A quality evaluation methodology of health web-pages for non-professionals

VINCENZO CURRÒ†*, PAOLA SABRINA BUONUOMO†,
ROBERTA ONESIMO†, PAOLA DE ROSE†,
ANDREA VITUZZI‡, GIAN LUCA DI TANNA§ and
ALESSANDRO D'ATRI‡

†Istituto di Clinica Pediatrica, Università Cattolica del Sacro Cuore – Roma
‡Centro di Ricerca sui Sistemi Informativi, LUISS ''Guido Carli'' – Roma
§Scuola di Specializzazione in Statistica Sanitaria, Università di Roma ''La Sapienza''

Abstract. *Primary objective*: The proposal of an evaluation methodology for determining the quality of healthcare web sites for the dissemination of medical information to non-professionals.
Methods and procedures: Three (macro) factors are considered for the quality evaluation: medical *contents*, *accountability* of the authors, and *usability* of the web site. Starting from two results in the literature the problem of whether or not to introduce a weighting function has been investigated. This methodology has been validated on a specialized information content, i.e., sore throats, due to the large interest such a topic enjoys with target users. The World Wide Web was accessed using a meta-search system merging several search engines. A statistical analysis was made to compare the proposed methodology with the obtained ranks of the sample web pages.
Main outcome and results: The statistical analysis confirms that the variables examined (per item and sub factor) show substantially similar ranks and are capable of contributing to the evaluation of the main quality macro factors. A comparison between the aggregation functions in the proposed methodology (non-weighted averages) and the weighting functions, derived from the literature, allowed us to verify the suitability of the method.
Conclusions: The proposed methodology suggests a simple approach which can quickly award an overall quality score for medical web sites oriented to non-professionals.

Keywords: Web quality; Scoring methods; Sore throat; Medical information; Non-professionals.

1. Introduction

The rapid growth of computer-based electronic communication, the fact that new generations are increasingly at ease with the electronic transfer of information, and the shift towards giving patients more responsibility for their health care decisions will have a deep influence on the delivery of health care [1, 2] and in particular for pediatrics [3, 4]. Despite its obvious benefits, it could also create many harmful effects on both consumers and health professionals who do not use the new technology the right way. The problem is not that there is too little information but too much of it, and with vast portions thereof incomplete, misleading and inaccurate [5]. Typing the word 'health', altavista.com retrieves about 25 million web pages and google.com 68 million in January 2002! Many of

*Author for correspondence; CeRSI – LUISS Guido Carli University – Rome, Via 0.Tommasini 1, 00162 Rome – Italy; e-mail: avituzzi@luiss.it

these results are individual pages with limited aims while others are very large web sites containing massive compendia. Size, scope, content, and funding source – each characteristic you can associate with a publication – vary tremendously amongst sites. People are likely to be better informed about their health, but they can be confused by conflicting reports, varied opinions and contradictory recommendations [6–8].

In an Italian survey [9], 19% of the selected families' sample had one member who regularly used the Internet, with a high number of parents accessing medical sites (35% of the Internet users). Most of them (84%) consider the information provided as being useful or very useful.

The use of the Internet for medical and health-related information is being monitored by the Health On the Net Foundation (www.hon.ch), the international non-profit organization that periodically carries out on-line surveys on the use of the Internet for medical/health purposes. The 5th survey's results – October/November 1999 – report that 98% of the 3276 respondents agreed with the statement 'I have found useful medical and health information on the Internet'. But finding quality health web sites among the thousands could be a problem and users must keep in mind that the contents of a page can, and likely will, change over time; this is one of the useful qualities of the web, but it is also a dangerous one. In May–July 2000, in the HON survey on the professional experience of the medical net, 28.82% of the 1129 respondents were 'dissatisfied with the information quality' and considered it as one of the 'obstacles to face when using the Web today'.

Many independent organizations and governmental bodies have developed voluntary standards/guidelines for publishers of scientific journals, e.g.: C-H-I-Q [10], NetScoring [11], and Mitretek [12]. Hence, there are several proposals in literature that investigate the methods for assuring high quality resources and informed usage of them by patients and not by health care professionals, e.g., the POPPi Guide [13], Discern [14], and Quick [15].

Although many attempts have been made to assess, control and assure the quality of web-based information [16], there is no standard strategy to evaluate web pages that are provided to non-professionals.

Our main objective is to provide an evaluation methodology, based on a reduced set of relevant criteria, that is easy to be implemented and understood by non-professional users. This method allows the ranking of medical web sites according to their quality and hence could be used to assist the decision-making processes during the navigation.

The scoring system utilizes a simple arithmetical average, and in this paper we compare it with some results in the literature by using some non-parametric type statistical tests.

2.   Materials and methods

2.1. *Selection of the medical topic for evaluating contents*

The topic we considered is pharyngitis since it is a very common disease in childhood and has a relevant impact for non-professionals (in particular parents) who frequently request information on this subject from pediatricians (hence also on the web). To evaluate the quality of medical contents we compared the web pages with the recommendations of the Committee on Infectious

Diseases of the American Academy of Pediatrics [17], considered as the gold standard.

2.2. *Scoring proposal*

The quality of every web site was scored (S_i) by computing the average value of a suitable set of items, where each item has been evaluated in a range between **0** (the minimum or worst value) and **1** (max or best value). In particular these items have been grouped into three main sections (macro-factors):

- **Medical contents** as the average of:
- *Adherence* of the page to the gold standard. Since the gold standard is composed of 5 factors (epidemiological, clinical, complications, diagnosis, therapy), each factor was independently evaluated: **0** (in case of errors or no adherence); **1/4** (medium-low adherence); **1/2** (medium adherence); **3/4** (medium-high adherence); **1** (full adherence). The average score among the values of the factors considered by the web page was used as the global evaluation of the adherence (note that the missed factors have not been taken into account in the average);
- *Completeness*; determined according to the number of considered/missed factors of the gold standard: **0** (the worst, in case of 1 factor), **1/4** (bad, 2 factors), **1/2** (medium, 3 factors), **3/4** (good, 4 factors), **1** (the best, all factors);
- *References*: **1** if the page lists some pertinent references about its contents, **0** otherwise;
- *Update*: **1** if the page contains the date of creation or of its last update, **0** otherwise;
- **Accountability**, as the average of:
- *author's name,* **1** if he/she is indicated, **0** otherwise (if more than one only the first is considered);
- *author's c.v. or position,* **1** if this additional information is given, **0** otherwise;
- *commercial sponsor,* **1** if there is no sponsorship, **0** otherwise;
- *certification*: **1** if the web site is certified by a certification authority, **0** otherwise;
- *non-profit*: **1** if the site belongs to a not-for-profit (e.g.: .org, .edu, .gov) organization (also if relative to funded projects), **0** otherwise;
- **Usability** as the average of:
- *E-mail*: **1** if the page contains the author's e-mail address (or an appropriate one, to have additional information about the page), **0** otherwise;
- *Aesthetics* **0** (the worst); **1/3** (fair), **2/3** (good), **1** (very good); To evaluate the aesthetics aspects we considered the design of the site and how the graphic elements supported the presentation of the contents.
- *Ease of use as the average of:*
    clicks: **0** if the page is not directly reachable by any navigation (hyperlink) from the home page of the web site, **1/2** if it is reachable by 3 or more clicks, **1** if it is reachable by 1 or 2 clicks;
    intuitiveness: **0** (worst), **1/4** (poor, e.g., search engine only); **1/2** (medium); **3/4** (good); **1** (very good):
    search engine: **1** if such a function is available, **0** otherwise;
- *Links* as the average of:

internal: **0** (no links, i.e., the worst), **1/4** (1 link), **1/2** (2 links), **3/4** (3 links), **1** (4 or more links),

external: similarly evaluated

Only the relevant links to other sources that give more information about the argument explained in the web page were considered in this criterion; for this reason a higher number of links push a site score up.

Finally, the simplest way to determine a global score is to give the same relevance to the three main sections by determining the quality as the average of the contents, accountability, and usability scores. We will denote it as the Average Global Score (AGS).

Furthermore the mean of author's name, author's c.v. or position, and commercial sponsor have been considered as factors in determining as an additional aggregated parameter the *authorship/disclosure*; the mean of certification and no profit in determining the second aggregated parameter *authority* of source, i.e., of the web site.

2.3. *Weighting functions based on the literature*

The second function we considered was obtained from the values of the frequency of explicit criteria for evaluation of health-related web sites by the criteria groups that are reported in [18]. In particular such values were proportionally modified in percentages restricted to the nine considered criteria to obtain for each $1 \leqslant i \leqslant 9$ a weight (W_i) associated with the above scores (S_i) to determine the 'Kim-based' Global Score (KGS):

$$\mathrm{KGS} = \sum\nolimits_{i=1,9} \mathrm{W_i S_i}$$

These weights are: content of site (includes quality, reliability, accuracy, scope and depth): 25.4% corresponds to a frequency of 30 out of 118 citations; design and aesthetics: 18.6%; disclosure of authors, sponsors, developers: 16.9%; currency of information (frequency of update): 11.9%; authority of source: 9.3%; ease of use: 7.6%; links: 4.2%; attribution and documentation (references): 4.2%; contact addresses or feedback mechanism (e-mail): 1.7%. Note that the interpretation of the 'Contents' given in [18] refers only to a subset of the aspects we considered above; in fact it does not contain the References and Update criteria groups which are weighted independently.

The third function 'Netscore' Global Score NGS is based on a (partial) usage of the NetScore [11] weights (restricted to the three main sections we considered). This scoring methodology assigns 79 points to contents (29.5%), 100 points to accountability (37.3%), and 89 to usability (33.2%).

2.4. *Selection of the evaluation sample for the methodology*

To obtain the sample for testing the methodology, the World Wide Web was searched (for the first time in December 2000) by using metaCrawler (www.metacrawler.com); this meta-search system combined several search engines. Because of the differences among the search engines (i.e., Infoseek is a random search tool, Alta Vista ranks the requested references in order of relevance, etc.) a meta-searcher combines such different approaches in a

unique 'independent' collection which could be linked to the way in which any parent with variable experience, culture and computer ability performs the search.

This search was done by asking for links to any of the keywords 'sore', 'throat', 'pharyngitis', and the Italian word 'faringite'.

The results displayed about 150 web sites. As occurs in all Internet searches, some inappropriate documents appeared: we excluded all pages in which the words of our search appeared in a non-medical contest, i.e., we excluded all pages created by commercial ventures or corporate sites sustained by advertising and sales of commercial products (candies, oral spray, herbal tea, etc.) as well as pages related to veterinary purposes. After the exclusion of the inappropriate sources, a total of 55 web sites were included in the study.

The complete alphabetically ordered list of these pages and the related URLs, is available on-line [19]. Since some of these pages are not available today the corresponding files can be obtained by following the related link.

The retrieved pages were created by health care authorities (e.g., American Medical Association, American Academy of Pediatrics), hospitals (e.g., Children's Hospital of Iowa, St. James Hospital), departments of Public Health (e.g., Hawaii, Illinois, Bethesda), Research Foundations (e.g., Mayo Clinic), and other heath care experts (e.g., WellnessWeb, MedicineNet, HealthyNet).

2.5. *Evaluation procedure*

The quality of the web pages' contents with respect to the gold standard was independently evaluated by two MDs (VC a senior pediatrician and PSB a resident). Disagreements were resolved soon after rating by discussion and joint re-reading of the contents. Two information systems scientists (AD′A and AV) similarly evaluated the accountability and usability aspects.

2.6. *Statistical analysis*

To evaluate the aggregated quality factors (e.g., *ease of use* as the average of *clicks, intuitiveness* and *search engine* items) and sections (e.g., *contents* as the average of *adherence, completeness, references,* and *update*) in the above scoring methodology, we based our analysis on the non-parametric statistics and, more precisely, to the family of rank-tests.

With the aim of summarizing the result for each of the factors analysed we calculated a rank variable and then compared the function in pairs using the Wilcoxon signed rank test whose null hypothesis is that both rank-distributions are the same.

Data management and statistical analysis were done using STATA 6 [20].

3. Results

3.1. *Privacy and on-line availability of results*

In order to avoid the disclosure of the individual quality values for each web site and to give general anonymous evaluations, the following results tables do not include the name and URL of the site and are ordered from the best to the worst global score.

In order to retrieve obtained results simply, the complete set of tables related to this paper have been stored on a web site as downloadable .pdf files. An index of

these files is available on-line [19]. In particular there is a file containing all the original collected data for each site and for each scoring variable.

3.2. *Wilcoxon analysis*

From the statistical analysis relative to the three sections of contents, accountability and usability, tests (whose null hypothesis is that the rank distributions are equal) have produced non-significant results (as indicated in the following tables). Therefore, the groupings of indicators we propose are suitable and calculating their arithmetic average can be a valid exercise (see table 1); one exception, which refers to a comparison between the sub-values 'authors' and 'kind of source' in the ambit of the accountability parameter ($p = 0.0132$), seems to be due to factors related to smallness of the sample and the dichotomic nature of the data gathered.

Table 1.   Wilcoxon analysis of contents, accountability and usability factors.

Accountability factors		Contents factors		Usability factors	
Authors vs.		Adherence vs.		Email vs.	
C.V.	0.1162	Completeness	0.6507	Aesthetics	0.9799
Sponsor	0.1432	Reference	0.9765	Ease	0.8536
Certification	0.1843	Update	0.8700	Links	0.8273
Kind of source	0.0132				
C.V. vs.		Completeness vs.		Aesthetics vs.	
Sponsor	0.0863	Reference	0.5184	Ease	0.709
Certification	0.1864	Update	0.8731	Links	0.9599
Kind of source	0.9185				
Sponsor vs.		Reference vs.		Ease vs.	
Certification	0.8641	Update	0.9653	Links	0.9566
Kind of source	0.1809				
Certification vs.					
Kind of source	0.1251				

Table 2.   Wilcoxon analysis of Kim-based Global Score (KGS) factors.

Wilcoxon analysis relative to the weighting functions used		Wilcoxon analysis of Average Global Score (AGS) and NetScore Global Score (NGS) factors	
AGS vs.		Contents vs.	
NGS	0.9131	Accountability	0.9199
KGS	0.5766	Usability	0.7155
NGS vs.		Accountability vs.	
KGS	0.6746	Usability	0.9766

Table 3.   Wilcoxon analysis of Kim-based Global Score (KGS) factors.

	Authority of source	Authorship disclosure	E-mail	Aesthetics	Ease	Links
Contents	0.9	0.8768	0.9532	0.8834	0.9399	0.7501
Authority of source		0.9131	0.2765	0.5735	0.7659	0.6029
Authorship disclosure			0.3964	0.7498	0.7853	0.8966
E-mail				0.9799	0.8536	0.8273
Aesthetics					0.709	0.9599
Ease						0.9566

Table 4.   Rankings of all the sites obtained with all functions used.

AGS	KGS	NGS	AGSRank	KGSRank	NGSRank	CQS	CQSRank
0.80	0.81	0.79	55	54	55	0.80	55
0.79	0.80	0.79	53	53	54	0.79	54
0.77	0.83	0.75	52	55	52	0.78	53
0.79	0.77	0.78	54	52	53	0.78	52
0.72	0.74	0.71	51	50	50	0.72	51
0.71	0.66	0.72	50	48	51	0.70	50
0.69	0.67	0.70	49	49	49	0.69	49
0.65	0.74	0.64	47	51	46	0.68	48
0.67	0.61	0.67	48	44	48	0.65	47
0.62	0.64	0.62	44	47	44	0.63	46
0.64	0.60	0.65	46	41	47	0.63	45
0.62	0.60	0.64	45	42	45	0.62	44
0.58	0.63	0.57	36	46	37	0.59	43
0.58	0.61	0.58	37	43	38	0.59	42
0.59	0.63	0.55	38	45	33	0.59	41
0.60	0.54	0.61	43	34	42	0.59	40
0.59	0.58	0.58	41	39	39	0.58	39
0.59	0.54	0.62	42	33	43	0.58	38
0.59	0.57	0.59	39	38	41	0.58	37
0.59	0.54	0.59	40	35	40	0.57	36
0.55	0.55	0.56	33	36	35	0.55	35
0.55	0.58	0.53	31	40	29	0.55	34
0.56	0.52	0.56	34	29	34	0.55	33
0.54	0.54	0.54	30	32	31	0.54	32
0.57	0.48	0.57	35	22	36	0.54	31
0.55	0.52	0.54	32	28	32	0.54	30
0.53	0.52	0.54	29	30	30	0.53	29
0.52	0.50	0.53	28	25	28	0.52	28
0.50	0.53	0.49	26	31	26	0.51	27
0.51	0.49	0.50	27	24	27	0.50	26
0.47	0.50	0.48	25	26	25	0.48	25
0.44	0.57	0.43	19	37	15	0.48	24
0.46	0.51	0.44	23	27	18	0.47	23
0.45	0.49	0.46	22	23	24	0.47	22
0.44	0.46	0.45	20	19	20	0.45	21
0.44	0.45	0.44	18	18	19	0.44	20
0.47	0.41	0.46	24	13	22	0.44	19
0.43	0.47	0.43	15	20	16	0.44	18
0.43	0.42	0.45	17	15	21	0.44	17
0.45	0.39	0.46	21	8	23	0.43	16
0.42	0.42	0.44	14	14	17	0.42	15
0.41	0.44	0.39	13	17	13	0.41	14
0.43	0.38	0.42	16	6	14	0.41	13
0.39	0.47	0.36	12	21	10	0.41	12
0.39	0.39	0.38	11	9	11	0.38	11
0.35	0.42	0.34	6	16	8	0.37	10
0.35	0.41	0.33	9	12	6	0.36	9
0.37	0.32	0.39	10	5	12	0.36	8
0.35	0.38	0.34	8	7	7	0.36	7
0.35	0.31	0.36	7	4	9	0.34	6
0.32	0.39	0.30	5	10	4	0.34	5
0.31	0.40	0.30	4	11	3	0.34	4
0.30	0.25	0.32	3	3	5	0.29	3
0.19	0.17	0.20	2	1	2	0.19	2
0.18	0.18	0.18	1	2	1	0.18	1

Similar results of substantial equality in the distribution range are shown in the rank test tables relative to the weighting functions we used to determine the global quality score.

Therefore, an estimate of compatible ranges was found with respect to the three variables taken into consideration by us for the determination of global quality according to the AGS and NGS weighting functions (see table 2).

A similar estimate was found also with respect to the nine variables for the determination of overall quality according to the KGS function (see table 3).

3.3. *Global quality evaluation*

NGS and KGS turned out to be roughly equal and a similar parity was obtained for AGS (see table 4).
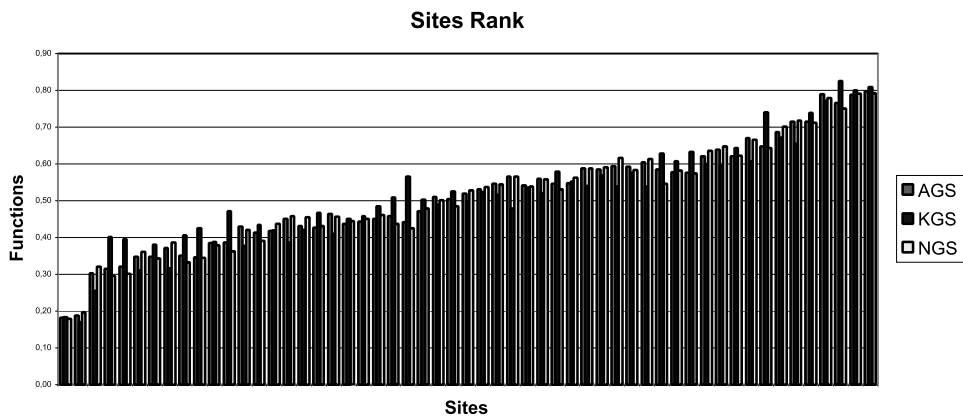
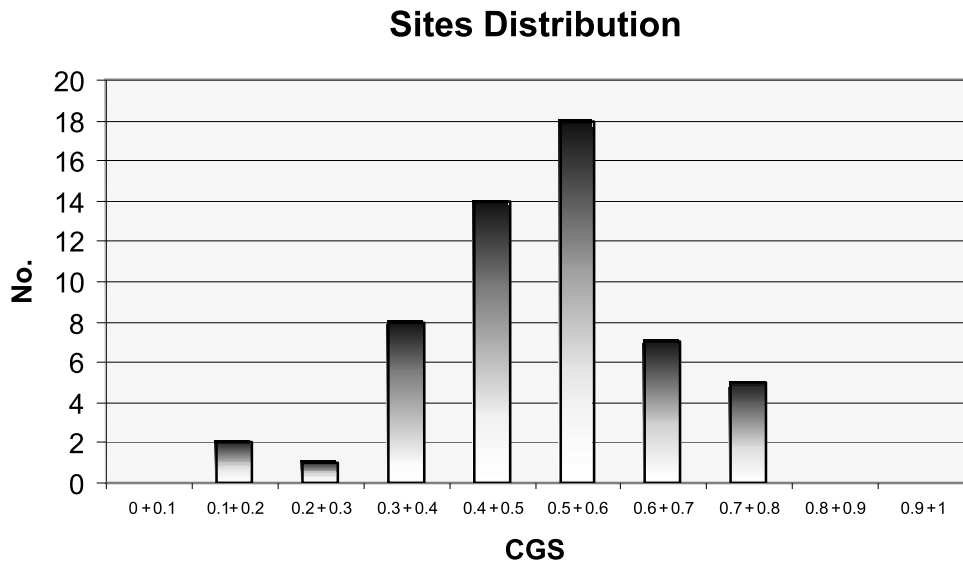Figure 1.   Comparison of the rankings obtained with the three weighting functions.

Figure 2.   Site distribution according to the CGS.

This table and the related graph (see figure 1) show in detail the comparison between the rankings of all the sites for the purposes of best representing the results we obtained.

In particular, to obtain a 'Combined Quality Score' (CQS) we considered in the same table the usage of the average of the scores given by the three weighting functions.

Finally, we evaluated the distribution of the sites (see figure 2) in the sample examined on the basis of the quality values (number of sites that have a 'similar' quality in intervals of 0.1 on the basis of CQS); in such a distribution the threshold values for 10%, 50%, 90% are 0.34, 0.5, and 0.70 respectively.

4. Discussion

Quality benchmarking of health-related resources will always depend on a human assessor, and as such, it will always have attached to it an inevitable element of subjectivity that cannot be simply corrected. Any rating activity gives rise to various risks, including making either falsely positive or falsely negative evaluations; the fact that serious prejudice can be caused to either the users or the author of the site [21] in both cases justifies the attention that has been given to rating activities by various studies.

Current literature is overflowing with references to useful standards to be employed in determining the validity of medical information on the Internet.

One of the most authoritative initiatives is the DISCERN project created by the Division of Public Health and Primary Care at the Institute of Health Sciences of the University of Oxford, England. DISCERN [14] was born with the aim of providing a tool that allows both users and authors to judge the merit of a publication as a source for making a decision in choosing treatment. For this reason a questionnaire was drawn up which considers the aspects relative to the reliability of a publication and the quality of its contents. That initiative suffers from the limitation of non-consideration of usability criteria, which represent a fundamental aspect where medical information on the web is concerned, especially that which is directed at parents.

Another much-cited initiative in this field is OMNI (Organizing Medical Networked Information), which is based on experts' evaluation of on-line medical information through the use of a list of criteria known as the OMNI 'Evaluation Guidelines' [22].

An organization which ranks as one of the pioneers in web-quality in the medical field is HON [23]. The HON Code of Conduct (HONcode) is a self-regulatory, voluntary certification system; it does not rate the medical accuracy, validity or appropriateness of the information itself, but simply defines a set of rules to hold Web site developers to basic ethical standards in the presentation of information and ensure that readers always know the source and the purpose of the information they are reading. The HON code addresses eight points: authority of the information provided, complementarity, confidentiality and privacy, proper attribution of sources, justifiability, transparency of authorship, transparency of financial sponsorship and honesty in advertising and editorial policy with emphasis on the importance of clearly separating advertising from editorial content.

Another frequently cited contribution is that of Silberg [24] who suggests the following four core criteria for measuring the quality of medical Web sites, based

completely on 'accountability – that an identifiable person or groups of people stand behind what is being published on the web'. *Authorship*: the author(s) of a Web page, along with their affiliations and credentials, should be clearly stated, and ideally, there should be the facility to contact the author(s) by e-mail; *Attribution*: if a Web site is quoting research or evidence then the source of this data must be explicitly stated; *Disclosure*: the owner of the Web site must be prominently displayed, along with any sponsorship or advertising deals that could constitute a potential conflict of interest; *Currency*: web pages should indicate when the page was created, and when it was last updated.

Some literature establishes measuring systems (associating numerical points to such parameters) and then provides weighting mechanisms to add up the points in various ways, e.g., Netscoring [11] envisages the use of a set of 49 criteria that can be consistently used to assess the quality of health information on the Internet. These criteria are weighted as 'essential' (up to 10 points), 'important' (up to 5 points) or minor (max 2 points) and are divided into eight categories: credibility, content, links, design, interactivity, quantitative aspects, ethics and accessibility. In our opinion, this (complex) kind of scoring is very useful as self-assessment criteria covering the largest number of parameters for the designer of a new web site but is not easy to use in the evaluation of an already existing site during user navigation since there are aspects (e.g., those related to 'conflict of interest', the 'editorial review process', 'regular verification of hyper-link functionality', 'number of machines visiting the site . . . ') that are often unavailable unless one obtains them directly from the publisher (or from third parties).

Kim *et al*. [18] reviewed literature regarding criteria for specifically evaluating health-related web sites with the aim of identifying areas of consensus. They found 12 main, most frequently cited, categories (content, design and aesthetics of the site, disclosure of authors, sponsors, currency of information – including frequency of update, freshness, maintenance of site – authority of source, ease of use, accessibility, availability, . . .). They suggested identifying and assessing a clear, simple set of consensus criteria that the general public can understand and use.

On the basis of the aforementioned need for simplification, we limited the choice to 16 basic evaluation criteria partially coinciding with those ('measurable' and indicated as essential) in NetScoring [11].

The scoring system we propose traces its origins from suggestions made by Abbott [25], who advocates the following main categories: 'contents', 'authorship' (expanded by us into 'accountability' including also aspects linked to possible conflicts of interest due to the presence of commercial sponsors, or non-profit nature), and 'page aesthetics' (expanded by us into 'usability' including also ease of use factors).

A simple scale was adopted for the awarding of points to each category: from dichotomic values (yes/no) up to a maximum of five values, like the well-standardized Lickert scale [26].

As the group of criteria was limited to those considered essential, all criteria were attributed the same weight: arithmetic averages were used to give partial and final results. This method of calculation was confirmed by the Wilcoxon test and by a comparison with other weighted scoring methods (Kim e NetScoring).

Therefore, we believe that when existing sites are to be evaluated, rapidity and cheapness of the resources utilized are fundamental requisites to encourage web

quality control, especially as there is a vast number of medical information sites addressed to non-professionals in contrast to the situation for professionals for which there are just a few well-known sites.

The decision to make frequency of reference in the literature a factor associated with quality parameters, assuming that increased frequency = greater importance attributed, and thus to adopt the values recorded by Kim [18], seemed to us a further significant way in which to validate the results obtained by the application of the simple averages method we proposed. In fact, the criterion is based on what happens to the Impact Factor [27] or to the measurement of the number of external links to a page in order to determine the importance of that page [28, 29].

Of the criteria mentioned by Kim the 'accessibility criterion' [18] has not been considered: the reason is that sites for non-professionals are rarely subject to access restrictions, and in our verification of this point we found that all sites were fully accessible (no fee for access) with only one exception.

The sample we utilized to evaluate the proposed quality measurement technique was obtained through Metacrawler.

By utilizing a multiple search engine the potential for finding information is much greater and reduces the time spent manually visiting a number of search engines and re-keying the query or in following dead or duplicate links [30].

5. Conclusions and further research

Internet technology and use are evolving rapidly and outpacing our ability to create standards and guidelines. For this reason, creating web quality evaluation criteria and certification procedures has to be an evolving process.

We support the current research on web quality and hope that it will quickly lead to a widespread agreement on a set of standards for information providers. We suggest that these standards be comprehensible to those parents accessing medical web sites who have a good level of education. In this way, they are personally able to directly evaluate and choose among web pages. Furthermore, a sort of 'seal of quality' for information sites should be introduced to the general public to ensure the high quality of the sites.

The increasing number of parents searching the Web for health information makes it necessary to involve them in the quality control process of the provided contents. Thus by doing further research, we are now testing the parents' ability to associate a quality score to a pediatrics information site. This test is based on the use of a web application to fill quality evaluation forms about five pediatric web sites.

However the web is a truly international press and is essentially unregulated. Anyone can publish on the web, a positive factor in a democratic society, but potentially problematic in professions such as medicine. However, both health professionals and health consumers must become critical users of the web. We have seen that the ability of a professional allows him/her to gain a better evaluation compared to a novice, who is the weakest link in the information dissemination process.

We know the Internet is a rapidly expanding universe, and trying to stop it would be in vain. We choose to spend time in making it better!

Acknowledgements

Conflicts of interest

The authors do not have any personal financial interest related to the subject matters discussed in the manuscript.

References

1. LA PORTE, R. E., 1994, Global public health and the information superhighway. *British Medical Journal*, **308,** 1651–1652.
2. KASSIRER, J. P., 1995, The next transformation in the delivery of health care. *New England Journal of Medicine*, **332**(1), 52–54.
3. FIKAR, C. R., 1996, The Internet and the pediatrician: should there be a connection? *Clinical Pediatrics*, **35**(5), 229–235.
4. IZENBERG, N. and LIEBERMAN, D. A., 1998, The Web, communication trends, and children's health. *Clinical Pediatrics*, **37,** 153–158.
5. EYSENBACH, G. and DIEPGEN, T. L., 1998, Towards quality management of medical information on the Internet: evaluation, labelling, and filtering of information. *British Medical Journal*, **317,** 1496–1502.
6. IMPICCIATORE, P., PANDOLFINI, C., CASELLA, N. and BONATI, M., 1997, Reliability of health information for the public on the World Wide Web: systematic survey of advice on managing fever in children at home. *British Medical Journal*, **314,** 1875–1879.
7. MCLUNG, H. J., MURRAY, R. D. and HEITLINGER, L. A., 1998, The Internet as a source for current patient information. *Pediatrics*, **101**(6), URL: http://www.pediatrics.org/cgi/egi/content/full/101/6/e2
8. PANDOLFINI, C., IMPICCIATORE, P. and BONATI, M., 2000, Parents on the web: risks for quality management of cough in children. *Pediatrics*, **105**(1), URL: http://www.pediatrics.org/cgi/egi/content/full/105/1/e1
9. CURRÒ, V., MAURO, V., BUONUOMO, P. S., MASTROIACOVO, P., BERNABEI, A. and D'ATRI, A., 2000, Accessing information on medical/health web sites by Italian families: a pilot study. *Clinical Pediatrics*, **39**(1), 68–69. [Medline]
10. C-H-I-Q. Guidelines for producing and maintaining patient information on conditions and treatments. URL http://www.hfht.org/chiq/guidelines.htm [last access 2002 Feb 28]
11. NetScoring: criteria to assess the quality of Health Internet information. URL: http://www.chu-rouen.fr/dsii/publi/netscoring.html [last access 2002 Feb 28]
12. HEALTH INFORMATION TECHNOLOGY INSTITUTE, MITRETEK SYSTEMS. Criteria for assessing the quality of health information on the Internet; URL: http://hitiweb.mitretek.org/docs/criteria.html [last access 2002 Feb 11]
13. DUMAN, M. and FARRELL, C., 2000, *The POPPi Guide. Practicalities of producing patient information* (King's Fund).
14. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. URL: http://www.discern.org.uk/journalepidemiology.htm [last access 2002 Feb 28]
15. QUICK. The Quality Information Checklist. URL: http://www.quick.org.uk/menu.htm [last access 2002 Feb 28]
16. RISK, A. and DZENOWAGIS, J., 2001, Review of Internet Health Information. *Quality Initiatives*; URL: http://www.jmir.org/2001/4/e28/index.htm
17. COMMITTEE ON INFECTIOUS DISEASES. *AAP. Group A Streptococcal Infections, 1997 Red Book: Report of the Committee on Infectious Diseases*, 24th ed.: 483–494.
18. KIM, P., ENG, T. R., DEERING, M. J. and MAXFIELD, A., 1999, Published criteria for evaluating health related web sites: a review. *British Medical Journal*, **318,** 647–649.
19. Centro di Ricerca sui Sistemi Informativi LUISS ''Guido Carli'' – Material for this paper URL: http://cersi.luiss.it/ped/meth/
20. STATA CORP. (1999): STATA Statistical Software: release 6.0 (College Station, TX: Stata Corporation).
21. TERRY, N. P., 2000, Rating the ''Raters'': Legal Exposure of Trustmark Authorities in the Context of Consumer Health Informatics. URL: http://www.jmir.org/2000/3/e18/index.htm
22. HEALTH ON THE NET FOUNDATION. HON Code of Conduct (HONcode) for medical and health Web sites. URL: http://www.hon.ch/Survey/analysis.html [last access 2002 Feb 28]
23. OMNI: The UK's Gateway to High Quality Internet Resources in Health and Medicine. URL: http://omni.ac.uk [last access 2002 Feb 28]

24. SILBERG, W. M., LUNDBERG, G. D. and MUSACCHIO, R. A., 1997, Assessing, controlling and assuring the quality of medical information on the Internet. *JAMA*, **277**(15), 1244–1245 [Medline]
25. ABBOTT, V. P., 2000, Web page quality: can we measure it and what do we find? A report of explanatory findings. *Journal of Public Health Medicine*, **22**(2), 191–197. [Medline]
26. LIKERT, RENSIS, 1932, A Technique for the Measurement of Attitudes, *Archives of Psychology,* **140** (June 1932), 5–53.
27. ISI Web of Knowledge, URL: www.isinet.com [last access 2002 Dec 5]
28. HERNÁNDEZ-BORGES, A. A., MACÍAS-CERVI, P., GASPAR-GUARDADO, M. A., TORRES-ÁLVAREZ DE ARCAYA, M. L., RUIZ-RABAZA, A. and JIMÉNEZ-SOSA, A., 1998, Can Examination of WWW Usage Statistics and other Indirect Quality Indicators Help to Distinguish the Relative Quality of Medical websites? URL: http://www.jmir.org/1999/1/e1/index.htm
29. CUI LEI, 1999, Rating Health Web sites using the principles of Citation Analysis: A Bibliometric Approach, URL: http://www.jmir.org/1999/1/e4/index.htm
30. KAMEL BOULOS, M. N., ROUDSARI, A. V., GORDON, C. and MUIR GRAY, J. A., 2001, The use of quality benchmarking in assessing web resources for the Dermatology Virtual Branch Library of the National electronic Library for Health (NeLH). *Journal of Medical Internet Research*, **3**(1), e5 URL: http://www.jmir.org/2001/1/e5/index.htm