

Assessment of A French Code of Ethics for Health Teaching Resources on the Internet

F. Le Duff^d, SJ Darmoni², J-C Dufour³, M. Joubert³,
M. Fieschi³, J. Benichou², P. Le Beux¹

¹ *Laboratory of Medical Informatics,
Medical School 2 rue Henri le Guilloux 35000 Rennes, France*
² *Medical School, Rouen University, France*
³ *LERTIM, Marseille Medical School, France*

Franck.Leduff@univ-rennes1.fr

Summary

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 scientists selected nine criteria from previously established codes of ethics from Europe and the USA. This instrument was tested on a sample of 24 health French-speaking Internet teaching resources. Results: For the panel of experts, we analyzed assessments with non parametric tests. Expert assessment was analyzed with non-parametric tests. This analysis demonstrated a strong agreement among the raters. Conclusion: It seems possible to produce an analysis summary to evaluate teaching material available on the Internet.

INTRODUCTION

Each year, the availability of health Internet tools and services increases at a phenomenal rate. However healthcare professionals and consumers, do not have means to easily assess the quality of health information on the Internet. Nevertheless, more specialized and focused access to the right information at the right moment is essential¹. Health resources range from general keep fit questions or preventing and managing disease, to making major decisions in surgery and medicine². Also, the format of documents may vary from general data to simple text or complete audio and video. Constant assessment of 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 and uncontrolled, health information can be launched by anyone.

Nonetheless, some health information sources can be considered valid, 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 damaging³.

Over the past years several world-wide initiatives have been undertaken to define criteria to assess the quality of the increasing amount of health information on the Internet²⁻⁹.

A lot of articles have studied the quality of the health content available on the Internet, and most of the authors have used clinical guidelines as their standard reference¹⁰⁻¹⁵. Nonetheless, it is important to separate the criteria used to only assess a Web site itself and the criteria to assess its health content. In June 1999, a French Medical Virtual University (FMVU) consortium¹⁷ was created to test various tools and methods necessary to build a medical 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. 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 medical decision making (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 interventions¹⁶.

Assessment of a French Code of Ethics for health teaching resources on the Internet has been included in the working package devoted to quality criteria of health information available for students on the Internet. Three teachers in medical informatics from the FMVU consortium (Marseille, Rennes, and Rouen)

proposed a French Code of Ethics for medical students. The goal of this study was to test the reproducibility of this code of ethics in measuring the inter-expert variability in twenty four French-speaking teaching resources, located in France, Belgium, Switzerland, and Canada.

This paper details the criteria involved in the study and the statistical evaluation performed after the rating process operated by the judges.

MATERIAL AND METHODS

The French Code of Ethics was defined after a consensus has been reached by three medical scientists of the FMVU consortium in charge of the working package: "Quality of health information on the Internet". This code of ethics was structured as simply as possible to facilitate its reproducibility and its accessibility by its future users (students and 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 2,500 Web health sites in the world as well as three French codes: Net Scoring⁷, French Ministry of Health, and Marseille¹⁸. Following analysis of the literature and semantic analysis of the employed vocabulary, nine criteria were selected according to their frequency of use in the previous codes. These criteria describe nine characteristics of the medical contents for pedagogical use : three regarding the source of the information, three on the content and three on man-machine interface. These criteria are:

- Source: title, author name, logo of the organization
- Editorial board: list and links in the document, available in home page
- Target: intended students
- Update: date of creation and/or update on the document
- Feedback: e-mail address(es) of the author(s)
- References: available with links and correct format
- Links: external and internal links available
- Navigability: table of content, help
- Aesthetics and design: very subjective criterion

To study the reproducibility of this code of ethics, 24 teaching resources were selected from the 1,500 available in Doc'CISMeF¹⁹ one of the major search tools chosen by the FMVU consortium. One document has been randomly selected from each medical school web site. For each rater and resource, a global score was obtained by summing rater's grading (from 1 to 4) over all nine criteria. Therefore, among the 24 studied resources, some of them were teaching documents in HTML format such as "Blood pressure and its clinical measurement"^A and others were in PDF format such as

^A Marseille Medical School : <http://medidacte.timone.univ-mrs.fr/learnet/webcours/hta/pressionart/index.htm>

"Sexuality and Pharmacology"^B. All the documents were intended for initial training.

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 even number of occurrences would oblige the evaluators to take a stand between "good" and "bad" evaluation. Prior to formal evaluation, the three experts 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 24 teaching resources evaluated.

There are several ways to analyze the agreement of judges rating the threads from the Internet: kappa, gamma, and Kendall's W²¹. Perhaps the most familiar to medical researchers and practitioners is Cohen's kappa. However, this statistic assumes the data is nominal in measurement. The data we have is ordinal and so Cohen's kappa, although familiar and often used, was inappropriate for this study. Rather than comparing absolute scores given by the judges, we preferred to test the difference between scores attributed by experts according to the given criteria. Inter-rater agreement was assessed by calculating a two related samples test. The Wilcoxon's non parametric test for pair data was used to assess differences between raters. In order to have a global view of the scoring made by all the raters, we applied on the differences of scores a variance analysis the advantage of which is to preserve the notion of what is tested. This means that the statistical test allows to measure the difference of scoring for each studied document. The statistical work has been done with SPSS (SPSS v10 for Windows).

RESULTS

As an illustration of the scoring process are listed here the scores of two documents evaluated by the judges. The first is one for which scores were very similar^C. It is a document located on the web site of the University of Rennes, France. It is a PDF file scored respectively 20, 20 and 19 by the experts on a scale going from 9 to 36. The second is one for which scores were very different^D. Its respective scores are 17, 22, and 27. In this case it is a University site in Laval, Canada. These two examples show that raters may have a different way to judge a document and a complete site in locations where teaching rules are not the same.

Nevertheless, strong agreement was observed among raters. The Wilcoxon test does not show significant differences between the scores attributed by paired

^B Rennes Medical School : <http://www.med.univ-rennes1.fr/etud/pharmaco/pdf/sexualite.PDF>

^C <http://www.med.univ-rennes1.fr/etud/pharmaco/pdf/sexualite.PDF>

^D <http://www.fmed.ulaval.ca/med-17105/index.htm>

raters. Probabilities are respectively: 0.667, 0.767, and 0.353.

A variance analysis on repeated measures evaluates the effect between judges regarding each document. This test reveals no significant difference between them ($p=0.589$). A variance analysis on repeated measures taking into account the interaction between judges and criteria shows: 1) there is a significant difference between criteria ($p<0.001$), and 2) there is no significant difference when the effect between judges is considered ($p=0.784$)

From this study we can conclude that:

- The used criteria have not the same impact in the scoring process (have judges the same understanding of each of them?)
- Pair wise, judges do not score each document differently in a significant way
- Globally, taking into account both judges and documents, there is no significant statistical difference.

This entails us to conclude that the few “common sense” criteria proposed to form a grid for medical training web resources evaluation may be used successfully in the framework of our project of FMVU.

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 care².

This topic has previously been exhaustively studied : we found more than 100 articles when searching the following PubMed request 'Internet [MeSH] and quality control [MeSH]' over the past four years, (2001-10-30).

The results of this current study which tested the reproducibility of a French Code Ethics concerning quality of health teaching information on the Internet is highly encouraging. To our knowledge, it is the second study reported in the literature after the Discern study in 1999⁵. However, the aim of the Discern study was specifically consumer health information on therapeutics. Moreover, the evaluation focused on published resources and not only the Internet, which could explain the different results. 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. 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 results. The main point is the time passed to obtain an agreement grid analysis between raters to more precisely discuss about each possible occurrence of every criterion. Some of these criteria are more subjective: for example aesthetic but more of them are objective and it was possible to define each grade of the Likert's scale.

Furthermore, a code of ethics to evaluate quality of health information on the Internet must not only deal with Web site quality but with scientific health content. Nonetheless, several studies have tested the hypothesis that a set of criteria to assess Web sites quality may also be an indirect indicator of the overall quality content of the health information^{2, 7}. Moreover, studies showed that the correlation between quality of the site and quality of the health content still remains controversial in 2001. Sandvik¹⁰ 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.¹¹ 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). A recent paper²¹ showed poor agreement between five medical doctors which rated the information contained in newsgroup threads using a 6-point scale. Several tests were used: Cohen's kappa, gamma, Kendall's W, and Cronbach's alpha and the paper pointed out the difficulties to select a good statistics method to measure the agreement of judges. In contrast, Hernández-Borges and coll.^{13,14} found a significant correlation between quality of content and technical quality in pediatric sites. With an increasing number of people accessing and a growing amount of health information on the Internet, publishers of information have a major ethical obligation to help their readers (health professionals but more so, Netizens) to identify high quality documents.

One of the main interest of this current work is also the creation of a quick assessment grid easy to use. In contrast with other ethics codes of evaluation, this one is simple to use and could explain the results and the strong agreement between experts. Moreover, this code of ethics could be useful for evaluating other medical contents such as entire website for pedagogical use or in general for all content for medical use. In particular, it should be interesting to evaluate this French code of ethics for health content intended for patients. In the context of the consumer health information, this code should allow to produce quality tutorials or simulations.

CONCLUSION

Reproducibility of the French Code of Ethics chosen by the FMVU¹⁷ consortium was improved by this study. This work is highly interesting for the FMVU Consortium because it should allow to produce a charter for medical and pedagogic resources. These

contents will be published on the medical virtual campus with the approval of the raters using the same analysis in different Medical School.

ACKNOWLEDGMENTS

French Medical Virtual University (URL: <http://www.umvf.prd.fr>) was partially granted by the Health Technology National Network program of the French Ministry of Research. The authors thank Dr Roch Giorgi, Marseille University, who kindly contributed to the statistical analyses. The authors thank Richard Medeiros, Rouen University Hospital Editor, for his valuable advice in editing the manuscript.

REFERENCES

1. Brunetaud JM, Darmoni SJ, Souf N, Dufresne E, Beuscart R. A Resource Server For Medical Teaching. *Methods Inf Med* 2001 (in press).
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.mitrotek.org/hswg/> & <http://hitiweb.mitrotek.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, Flanagan 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. Pandolfini C, Impicciatore P, Bonati M. Parents on the web: risks for quality management of cough in children. *Pediatrics.* 2000 Jan;105(1):e1.
12. 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.
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* 2001 Jun 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. 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.
21. Mark Craigie; Brian Loader; Roger Burrows; Steven Muncer. Reliability of Health Information on the Internet: An Examination of Experts' Ratings *Journal of Medical Internet Research* 2002;4(1):e2.