-1) (0) y (+1)  
Disclosure : n (-1) (0) y (+1)  
Interest Disclosure : n (-1) (0) y (+1)  
Balanced : n (-1) (0) y (+1)  
LEvel of evidence\*: e (-1) (0) t (+1)  
\*[e=experiential, t=trials]

CREDIBLE score:.....

Document B

Current : n (-1) (0) y (+1)  
References : n (-1) (0) y (+1)  
Explicit purpose : n (-1) (0) y (+1)  
Disclosure : n (-1) (0) y (+1)  
Interest Disclosure : n (-1) (0) y (+1)  
Balanced : n (-1) (0) y (+1)  
LEvel of evidence\*: e (-1) (0) t (+1)  
\*[e=experiential, t=trials]

CREDIBLE score:.....

Document C

Current : n (-1) (0) y (+1)  
References : n (-1) (0) y (+1)  
Explicit purpose : n (-1) (0) y (+1)  
Disclosure : n (-1) (0) y (+1)  
Interest Disclosure : n (-1) (0) y (+1)  
Balanced : n (-1) (0) y (+1)  
LEvel of evidence\*: e (-1) (0) t (+1)  
\*[e=experiential, t=trials]

CREDIBLE score:.....

Eliminate sites with score 2 or less, compare answers on remaining sites:  
☐ no consensus  
☐ consensus answer: .....

**Step 3:**  
check the source trustworthiness (reputation)  
Enter source/author/organization in Google

URL: http://.....  
Quote about source A: .....  
Deems A not reputable (-1) / neutral (0) / reputable (+1)

URL: http://.....  
Quote about author A: .....  
Deems A not reputable (-1) / neutral (0) / reputable (+1)

URL: http://.....  
Quote about organization A: .....  
Deems A not reputable (-1) / neutral (0) / reputable (+1)  
Reputation Score: .....

URL: http://.....  
Quote about source B: .....  
Deems B not reputable (-1) / neutral (0) / reputable (+1)

URL: http://.....  
Quote about author B: .....  
Deems B not reputable (-1) / neutral (0) / reputable (+1)

URL: http://.....  
Quote about organization B: .....  
Deems B not reputable (-1) / neutral (0) / reputable (+1)  
Reputation Score: .....

URL: http://.....  
Quote about source C: .....  
Deems C not reputable (-1) / neutral (0) / reputable (+1)

URL: http://.....  
Quote about author C: .....  
Deems C not reputable (-1) / neutral (0) / reputable (+1)

URL: http://.....  
Quote about organization C: .....  
Deems C not reputable (-1) / neutral (0) / reputable (+1)  
Reputation Score: .....

Eliminate sites with negative reputation, compare answers on remaining sites:  
☐ no consensus -> repeat search or add hits  
☐ consensus answer: .....

Figure 1 - The FA4CT Algorithm (Worksheet for Consumers)

A recent review of 273 instruments which can be used by patients and consumers to assess the credibility of health information has concluded that “few are likely to be practically usable by the intended audience” [7].

Many or all of today’s tools are cumbersome and time-consuming checklists. They do not adequately take into account the unique features of a digital networked environment, but are still guided or influenced by our thinking about credibility in the “offline”, printed world. The DISCERN instrument, developed for *printed* patient education brochures but advocated by its developers as an evaluation tool for web-based information [8;9] is a prime example. The claim of the DISCERN authors that “there’s nothing radically different about information on the web” [10] illustrates a failure to recognize and to capitalize on the advantages of the Web to use the networked environment itself to assess the credibility of (health) information.

A second generation of educational tools – beyond checklists of authorship and content criteria of web documents – is needed, one that takes into account that consumers are in a *networked, digital* environment, and that credibility evaluation in this medium is a dynamic, interactive, and iterative process. Advantages of a networked environment,

which should be exploited and utilized by educational and technological tools include the ability for users not to rely on only a single source, but to cross-check information on other websites and to check the credibility and reputation of the source using the Web itself. As Meola noted, rather than promoting a mechanistic way of evaluating Internet resources, a contextual approach is needed, which includes for example the possibility to corroborate information on the Web from other sources [11].

## Methods

### The FA4CT approach

In this paper we propose and pilot test a second generation educational model and approach which we call the FA4CT model. This educational model was originally developed in the context of an Internet school for cancer patients (I3MPACT project: Impact of Internet Instructions on Men with Prostate Cancer). FA4CT is intended for use by consumers to find and check medical facts on the Internet. In contrast to earlier approaches, FA4CT is not a checklist, but an intuitive process (or algorithm) which users are instructed to follow when assessing health information on the Web. The algorithm mimics the process expert search-

ers use for information retrieval and fact checking on the Web. For example, journalists use the technique of cross-checking facts from multiple sources to verify the credibility of their sources.

FA4CT (or FACCCT) is an acronym for the three steps suggested in the algorithm: 1) Find Answers and Compare [information from different sources], 2) Check Credibility [of sources, if conflicting information is provided], 3) Check Trustworthiness (Reputation) [of sources, if conflicting information is provided].

The model also recognizes that consumers are usually not primarily interested in assessing the credibility of an entire “website” (or page, or document) as the unit of evaluation, but usually in the credibility of a specific health claim (fact). Thus, in order to use FA4CT, a consumer seeking information on the Internet is instructed to first formulate his factual question as clearly as possible, preferably in a way that allows a yes/no answer. He is then instructed to translate this question into search terms and to conduct an initial Google search query to locate three web sites that contain an answer to their specific medical question. The first key step (**step 1**) for making sure that the information found on the Web is “accurate” is to *compare* (cross-check) the information found on multiple websites. This is a major shift from previous approaches such as DISCERN, where checklists are used to check the credibility of the source and the information itself. In contrast, the FA4CT algorithm suggests a source/information credibility assessment based on a checklist only as a second step, and only if there is no consensus in the three answers provided. In this case, **step 2** suggests to assess each web site using the *CREDIBLE* criteria [12]. The acronym CREDIBLE refers to Current, References, Explicit purpose, Disclosure of sponsors, Interest disclosed and no conflicts found, Balanced, and Level of Evidence. These criteria are based on empirical studies and reflect markers which have in multivariate regression models been shown to be independent predictors for accuracy [12].

Each of the seven criteria has three simple rating options “not fulfilled” (scored with -1), “neutral” (0), and “fulfilled” (+1) with a total possible credible score ranging from -7 to 7.

If after elimination of less “credible” web sites according to these criteria there is still no consensus among the remaining websites, users are in **step 3** asked to enter the name of the source into Google to check what others on the Web have to say about the source, arriving at a *reputation* score. To assess the reputation, for each web page in question the source, author or organization are entered into Google and three new sources commenting on the reputation of the source in question are identified and a quote commenting on the reputation of the source is recorded. Reputation is scored “+1” if there was an explicit statement of trustworthiness, “0” if neutral or “-1” if there was an explicit statement of untrustworthiness. Figure 1 shows the algorithm as worksheet for users. In addition, a more detailed instruction sheet (not shown) is made available to users. It should be noted that the algorithm is designed for educational purposes or for implementation in automated

tools assisting users. Users are not expected to go through these detailed calculations each time they check a fact, rather, they should - by applying the algorithm a few times with an instructor - develop and internalize the process on a more intuitive basis.

### Formative evaluation and ROC evaluation of CREDIBLE checklist

As part of the formative evaluation of the FA4CT algorithm we had to establish 1) how many websites consumers should cross-check to arrive at a valid assessment on the accuracy of a fact, 2) what the optimal cut-off point of the CREDIBLE score from step 2 is, using a ROC (receiver-operating characteristic) curve approach.

Four questions related to a medical fact were used to pilot the FA4CT algorithm, for each question the first six websites resulting from a Google search containing the answer were assessed, resulting in a total of 24 evaluations.

The searches took place on March 3, 2006. The following are the four pilot questions used and their associated answers from gold-standard evidence based resources.

- 1) Do exclusively breastfed babies require vitamin D supplementation? The search terms entered were “vitamin D” “breastfeeding”. The answer as derived from the CMA clinical practice guidelines developed by the Canadian Pediatric Society, which recommends that breastfed babies be given a daily vitamin D supplement until their diet includes a reliable source or they are one year of age.
- 2) Does vaccination cause autism? The search terms used were “vaccination” “autism”. According to the Cochrane database of systematic reviews there is no credible evidence of a link between the MMR vaccine and autism.
- 3) Does Echinacea cure colds? The search terms entered were “Echinacea” “colds”. According to a review from the Cochrane database of systematic reviews there is no clear evidence that Echinacea prevents colds.
- 4) Should statins be taken for high cholesterol? This question was searched using “statins” “cholesterol”. According to the recommendations of the CMA clinical practice guidelines people at high risk should be treated with the equivalent of 40 mg/d of simvastatin.

### Pilot usability test with end-users

Eight participants were recruited using advertising posters distributed throughout three Toronto hospitals as well as the Consumer Health Information kiosk at the Toronto Reference Library. All participants attended an one hour session at the Centre for Global eHealth Innovation, University Health Network. The sessions were conducted from July 25-28, 2006. Computer sessions were held in a usability lab that enabled video, audio and computer recording, in addition to a one-way mirror for one observer to take notes. The session was recorded using Morae software and captured computer screen and key strokes as well as video and audio of the participant. Participants were

encouraged to speak out loud creating a narrative of their actions and decisions. Post session interviews were recorded with a simple hand held recorder.

Each computer session consisted of three tasks. Due to technical problems only five of the eight participants were included in the analysis of task two. All participants received a brief training at the beginning of the session.

**Task One.** This task was designed to test a “forced” step 2 of the FA4CT algorithm (regardless of whether step 1 would have triggered step 2). Participants were asked to rate three pre-selected websites using the CREDIBLE criteria. Each website provided an answer to the dichotomous question ‘Do exclusively breastfed babies require vitamin D supplementation?’ To retrieve these webpages the search terms “vitamin D” “breastfeeding” were entered into Google. Two of the three websites retrieved provided the ‘correct’ answer. This search took place on March 3, 2006. Each participant received a copy of the FA4CT algorithm and a list of the CREDIBLE criteria definitions.

To determine the reliability of the CREDIBLE criteria when used by multiple raters an inter-rater reliability score was calculated for each criteria. This measure was calculated using the Fleiss variation of the Kappa coefficient. Calculations were completed using SAS version 9.1 and the SAS MAGREE macro.

**Task Two.** Participants were asked to use the FA4CT tool for a health-related question of their own choice.

Both tasks were followed by a short semi structured interview and questionnaire to elicit participant feedback regarding the use of the FA4CT tool.

## Results

### Formative evaluation and ROC evaluation of CREDIBLE checklist

A total of 24 websites were assessed representing four different questions queried, 33% (8/24) of these web pages were determined to provide the “wrong” answer as compared to the evidence-based gold-standard, with “breastfeeding and vitamin D” having the most wrong answers at (3/6).

Although the FA4CT algorithm stipulates that step 2 (CREDIBLE evaluation) should only be carried out if there was no consensus among the first 3 websites, for this pilot evaluation a CREDIBLE score was calculated for all 24 webpages in order to determine the optimal cut-off point. The best cut-off point seems to be a threshold of 2 (sites meeting only 2 or less CREDIBLE criteria are considered not credible), where 87.5% of all web pages that contained the correct answer were correctly deemed credible and only 12.5% of web pages that contained the wrong answers were incorrectly labeled as credible.

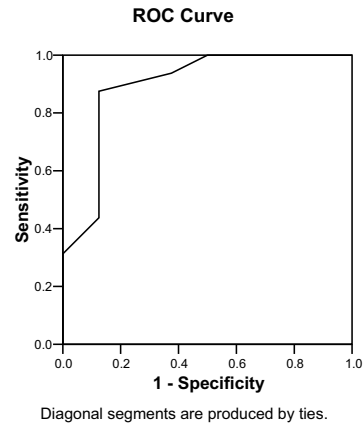


Figure 2 - ROC curve for CREDIBLE score of 24 webpages

Using 3 websites as a starting point, and a CREDIBLE score cut-off point of 2, the FA4CT algorithm performed well. Only 1/8 question sets did not result in a consensus answer at the end of the entire algorithm. Three question sets were correctly answered immediately in step 1 and a further four sets were correctly answered after employing the CREDIBLE criteria in step 2.

### Usability test with end-users

The pilot user sample consisted of an equal number of males and females, with four participants aged 21-40, two participants aged 41 to 60 years and two participants over 61 years. In terms of education, three participants had completed post graduate training, four completed college or university and one completed high school. When asked about their level of confidence using the computer for the session, four reported feeling very confident, two reported that they felt confident and two reported to have some confidence. All participants had at least one year computer experience with six participants reporting more than ten years experience. All participants reported looking for health information on the internet.

### Task one: The CREDIBLE criteria

All eight participants completed task one. On average the time to score one website using the CREDIBLE criteria was 10.25 minutes per page with a range of 5.1 to 18.1 min/page. Because two of the three websites provided a correct answer, it was of interest to identify whether participant criteria scores reflected this distinction. Three participants awarded final criteria scores that correctly classified all three websites therefore awarding passing scores to the two websites with the correct answer and a failing score to the website that provided an incorrect answer. Three participants correctly classified two out of three sites and two participants correctly classified only one site. The site that received the least correct classifications was the site that provided the incorrect answer, according to the gold standard.

Kappa statistics were calculated for each of the seven criteria scored on three webpages by eight raters. Generally the kappa results were satisfactory falling between  $k=0.52$ - $0.79$  (Table 1).

Table 1 - Kappa agreement scores from 8 consumers

CREDIBLE Criterion	Kappa Score
Current	0.62
Reference	0.62
Explicit purpose	0.79
Disclosure of sponsors	0.68
Interest disclosed and no conflicts found	0.68
Balanced	0.53
Level of Evidence	0.52

### Task two: using the FA4CT tool

Five participants properly completed task two, using the entire FA4CT algorithm for their own health-related question. Each participant successfully asked a dichotomous medical question, completed a Google search and located three websites that answered their question. Upon evaluation of the three websites all five participants deemed a consensus answer to have been reached, i.e. no participant had to proceed to step 2 or 3. The average total time spent using the tool was 17.58 minutes with a minimum 10.1 and a maximum 27.2 minutes.

Participant search strategies may have influenced the outcome of the algorithm operation. Four out of five participants typed their medical question into Google using a full sentence complete with a question mark. Only one participant used keywords to search. This may have affected the retrieval of websites by listing first the websites that contained that specific sentence, as opposed to listing websites that contained the search key words without an exact sentence match.

### Discussion

While the FA4CT approach still needs to be refined, the overarching model constitutes a major paradigm change from previous approaches. First, it is a *process*, rather than a mere checklist. While a checklist (the CREDIBLE criteria) is a part of the process, this checklist is only used as second step to eliminate less credible websites in case of lack of consensus among the first 3 sources in step 1, and is in practice rarely needed. Secondly, the approach teaches an evaluation on a fact-level, rather than a “web-site” or document level (though consumers could be taught to evaluate a couple of facts using FA4CT to arrive at a website/document rating). Thirdly, it takes into account the major advantages of information retrieval on the web, which is the ability to cross-check facts using different sources and to check the reputation of sources. Thus it reflects what experienced searchers do when they check the credibility of a (medical or non-medical) claim on the Internet. Consumers using the FA4CT approach are encouraged to check multiple sources and websites to arrive at an answer. They learn to eliminate clearly non-credible sources. In relatively rare instances, discordances will remain after elimination of non-credible sources, leading to the teaching point that in medicine there is often more than one answer, and sometimes even reputable sources contradict each other, which is often a sign of conflicting evidence in the literature.

Our initial experiments with the FA4CT approach have been encouraging. A caveat is that detailed instructions on

how to formulate a search query in a neutral way should be part of the tool, to avoid that people enter preconceived opinions in form of full sentences into Google to find only one-sided answers. Initial findings from user testing also led to fine-tuning of the instrument, in particular related to the CREDIBLE scoring. Checking reputation (step 3) was rarely required in our initial experiments, as in most cases step 1 and step 2 already led to an accurate result, and will require further testing. Reputation checking on Google requires some more advanced search strategies, but might be facilitated by future tools specifically built for checking the reputation of sources (for some time, Google Labs offered such a feature, which is now disabled).

While the algorithm was initially designed to be used as part of educational interventions, it is conceivable that a similar algorithm based on a mix of cross-checking facts and checking of credibility markers and reputation could be part of future automated tools that help consumers to find trustworthy information on the Internet.

### References

- [1] Fox, S. and Fallows, D. Internet Health Resources. <http://www.webcitation.org/5I2STSU61>. 2003.
- [2] Eysenbach G, Kohler C. What is the prevalence of health-related searches on the World Wide Web? Qualitative and quantitative analysis of search engine queries on the Internet. *Proc AMIA Annu Fall Symp* 2003;225-9.
- [3] Eysenbach G, Köhler C. Health-Related Searches on the Internet. *JAMA: The Journal of the American Medical Association* 2004;291:2946.
- [4] Spink A, Yang Y, Jansen J, Nykanen P, Lorence DP, Ozmutlu S et al. A study of medical and health queries to web search engines. *Health Information and Libraries Journal* 2004;21:44-51.
- [5] Pass, G., Chowdhury, A., and Torgeson, C. A Picture of Search. <http://www.webcitation.org/5I2Tms4Qk> INFOSCALE '06 Proc of The First International Conference on Scalable Information Systems, Ma 29-June 1, 2006, Hong Kong. 2006.
- [6] Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. *Journal of Medical Internet Research* 2006;8(4):e27.
- [7] Bernstam EV, Shelton DM, Walji M, Meric-Bernstam F. Instruments to assess the quality of health information on the World Wide Web: what can our patients actually use? *Int J Med Inform* 2005;74:13-9.
- [8] Charnock D, Shepperd S. Learning to DISCERN online: applying an appraisal tool to health websites in a workshop setting. *Health Education Research* 2004;19:440-6.
- [9] Shepperd S, Charnock D, Cook A. A 5-star system for rating the quality of information based on DISCERN. *Health Info. Libr. J* 2002;19:201-5.
- [10] Shepperd S, Charnock D. Against internet exceptionalism. *The BMJ* 2002;324:556-7.
- [11] Meola M. Chucking the Checklist: A Contextual Approach to Teaching Undergraduates Web-Site Evaluation. *Libraries and the Academy* 2004;4:331-44.
- [12] Eysenbach G. Infodemiology: The Epidemiology of (Mis)information. *Am J Med* 2002;113:763-5.

### Acknowledgements

The Internet school was partly funded by the Change Foundation, Toronto (I3MPACT project).

### Address for correspondence

Gunther Eysenbach MD MPH, Centre for Global eHealth Innovation, 190 Elizabeth Street, Toronto M5G2C4, Canada