Level of Evidence as a Future Gold Standard for the Content Quality of Health Resources on the Internet

A Preliminary Study

S.J. Darmoni1, E. Amsallem2, M. Haugh2, B. Lukacs3, V. Leroux4, B. Thirion5, J. Weber6, J.P. Boissel2

1Computer and Networks Department, 2Centre Cochrane Français, Lyon, France, 3Tenon Hospital, Assistance Publique de Paris, 4Centrale Santé, Paris, France

Summary
Objective: An assessment of the quality of health information on the Internet is an absolute necessity. In this study, ‘sensitive’ information was defined as information found in documents published on the Internet, which could be used in a medical decision. For sensitive information, the main criterion chosen for the quality of the information was an indication of the level of evidence. A survey was conducted using the CISMeF health catalogue to assess how often a score of the level of evidence is mentioned in the information accessible on the Internet in French-language health resources.

Methods: Since 1999, members of the CISMeF team have systematically been searching for all documents containing ‘sensitive’ information and verifying whether the level of evidence was explicitly indicated as a score at least once in the document.

Results: As of June 2001, 10,190 resources were included in CISMeF, including 2964 textual ‘sensitive’ resources (29.1%). Out of all these resources, only 4.7% (95% confidence interval: 4.0 - 5.5%) indicated the level of evidence. A statistically significant difference in the prevalence of indicating the level of evidence according to resource types (e.g., 18.1% for guidelines compared to 0.0% for teaching material), year of publication (almost three times greater in 1997-2001 compared to 1990-1996) and publishers was observed.

Conclusion: As the number of people accessing the growing amount of information on the Internet is increasing daily, publishers have an ethical obligation to inform their readers about the validity of ‘sensitive’ information their sites contain. However, the vast majority of the French language Internet resources that were surveyed do not mention a score of the level of evidence for their sensitive information.

Keywords
Internet, quality control, evidence-based medicine, level of evidence

Introduction

Internet technology, tools and services currently provide access to an enormous volume and broad spectrum of scientific and health information. Health Internet sites include information for staying fit, preventing and managing disease, and making other decisions related to health care (1). They also include information for making decisions about health products and health services. These sites vary in their format, from general data to simple text to audio or to video.

Rating the quality of health information on the Internet is an absolute necessity because unlike scientific journals, this media is lacking peer review. As creating a Web site is relatively easy (any Netizen is a potential Gutemberg), health information can potentially be accessed by more than 450 million people. 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 (see Appendix (2-7)). Over 100 articles have described the quality of the health content available on the Internet, and most of these articles have used clinical guidelines as their reference standard (8-13).

In May 2000, the French Ministry of Health and the National Council of Physicians launched an initiative to define a French code of Ethics for e-health sites devoted to lay people. To examine the different aspects of the problem four working groups were created. These groups brought together three categories of participants: (a) public and private providers of e-health information, (b) end-users associations and 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 since anyone with Internet access can read this information. It can be difficult for users (health professional or lay people) to determine which part of this information is usable and credible; how it can be critically appraised or verified. Therefore, Internet health sites consumers should have a careful and critical attitude towards health information available on the Internet. Netizens could play a useful role participating in quality feedback. The difference between Internet and other media is that, in 2002, it can potentially be accessed by more than 450 million people. Over the past few 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 (see Appendix (2-7)). Over 100 articles have described the quality of the health content available on the Internet, and most of these articles have used clinical guidelines as their reference standard (8-13).

In May 2000, the French Ministry of Health and the National Council of Physicians launched an initiative to define a French code of Ethics for e-health sites devoted to lay people. To examine the different aspects of the problem four working groups were created. These groups brought together three categories of participants: (a) public and private providers of e-health information, (b) end-users associations and...
(c) third-parties in charge of the e-health information assessment. The goal of the respective three other working groups were (1) to identify potential risks of e-health information for lay people; (2) to define a French code of ethics and transparency for e-health Web sites and services; (3) to create an organization in charge of the management of the overall national initiative and to cooperate with other European and US initiatives.

The goal of our working group was to define criteria to assess the quality of health «content» on the Internet which is clearly different from the quality of the site itself. This group decided that it was imperative to distinguish between information on the sites which is ‘sensitive’, e.g., information concerning the efficacy or toxicity of healthcare interventions, and that which is ‘non-sensitive’, such as addresses for doctor or patient associations. We defined ‘sensitive’ information as information found in documents published on the Internet, which could be used in a medical decision, in particular evidence-based medicine resources. This information is contained in documents such as clinical guidelines, consensus conference reports, technical reports and teaching material.

For sensitive information, the group suggested that the main criterion chosen should be an indication of the level of evidence for all information, particularly information on efficacy and toxicity of healthcare interventions (14). It is obvious that this proposal is not applicable to all health resources on the Internet. For example, patients’ association web sites will not need to indicate the level of evidence for information presented on their sites, if this information will not be used in a medical decision.

To explore to what extent existing Internet resources already comply with this proposal, we conducted a survey using CISM eF to assess how often an indication of the level of evidence has been systematically indicated in French language Internet health resources. CISM eF (15) is a Web site created in February 1995 at the Rouen University Hospital, in France to build a catalogue of French language health resources on the Internet (http://www.chu-rouen.fr/cismef or http://www.cismef.org).

Material and methods

CISM eF uses a rigorous four-step process to build the catalogue: resource collection, selection, description and indexing. One deputy medical librarian performs the resource collection and the information surveillance. Health resources are selected using the Net Scoring initiative main criteria (source of information, disclosure, editorial review process, date of last update, and mechanism for feedback) for inclusion in CISM eF. The selection process should be mandatory to create a quality health Internet directory (16). Currently, Net Scoring is used by many prominent French institutions (17). Resources that do not respect basic, particularly ethical, criteria are not included in the CISM eF database. Two deputy medical librarians describe and index the resources. The chief medical librarian is a ‘super-indexer’ in charge of verifying the indexing. The last two stages are based on a similar procedure used by the US National Library of Medicine to build and update MEDLINE.

CISM eF uses two standard tools for organizing information: the MeSH (Medical Subject Headings) thesaurus from the US National Library of Medicine and the Dublin Core metadata format (18) which is included in the IEEE 1484 Learning Objects Metadata standard developed by IEEE Learning Technology Standards Committee (URL: http://www.manta.ieee.org/groups/ltscc/). To describe resources, the CISM eF team use the following 10 elements, out of the 15 from the Dublin core project (URL: http://purl.org/DC/elements_set.html): title, author or creator, subject and keywords, description, publisher, date, resource type, format, identifier, and language (19). The following fields are specific to CISM eF: institution, city, province or state, country, target, cost of access, type of sponsorship. Each resource included in CISM eF is indexed with MeSH terms (keywords and qualifiers) which are semantically linked to medical specialties (i.e., the MeSH keyword ‘neoplasms’ or the qualifier ‘secondary’ are linked to the medical specialty ‘medical oncology’ (20). In CISM eF, ‘sensitive’ documents were defined as those indexed with the following resources types: clinical guidelines as the main evidenced-based medicine resources freely available on the Internet, consensus conference reports, technical reports and teaching material. Clinical guidelines were defined as «the official statement or policies of major organizations and agencies on the proper indications for performing a procedure or treatment or the proper management for specific clinical problems» (21).

Since 1999, members of the CISM eF team have systematically been searching for all documents containing ‘sensitive’ information and verifying whether the level of evidence was explicitly indicated as a score (e.g., a rank or a pictogram) at least once in the document. This search is simple, manual and pragmatic. It first involves a search for the string ‘level of evidence’ (in French, ‘niveau de preuve’ (22-23)) in the document which may be accessed in a HTML or PDF format. All other synonyms or expressions of related concepts were also searched for (e.g., «force de preuve», «force de démonstration», «niveau de confiance», etc. with approximate equivalent terms in English evidence level, rate of evidence, score, grade, etc.). Nonetheless, there is a lack of a reference method for evaluating the level and over 20 different methods are used worldwide (23) and several in France only: in particular, the most important of these methods has been defined by the French Agency of Health Acreditation and Evaluation (A NA E S) (Table 1). Therefore, for the resources which indicate a level of evidence, CISM eF also displays the method used to evaluate it. Since there is no standard scale, we restricted our task to the identification of documents showing an indication of the level of evidence, whether it applied to the whole document, to a specific guideline, or to the data the recommendation or the supplied information are based on.

To study the false negative rate (also called silence in information science) of this method, we reviewed retrospectively each clinical guideline from the French Agency of Health Acreditation and Evaluation (A NA E S) where the level of evidence has been systematically indicated since 1997.

Methods Inf Med 1/ 2003
The prevalence (and its 95% confidence interval) of ‘sensitive’ documents indicating level of evidence was computed with the binomial exact method (24). A Pearson Chi2 test was used to compare categorical data. All tests were two-tailed and \( p < 0.05 \) was considered statistically significant. Analyses were performed using SPSS for Windows, Version 10.0 (25).

**Results**

In June 2001, 10,190 resources had been included in CISMeF; among them, 2,964 textual ‘sensitive’ resources (29.1%), with only 4.7% (95% confidence interval: 4.0-5.5%) indicating the level of evidence. The different types of textual resources were: clinical guidelines (22.5%), consensus conferences (5.1%), technical reports (14.1%), teaching material (31.7%) and others, such as comparative studies (see URL: http://www.chu-rouen.fr/documed/typeeng.html) (26.6%). The prevalence of documents which indicate the level of evidence by resource type is described in Table 2. There was a significant difference between resource types (\( \chi^2 = 382.9; df = 4; p < 0.001 \)). Moreover, significantly more guidelines indicated the level of evidence than did consensus conferences (\( \chi^2 = 4.9; df = 1; p = 0.027 \)).

Table 3 shows the prevalence of indicating the level of evidence according to the year of publication of the document for guidelines and consensus conferences.

The time periods we chose were: 1990-1996 and 1997-2001, because in 1996 the French government decided to standardize the use of guidelines by physicians. A statistically significant increase was observed in the number of documents indicating the level of evidence in the later period (\( \chi^2 = 26.5; df = 1; p < 0.001 \)). The top 10 medical publishers were selected (according to the number of ‘sensitive’ textual resources indexed in CISMeF).

A statistically significant difference when indicating the level of evidence was observed according to the publishers of the document (Table 4; \( \chi^2 = 138.5; df = 9; p < 0.001 \)).

Furthermore, CISMeF includes a search tool named Doc’CISMeF (26) which may limit any request to documents that explicitly mention the level of evidence. For example, a Doc’CISMeF search for “asthma” found 26 resources, 19 ‘sensitive’ documents and 3 which explicitly indicated the level of evidence. To explore the sensitivity of our search procedure we checked the 60 resources indexed in CISMeF as clinical guidelines from the French Agency of Health Accreditation and Evaluation (ANAES). Since 1997, the ANAES procedure for guidelines contains a systematic and explicit mention of the level of evidence.

However, according to the CISMeF pragmatic detection, the level of evidence was only mentioned in 32 documents (ie 53% of documents). After Email and phone interaction with the ANAES, the CISMeF method did not omit any clinical

---

### Table 1: ANAES method to evaluate the level of evidence (for therapy issues)

<table>
<thead>
<tr>
<th>Level of evidence from the literature</th>
<th>Type of studies</th>
<th>Grade of the guidelines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>Well-designed randomized controlled trials (adequate sample size)</td>
<td>Grade A: explicit scientific evidence</td>
</tr>
<tr>
<td></td>
<td>Meta-analysis of randomized controlled trials</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Decision analysis based on well-designed studies</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Randomized controlled trials with poor power (non-adequate sample size)</td>
<td></td>
</tr>
<tr>
<td>Level 2</td>
<td>Well-designed non-randomized comparative studies</td>
<td>Grade B: assumption of scientific evidence</td>
</tr>
<tr>
<td></td>
<td>Cohort studies</td>
<td></td>
</tr>
<tr>
<td>Level 3</td>
<td>Case controlled studies</td>
<td>Grade C: weak level of scientific evidence</td>
</tr>
<tr>
<td></td>
<td>Comparative studies with important bias</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Retrospective studies</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Case-report studies</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Descriptive epidemiological (transversal or longitudinal) studies</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2: Prevalence of ‘sensitive’ textual documents indicating the level evidence by resource types

<table>
<thead>
<tr>
<th>Resource types</th>
<th>Number of resources</th>
<th>Number of resources indicating the level of evidence</th>
<th>Prevalence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guidelines</td>
<td>668</td>
<td>121</td>
<td>18.1</td>
</tr>
<tr>
<td>Consensus conferences</td>
<td>150</td>
<td>16</td>
<td>10.7</td>
</tr>
<tr>
<td>Technical reports</td>
<td>417</td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>Teaching material</td>
<td>941</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>Other</td>
<td>788</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td><strong>All</strong></td>
<td>2964</td>
<td>139</td>
<td>4.7</td>
</tr>
</tbody>
</table>

### Table 3: Prevalence of guidelines and consensus conferences indicating the level of evidence by year of publication

<table>
<thead>
<tr>
<th>Years</th>
<th>Number of guidelines and conference consensus (GCC)</th>
<th>GCC with indication of level of evidence</th>
<th>Prevalence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997-2001</td>
<td>523</td>
<td>114</td>
<td>21.8</td>
</tr>
<tr>
<td>1990-1996</td>
<td>295</td>
<td>23</td>
<td>7.8</td>
</tr>
<tr>
<td><strong>All</strong></td>
<td>818</td>
<td>137</td>
<td>16.7</td>
</tr>
</tbody>
</table>
guidelines which explicitly mentioned the level of evidence. The 28 resources indexed in CISMeF as clinical guidelines without explicit mention of the level of evidence were indexed in the ANAES with other resource types, primarily evaluation studies.

Discussion

As health information has both the potential to improve health as well as to cause harm, organizations and individuals who provide health information on the Internet have an ethical obligation to be trustworthy, provide high quality content, protect users' privacy, and adhere to the standards of the best practices in health care (2). This topic has already been extensively studied; we found 104 articles with 'Internet (MeSH) and quality control (MeSH), (2001-09-30) on PubMed.

With the exception of the process used for scaling the level of evidence, very few sites (4.7%) routinely mention the validity of the information they deliver to professional or lay users through a score of the level of evidence. The results of this survey demonstrate that there is an urgent need to encourage publishers of 'sensitive' health information to give an indication of the level of evidence for each piece of information they provide. Nonetheless, some publishers already work with this approach (e.g., the French National Federation of Anti-Cancer Centres (FNCLCC) which indicates the level of evidence in almost 90% of the guidelines it has published) (27-28).

A positive trend has been observed for indicating the level of evidence in documents over the past five years, compared with the 1990-1996 period.

In June 1999, a French Medical Virtual University (FMVU) consortium (29) was created to experiment with various tools and methods necessary to build a virtual university. In 2001, the FMVU consortium has scheduled six 2-day training sessions to explain the basic concepts of e-learning, such as computer-assisted problem-based learning or virtual tutorials. During these seminars, the FMVU consortium promotes the necessity to indicate the level of evidence in 'sensitive' teaching material which is currently completely absent, at least in the French language documents that were surveyed (Table 2). Similar results were obtained in a previous study focusing on cardiology teaching material, in 43 French medical schools (30).

Table 4

<table>
<thead>
<tr>
<th>Publisher</th>
<th>Number of guidelines and conference consensus (GCC)</th>
<th>GCC with indication of level of evidence</th>
<th>Prevalence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>French Federation of Anti-Cancer Centres</td>
<td>29</td>
<td>26</td>
<td>89.7</td>
</tr>
<tr>
<td>(FNCLCC)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>French Society of Cardiology (SFC)</td>
<td>6</td>
<td>3</td>
<td>50.0</td>
</tr>
<tr>
<td>Society of Obstetricians and Gynaecologists of Canada (SOGC)</td>
<td>34</td>
<td>13</td>
<td>38.2</td>
</tr>
<tr>
<td>French Agency of Health Accreditation and Evaluation (ANAES)</td>
<td>97</td>
<td>35</td>
<td>36.1</td>
</tr>
<tr>
<td>Canadian Medical Association (CMA)</td>
<td>26</td>
<td>8</td>
<td>30.8</td>
</tr>
<tr>
<td>French Society of Anesthesiology and Reanimation (SFAR)</td>
<td>13</td>
<td>2</td>
<td>15.4</td>
</tr>
<tr>
<td>Health Canada</td>
<td>92</td>
<td>18</td>
<td>19.6</td>
</tr>
<tr>
<td>French Health Products Safety Agency (AFSSAPS)</td>
<td>87</td>
<td>6</td>
<td>6.9</td>
</tr>
<tr>
<td>Quebec Physicians College (CMQ)</td>
<td>17</td>
<td>1</td>
<td>5.9</td>
</tr>
<tr>
<td>Canadian Paediatric Society (CPS)</td>
<td>106</td>
<td>1</td>
<td>0.9</td>
</tr>
<tr>
<td>All</td>
<td>507</td>
<td>113</td>
<td>22.3</td>
</tr>
</tbody>
</table>

Although, it is true that there is a lack of a reference method for evaluating the level of evidence, this should not be used as a justification for not tackling the problem. There is a real and urgent need for a universal rating scheme. In the absence of an accepted reference method, a simple graded score of the level of evidence could be used (e.g., A, B, C or I, II, III) with simple explanations of the meaning of each category so that consumers can understand the relative values of the level of evidence (Table 1, the ANAES method to evaluate the level of evidence). Whatever the method used, consumers should be made aware of the need for 'sensitive' documents to explicitly mention the level of evidence for each 'sensitive' statement as an indication of the quality of 'sensitive' content on health Internet sites.

Some specialized catalogues, such as the American National Guideline Clearinghouse (URL: http://www.guideline.gov) or the Canadian Medical Association Info-base – Clinical Practice Guidelines (URL: http://www.cma.ca/cpgs/index.asp) indicate the level of evidence of indexed guidelines. To our knowledge, CISMeF is the first Web quality-controlled health gateway to explicitly indicate whether the level of evidence is mentioned for each 'sensitive' document indexed including guidelines, consensus conferences technical reports and teaching material using a simple and pragmatic process to check the indication of level of evidence. This method showed no false negatives which would requires constant verification by e-mail with the publisher. Furthermore, this criterion is rapidly searchable using the Doc‘CISMeF tool. One drawback to our study is the lack of a reference method to describe the level of evidence.

This study has some limitations: It is based on a catalogue which is not exhaustive by nature although the CISMeF team continues to greatly focus on indexing guidelines and more generally evidence-based medicine resources. Although it is currently only based on French language documents,
this study should also be performed on English or Spanish-language resources.

Moreover, the method we used to identify scored documents might have omitted some documents with level of evidence score which was not detected by the search string “level of evidence”. The verification of A N A E S guidelines suggests it was unlikely that we overlooked a sufficient number of documents with explicit mention of the level of evidence to drastically change the overall conclusion. A third limitation could be our definition of sensitive information and the method we used to assess the survey material, although we checked that all the documents containing sensitive information were correctly classified. Nevertheless, this study should be followed by further research, in particular the validation of this indicator in medical practice using content assessment, a study to evaluate the percentage of ‘sensitive’ information with the highest level of evidence and the use of objective criteria from the different existing methods to calculate the level of evidence (e.g. the method to retrieve as many relevant references as possible). These other studies will be carried out in cooperation with the French Agency of Health Accreditation and Evaluation (A N A E S).

Indicating the level of evidence after each ‘sensitive’ statement could be an effective way to assess the quality content of a Web site as this approach would use selected indicators to highlight relevance and utility. We consider that a Web site with the highest level of evidence and the use of objective criteria from the different existing methods to calculate the level of evidence (e.g. the method to retrieve as many relevant references as possible) should be of high quality regarding technological and health aspects as well as content as suggested by the RAND Corporation more recently by the RAND Corporation (ANAES).

In Europe:

In the USA:
Code of ethics of the Internet Healthcare Coalition (URL: http://www.ichealthcoalition.org/ethics/ethics.html) (1)
Hi-E thics (URL: http://www.hiethics.com)

In Japan:
With the Japanese version of the e-Health Code of Ethics created by the JIMA (Japan Internet Medical Association) (URL: http://www.jima.or.jp/trust/eh/health/ethics_jp1.pdf) (7).

Appendix

Initiatives recently available to assess the quality of information on Internet health sites:

In Europe:

In the USA:
Code of ethics of the Internet Healthcare Coalition (URL: http://www.ichealthcoalition.org/ethics/ethics.html) (1)
Hi-Ethics (URL: http://www.hiethics.com)

In Japan:
With the Japanese version of the e-Health Code of Ethics created by the JIMA (Japan Internet Medical Association) (URL: http://www.jima.or.jp/trust/eh/health/ethics_jp1.pdf) (7).

References

Methods Inf Med 1/2003

Correspondence to:
S. J. Darmoni
Rouen University Hospital
1, rue Germont
F-76031 Rouen Cedex, France
E-Mail: Stefan.Darmoni@chu-rouen.fr