Medical Informatics Europe 2002
G Surján et al. (eds)
IOS Press 2002

A Preliminary Study to Assess a French Code of Ethics for Health Teaching Resources on the Internet

SJ DARMONI (a), F. LE DUFF (b), M. JOUBERT (c),
P. LE BEUX (b), M. FIESCHI (c), J. WEBER (a), J. BENICHOU (a)

(a) Medical School, Rouen University, France
(b) Laboratory of Medical Informatics, Rennes Medical School, France
(c) LERTIM, Marseille Medical School, France

Abstract.
Background: Constant assessment of the quality of health information on the Internet is an absolute necessity as peer review is often lacking in this media. *Objective:* To develop a simple and easy French Code of Ethics, which will enable medical students to judge quality of health information in teaching material available on the Internet. *Design:* Three medical informaticians selected ten criteria from previously established codes of ethics from Europe and the USA. This instrument was tested on a sample of 30 health Internet teaching resources. *Results:* For the panel of experts, chance corrected inter-observer agreement (kappa) for quality rating ranged from k=-0.19 and k=0.33, demonstrating poor agreement among the raters. *Conclusion:* If negative results of this preliminary study are confirmed by further research, this finding may detrimentally affect projects in Europe to accredit or certify Internet health resources.
Key-words: Ethics; evaluation studies; France; Internet; quality control; students

1. Introduction

Each year, the availability of health Internet tools and services has increased at a phenomenal rate. However, end-users have no accepted criteria to assess the quality health information on the Internet. Indeed, more specialized and focused access to the right information at the right moment is essential [1]. Health resources range from general keep fit questions or preventing and managing disease, to making major decisions in surgery and medicine [2]. Also, the format may vary from general data to simple text or complete audio and video. Constant assessing the quality of health information on the Internet is an absolute necessity as peer review is often lacking throughout this media as compared to scientific journals. As creating a Web site is relatively easy, health information can be launched by anyone with access.

Nonetheless, some health information sources can be considered valid per se, such as clinical guidelines from national agencies, or articles from a scientific journal, which have already been peer-reviewed. Unfortunately, the Internet also produces a large amount of advertising or the latest medical rumor, or even the most sophisticated pseudo-scientific scam. There is no other field in which inaccurate, incomplete, or biased information is potentially more damaging [3].

Over the past four years several world-wide initiatives have been undertaken to define criteria to assess the quality of the increasing amount of health information on the Internet [2-9].

More than 100 articles in the past five years have examined the quality of the health content available on the Internet, and most of these authors have used clinical guidelines as their standard reference [10-15]. Nonetheless, it is important to separate the criteria used to only assess the Web site itself and the criteria to assess the health content. In May 2000, the French Ministry of Health and the National Council of Physicians launched an initiative to define criteria to assess the quality of the scientific health "content" on the Internet. "Sensitive" information was defined as information found in documents published on the Internet, which could be used in making a medical decision (e.g. efficacy and toxicity of medical interventions). This information is contained in documents such as clinical guidelines, consensus conference reports, technical reports and teaching material. For "sensitive" information, the group recommended that the main criterion chosen should be an indication of the level of evidence for all information, particularly as regards efficacy and toxicity of healthcare interventions [16].

In June 1999, a French Medical Virtual University (FMVU) consortium [17] was created to test various tools and methods necessary to build a virtual university. Eight medical schools joined this consortium: Grenoble, Lille, Marseille, Nancy, Paris V, Paris VI, Rennes and Rouen. FMVU (URL: http://www.umvf.prd.fr) was partially granted by the Health Technologies National Network program of the French Ministry of Research. Eleven working packages (WP) were defined.

Assessment of French Code of Ethics for health teaching resources on the Internet has been included in the working package (WP8) devoted to the study of quality criteria to evaluate health information available for the student on the Internet. Firstly, three teachers in medical informatics from the FMVU consortium (Marseille, Rennes, Rouen) defined a French Code of Ethics for medical students. The goal of this study was to test the reproducibility of this code of ethics by measuring the inter-expert variability in thirty French-speaking teaching resources.

2. Material and Methods

The French Code of Ethics was defined by consensus between three medical informaticians of the FMVU consortium in charge of the working package: "Quality of health information on the Internet". This code of ethics was structured as simple as possible to facilitate its reproducibility and its accessibility by its future users (students & teachers).

To avoid the "reinventing the wheel syndrome", the French Ethical Code was primarily based on previous codes of ethics [2-9, 18], in particular the most-used code (code of Health on the Net) by over than 2,500 Web health sites in the world as well as three French codes: Net Scoring [7], French Ministry of Health & Marseille [18].

Following analysis of the literature and semantic analysis of the employed vocabulary, ten criteria were selected according to their frequency of use in the previous codes. These criteria are listed in Table 1: four regarding the source of the information, three on the content and three on man-machine interface.

To study the reproducibility of this code of ethics, 30 teaching resources were randomly selected from the 1,400 available in Doc'CISMeF [19] one of the major search tools chosen by the FMVU consortium (WP4). The concept of resources includes not only Web sites but also documents retrievable from these Web sites. Therefore, among the 30 studied resources, 15 were Web sites, e.g. the teaching Web site from the Geneva Medical School (URL: http://edumed.unige.ch/) and 15 resources were teaching documents, such as "Blood

pressure and its clinical measurement" (Marseille Medical School) (URL: http://medidacte.timone.univ-mrs.fr/learnet/webcours/hta/pressionart/index.htm).

To evaluate each criterion the three experts used a four point Likert scale (very good, good, bad and very bad). At least one of these occurrences was mandatory for each rater. The occurrence "average" was excluded because pair number of occurrences would oblige the evaluator to make a decision between "good" and "bad".

Prior to formal evaluation, the three raters agreed on the score to be assigned to each criterion with four possible occurrences of the Likert scale. Due to the structure of the study (in particular the source of the information), the evaluators were not blind to the 30 teaching resources evaluated.

Inter-rater agreement was assessed by calculating chance-corrected unweighted Cohen's kappa coefficient and its standard error for each pair of raters and for each of the 10 criteria. All 30 resources (i.e., 15 sites and 15 documents) were used for this analysis. Landis and Koch's classifications [20] were used to assess strength of agreement (i.e., poor for kappa value less than 0.00, slight for 0.00-0.20, fair for 0.21-0.40, moderate for 0.41-0.60, substantial for 0.61-0.80, almost perfect for 0.81+). Moreover, calculated kappa coefficients were compared to 0 using an exact test in order to assess the existence of an inter-observer agreement greater than what chance alone would have predicted under the assumption of independence between raters. As 30 tests were performed (10 criteria, 3 pairs of raters), Bonferroni's adjustment was used and only P values less than 0.0017 were considered as significant.

For each rater and resource, a global score was obtained by summing rater's gradings (from 1 to 4) over all 10 criteria. Mean values of the three raters over all 30 resources were compared using one-way analysis of variance.

These analyses were repeated separately for the 15 sites and 15 documents. Software StatXact-3 (Cytel Software Corporation) and SPSS (SPSS v10 for Windows) were used.

3. Results

Slight agreement was observed among raters (see Table 1). Out of the 30 calculated kappa values, 5 showed poor agreement, 17 slight agreement and 8 fair agreement. The highest value was 0.335 (standard error = 0.138) for the navigability criterion (raters 2 and 3). No kappa value was significantly greater than 0. One kappa value was significantly lower than 0 showing weaker agreement than what chance alone would have predicted for the "source of financing" criterion (kappa = - 0.185, standard error = 0.086, raters 1 and 3). Similarly, very mediocre agreement was also observed among raters when separately looking at sites and documents (data not shown).

Overall, means of global scores significantly differed between the three raters ($p = 0.043$). No significant differences were observed in separate analyses of sites ($p = 0.149$) and documents ($p = 0.241$), possibly due to lack of power.

4. Discussion

As health information has both the potential not only to improve health but also to be detrimental, organizations and individuals that provide health information on the Internet have the basic obligation to be trustworthy, provide high quality content, protect users' privacy, and adhere to standards of best practices in health care [2]. This topic has previously been exhaustively studied particularly over the past 4 years: In fact, we found 107 articles when searching the following PubMed request 'Internet [MeSH] and quality control [MeSH], (2001-10-30).

The limited results of this study which test the reproducibility of Code of Ethics concerning quality of health information on the Internet requires further research. To our knowledge, it is the second study reported in the literature after the Discern study in 1999 [5]. However, the aim of the Discern study was specifically consumer health information on therapeutics. Moreover, the evaluation focused on published resources and not the Internet, which would explain the different results of these two studies. Discern tool has 15 items: eight general criteria about quality information, very similar to our Code of Ethics and seven criteria concerning therapeutics. A five point Likert scale was used. For the overall quality, kappa was k=0.53 among the expert panel, k=0.40 among information providers and k=0.23 among self help group members. According to Discern cut off point for an acceptable level of agreement at kappa > 0.40, four of eight Discern general items were not classified as acceptable among expert (5 among information providers and 6 among self help group members): explicit aims, aims achieved, source of information and currency of information.

Several elements of our study could explain our limited results, in particular a lack of agreement between raters to more precisely discuss about each possible occurrence of every criterion. Some of these criteria are more subjective: for example, pertinence of hyperlinks vs. existence of a review process. Furthermore, it is easier to evaluate a document than a Web site, which may contain several hundred or thousand pages. In this case, the evaluation is only performed on a sample which is not standardized for each expert.

If these limited results are also confirmed by other studies, they should have important consequences on current approaches in Europe [6] to study accreditation and certification of e-health Web sites by third parties.

Furthermore, a code of ethics to evaluate quality of health information on the Internet must not only deal with the Web site quality but with the scientific health content. Nonetheless, several studies have tested the hypothesis that a set of criteria to assess Web site quality may also be an indirect indicator of the overall quality content of the health information [2, 7]. Moreover, studies show that the correlation between quality of the site and quality of the health content still remains controversial in 2001. Sandvik [11] did not find any significant correlation between quality of the content defined by clinical guidelines and the Web Impact Factor (WIF) (number of external links to a site) in a particular medical domain: female incontinency. The studies of Pandolfini and coll. [12] and Impicciatore and coll. [12] similarly did not find any significant correlation between quality of the content and technical quality of the site itself (respectively in child cough and child fever). In contrast, Hernández-Borges and coll. [14] found a significant correlation between quality of content and technical quality in pediatric sites. With an increasing number of people accessing an growing amount of health information on the Internet, publishers of this information have a major ethical obligation to help their readers (health professionals but more so, Netizens) to find only high quality documents.

Conclusion: Reproducibility of the French Code of Ethics chosen by the FMVU consortium was not proven by this first study. Our research team is conducting ongoing studies focusing on documents and not entire Web sites. Intra-expert variability will be also shortly tested on the same 15 teaching documents which will be saved to avoid any major problems when evaluating quality of health information on the Internet: time dependency.

Table 1: List of Criteria of the French Code of Ethics chosen by the FMVU consortium

Criteria	Expert 1 vs. Expert 2	Expert 1 vs. Expert 3	Expert 2 vs. Expert 3
Source			
Name. logo and references of the institution. and name and title of author on each document of the site	0.196 (0.175)	0.266 (0.171)	0.177 (0.122)
Source of financing. independence of the author(s). Conflict of interest. Influence. bias	-0.058 (0.096)	- 0.185 (0.086)	0.105 (0.173)
Editorial review process. existence of a webmaster	0.262 (0.093)	0.114 (0.079)	0.167 (0.111)
Original Source Stated	0.120 (0.096)	0.143 (0.100)	0.152 (0.101)
Content			
Target of the Web site	0.010 (0.103)	0.118 (0.131)	0.081 (0.117)
Updating: actualization of the site including date of creation. date of last update et eventually date of last version	0.230 (0.102)	0.294 (0.104)	0.068 (0.082)
Pertinence of hyper-links	-0.036 (0.062)	-0.028 (0.072)	0.208 (0.189)
Interface			
Logic organization (navigability): search engine. general index. "what's new" page. help page. map of the site	0.009 (0.092)	-0.001 (0.083)	0.335 (0.138)
Design of the Web site	0.160 (0.127)	0.159 (0.097)	0.062 (0.110)
Mechanism for feedback: Email of the author on every document	0.232 (0.118)	0.231 (0.128)	0.064 (0.093)

References

[1] Brunetaud JM, Darmoni SJ, Souf N, Dufresne E, Beuscart R. A Resource Server For Medical Teaching. Methods Inf Med 2002; 41(2):177-182.

[2] Ambre J, Guard R, Perveiler FM, Renner J, Rippen H. Health Information Technology Institute. Working Draft White Paper: Criteria for Assessing the Quality of Health Information on the Internet. Available from: URL: http://hitiweb.mitretek.org/hswg/ & http://hitiweb.mitretek.org/docs/policy.pdf

[3] Code of ethics of the Internet Healthcare Coalition. Available from: URL: http://www.ihealthcoalition.org/ethics/ethics.html

[4] Boyer C, Selby M, Scherrer JR, Appel RD. The Health On the Net Code of Conduct for medical and health Websites. Comput Biol Med. 1998 Sep;28(5):603-10.

[5] Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb;53(2):105-11.

[6] Eysenbach G, Diepgen T, Lampe K, Brickley D. EU-project medCERTAIN: Certification and Rating of Trustworthy and Assessed Health Information on the Net. Stud Health Technol Inform. 2000;77:279-83.

[7] Darmoni SJ, Leroux V, Daigne M, Thirion B, Santamaria P, Duvaux C. Critères de qualité de l'information de santé sur l'Internet. In: Albert A, Roger-France FH, Degoulet D, Fieschi M, editors. Santé et Réseaux Informatiques. Informatique et Santé. Paris: Springer Verlag France; April 1998. p. 162-74.

[8] Winker MA, Flanagin A, Chi-Lum B, White J, Andrews K, Kennett RL, DeAngelis CD, Musacchio RA.Guidelines for Medical and Health Information Sites on the Internet. Principles Governing AMA Web Sites. JAMA. 2000 Mar 22-29;283(12):1600-6. Available from: URL: http://jama.ama-assn.org/issues/v283n12/pdf/jsc00054.pdf

[9] Tatsumi H, Mitani H, Haruki Y, Ogushi Y. Internet medical usage in Japan: Current situation and issues. J Med Internet Res 2001;3(1):e12 Available from: URL: http://www.jmir.org/2001/1/e12/

[10] Sandvik H. Health information and interaction on the internet: a survey of female urinary incontinence. BMJ 1999 Jul 3; 319(7201):29-32.

[11] Impicciatore P, Pandolfini C, Casella N, Bonat M. Reliability of health information for the public on the world wide web: systematic survey of advice on managing fever in children at home. BMJ 1997 Jun 28;314(7098):1875-8.

[12] Pandolfini C, Impicciatore P, Bonati M Parents on the web: risks for quality management of cough in children. Pediatrics. 2000 Jan;105(1):e1.

[13] Hernández-Borges AA, Macías-Cervi P, Gaspar-Guardado MA, Torres-Álvarez de Arcaya ML, Ruiz-Rabaza A, Jiménez-Sosa A. Can Examination of WWW Usage Statistics and other Indirect Quality Indicators Distinguish the Relative Quality of Medical Web Sites? J Med Internet Res 1999;1(1):e5. Available from: URL: http://www.jmir.org/1999/1/e1/index.htm

[14] McClung HJ,. Murray RD, Heitlinger LA. The Internet as a source for Current Patient Information. Pediatrics. 1998 Jun;101(6):e2 Available from: URL: http://www.pediatrics.org/cgi/content/full/101/6/e2

[15] Griffiths KM, Christensen H. Quality of web based information on treatment of depression: cross sectional survey. BMJ 2000 Dec 16;321(7275):1511-5.

[16] Darmoni SJ, Haugh M, Lukacs B, Boissel JP. Quality of health information about depression on internet. Level of evidence should be gold standard. BMJ 2001Jun 2;322(7298):1367.

[17] LeBeux P, LeDuff F, Fresnel A, Berland Y, Beuscart R, Burgun A, et al. The French Virtual Medical University. Stud Health Technol Inform. 2000;77:554-62.

[18] Joubert M, Aymard S, Fieschi D, Fieschi M. Quality criteria and access characteristics of Web sites: proposal for the design of a health Internet directory. Proc AMIA Symp. 1999: 824-8.

[19] Darmoni SJ, Thirion B, Leroy JP, Douyère M, Lacoste B, Godard G et al. A search tool based on 'encapsulated' MeSH thesaurus to retrieve quality health resources on the Internet. Med Inform Internet Med 2001;26(3):165-78.

[20] Landis, JR, Koch, GG. The measurement of observer agreement for categorical data. Biometrics 1977:33;159-174.

[21] Shon J, Marshall J, Musen MA. The impact of displayed awards on the credibility and retention of Web site information. Proc AMIA Symp 2000;794-8.