Information in practice

Towards quality management of medical information on the internet: evaluation, labelling, and filtering of information

Gunther Eysenbach, Thomas L Diepgen

Editorial by Coiera

Unit for Medical Informatics, Epidemiology, and Public Health, Department of Dermatology, University Hospital Erlangen, Hartmannstrasse 14, 91052 Erlangen, Germany

Gunther Eysenbach, *resident*

Thomas L Diepgen, *consultant in dermatology*

Correspondence to: Dr Eysenbach Gunther.Eysenbach@ derma.med. uni-erlangen.de

The principal dilemma of the internet is that, while its anarchic nature is desirable for fostering open debate without censorship, this raises questions about the quality of information available, which could inhibit its usefulness. While the internet allows "medical minority interest groups to access information of critical interest to them so that morbidity in these rare conditions can be lessened,"[1] it also gives quacks such as the "cancer healer" Ryke Geerd Hamer a platform (http://www.geocities.com/HotSprings/3374/index.htm).[2–4]

Quality is defined as "the totality of characteristics of an entity that bear on its ability to satisfy stated and implied needs."[5] For quality to be evaluated, these needs have to be defined and translated into a set of quantitatively or qualitatively stated requirements for the characteristics of an entity that reflect the stated and implied needs. So how can we define consumers' "needs" in the case of medical information on the internet?

The quality of medical information is particularly important because misinformation could be a matter of life or death.[6] Thus, studies investigating the "quality of medical information" on the various internet venues—websites,[7] mailing lists and newsgroups,[8 9] and in email communication between patients and doctors[10]—are mostly driven by the concern of possible endangerment for patients by low quality medical information. Thus, quality control measures should aim for the Hippocratic injunction "first, do no harm."

Most papers published so far about the problem of quality of medical internet information focus on assessing reliability, but, as box 1 shows, this should be only one aspect of quality measures aiming for "first, do no harm." Another should be to provide context. Although these two problems are different in nature and different measures may be proposed to solve them, we discuss a common measure that could solve both aspects at the same time: assigning "metadata" to internet information; both evaluative metadata to help consumers assess reliability and descriptive metadata to provide context.

Benchmarks

Ideally, the success of methods of quality control and evaluation would be tested by their impact on morbidity, mortality, and quality of life. Such benchmarks would, however, be extremely difficult to measure.[12]

Summary points

The quality of information on the internet is extremely variable, limiting its use as a serious information source

A possible solution may be self labelling of medical information by web authors in combination with a systematised critical appraisal of health related information by users and third parties using a validated standard core vocabulary

Labelling and filtering technologies such as PICS (platform for internet content selection) could supply professionals and consumers with labels to help them separate valuable health information from dubious information

Doctors, medical societies, and associations could critically appraise internet information and act as decentralised "label services" to rate the value and trustworthiness of information by putting electronic evaluative and descriptive "tags" on it

Indirect "cybermetric" indicators of quality determined by computer programs could complement human peer review

Therefore, measures of process and structure[13] could be used as more indirect indicators of quality—for example, reliability, provision of context, qualification of authors, use or acceptance of this information by consumers, etc.

Filtering and selecting information

Table 1 shows different systems for quality control of information on the internet. If quality control at the time of production is not possible or not desirable,[14] it could be decentralised and consist of selecting the products complying to the quality requirements of a consumer. Such selection may consist of downstream filtering (by consumers) and upstream filtering (by an intermediary).

<div style="background:#cfe8f3">

Box 1: Why internet information is different from printed information

Characteristics of internet that make information and communication over this medium "special"
• Complete lack of quality control at stage of production, leading more easily to lack of reliability
• A "context deficit" leading to situation where information does not necessarily have to be false to harm

Examples of "context deficit"
• Less clear "markers" than in traditional publishing to allow patients to easily recognise a document as intended for professionals rather than for patients. Patients reading information intended for health professionals may misinterpret information,[6] leading to false expectations about treatment options, etc
• It is possible to read a web page without having seen context pages or the "cover" page containing disclaimers, warnings, etc
• Anonymity (of authors) may cause additional problems. Authors of web pages, news articles, emails, etc, sometimes remain unidentified
• Health information that is valid in a specific healthcare context may be wrong in a different one: "A free market of information will conflict with a controlled market in health care"[11]

</div>

Selection by third parties (upstream filtering)

Today, many reviewed indexes (review services) rate medical websites.[15 16] In this "upstream filtering" approach, third parties set quality criteria and also perform the evaluations, usually by means of a few human reviewers. This is one possible form of "distributed" quality management, but it has problems (see box 2).

Filtering by the user (manual downstream filtering)

An approach that circumvents some of the problems of upstream filtering (especially that of the volatility of internet information) is that of third parties communicating selection criteria to users (without any attempt to rate internet information themselves) to help consumers to evaluate ("filter") information "manually"

<div style="background:#cfe8f3">

Box 2: Drawbacks of upstream filtering

Volatility—The internet is too dynamic and rapidly changing to be reviewed by a few such filtering services. A solution for this problem could be that more and more highly specialised services could evolve, serving the special needs of certain user groups and focusing on certain internet venues, including newsgroups and mailing lists[8]

Questionable validity and reliability of rating instruments—A recent systematic review assessing 47 rating instruments for medical websites concluded that "many incompletely developed instruments to evaluate health information exist on the internet. It is unclear, however, whether they should exist in the first place, whether they measure what they claim to measure, or whether they lead to more good than harm."[15] Many of these services merely provide a badge or "seal of approval" or assign stars, medals, apples, thumbs, or sunglasses to websites,[15 16] which may, at best, give users a remote idea on the reliability of the website (leaving aside that the rating itself may be of questionable reliability and validity)

Rating cannot take into account users' context and needs—Quality criteria are fixed by third parties, and consumers may have different requirements than the reviewers. A link to a document written by an expert scientist and rated four stars by another expert may be useless for a patient. Equally, a document written for general practitioners may be of limited use for medical specialists

Users have to check a review service explicitly before or after reading a web page to check its rating—How many users who end up directly on a website because they used a search engine take the effort to make a second search of reviewed indexes for the rating of that site? How many users further try to obtain the ratings from different rating services in order to compare them and to estimate their reliability and interobserver variance? And if they did so, how should they interpret one service rating the website two stars and another rating it three sunglasses?

</div>

on their own.[17] The huge drawback of this approach is that it does not really help consumers to find high quality information quickly, as they have to check manually each entity (website, email, news article) against the given set of quality criteria.

Table 1 Different systems for quality control of information on the internet, ranging from present state of uncontrolled information to an unrealistic and undesirable state of full centralised control of information. In between are two decentralised filtering approaches: the present "upstream filtering" approach, and a possible future "downstream filtering" approach supported by software

	Uncontrolled, raw information	Intermediate state of labelled or filtered information ("bottom up" quality control)		Centrally controlled information ("top down" quality control)
		Present system	Possible future system	
Quality control	None	Decentralised control by a few third parties	Decentralised control by many third parties and users	Central control
Quality criteria	None	Set by third parties	Set by users	Set by central institution
Structure	Anarchic	In principle anarchic, but with option for users to take guidance by selecting rating services	In principle anarchic, but with option for users to take guidance by letting software automatically consult multiple rating services	Governed structure
Data	Raw data	Data evaluated by third party	Data labelled (categorised, rated, weighted) by author or third party, or both	Data edited or controlled
Filtering	None, unimpeded rivers of data	Upstream filtering by third parties (review services). Users cannot influence selection criteria	Downstream filtering by users and collaborative filtering. Users define selection criteria	Upstream filtering imposed by central body (such as government). Filtering criteria set by third party
Control of information	Chaotic, anarchic state with information of unclear quality	Information evaluated by third parties	High quality information easy to find. Data rated, labelled, and weighted according to users' criteria	"Censored" information

Filtering by the user supported by software (automatic downstream filtering)

We therefore propose to focus on a third approach, automatic downstream filtering. Here, quality criteria are set up by third parties and translated into a computer readable vocabulary, and the filtering is done, at least partly, by users' software.

A prerequisite for this approach is that internet information is labelled with "metadata" in a standardised format to allow software to search for and check information that is suitable for an individual user. Metadata can be provided by authors within the information itself, describing the contents and context of the information, but, more importantly, users' software could also request metadata from third parties (rating services) to see whether a rating service provides additional descriptive or evaluative information about the item retrieved. Software products (browsers) may be customised by clients in order to filter out any information that does not meet the personal quality requirements or interests of the user.

As both types of metadata (the authors' and those of third parties) can also be indexed in search engines, this approach also helps users to find information directly.

Electronic labels

The World Wide Web Consortium has recently developed a set of technical standards called PICS (platform for internet content selection)[18–21] that enable people to distribute electronic descriptions or ratings of digital works across the internet in a computer readable form. PICS was originally developed to support applications for filtering out pornography and other offensive material, to protect children. An information provider that wishes to offer descriptions of its own materials can directly embed labels in electronic documents or other items (such as images)—for example, such labels may indicate whether the content is appropriate for particular audiences such as minors, patients, etc.

Perhaps even more important, independent third parties, so called label services, can describe or evaluate material—human reviewers or automatic software (see below) rate websites and create electronic labels. An end user's software will automatically check at the label bureau(s) that the user is subscribed to while accessing a website or retrieving any other kind of digital information. The software further interprets the computer readable labels and checks them against the requirements defined by the user. It may then, for example, display a warning if the information is aimed at a different audience or if the website is known to contain misleading health information, etc.

The quality criteria (in PICS terms "rating categories") and their scales are together called rating vocabulary. We have developed a prototype core vocabulary, med-PICS, for possible use with medical information.[22] This vocabulary contains descriptive categories such as the intended audience (from "kids" to "highly specialised researcher"), which could be used by authors to provide "context," and evaluative categories such as "source rating" (from "highly trustworthy" to "known to provide wrong or misleading infor-

mation"), which could be used by third party label services.

The main advantages of automatic downstream filtering would be
● The exact quality requirements can be set by the user, not by the rating service alone. The rating service describes the information with values on defined scales in different categories, and the user determines the thresholds. For example, a user could tell the software, "I want only material that is suitable for patients, which relates to the healthcare setting in Britain, and which is rated of at least medium reliability"
● The software could automatically check one or more rating services in the background, without the user having explicitly to consult a rating service before or after entering a website or retrieving any other kind of information.

The idea of assigning standardised metadata to medical information on the internet is not new,[23] but the key difference of using an infrastructure such as PICS is that not only can authors include metadata but third parties can also associate metadata to all kinds of information (see table 2). Until now metadata were primarily thought of as descriptive (provided by authors), but in the future metadata could also be evaluative (provided by third parties).

Who should evaluate and how

PICS is merely an infrastructure for distributing metadata, not a method per se to evaluate information. The questions of who should evaluate and how still remain.

Today, most of the rating of medical information is done by organisations, publishers, and sometimes individuals. We think that in the future more people from the medical community should evaluate internet information while they surf the internet. We propose a collaboration of medically qualified internet users, consisting of volunteers who, for example, get a program or browser extension that allows them to rate medical websites in a standard format. These ratings could be transmitted to one or several medical label databases, which could be used by consumers.

If thousands of doctors continuously took part in a global rating project we might be able to keep pace with the dynamics of the internet. With this true "bottom up" approach, one could also easily evaluate the rating instruments in terms of variation among observers. Further, the heterogeneity of the reviewers would take account of the many different perspectives and backgrounds that consumers may have as well.

Beyond peer review: automatic and semiautomatic methods of assessing quality

Traditional peer review has many problems, such as that reviewers are human and can make factually incorrect judgments and that peer reviewing is very time consuming. We therefore propose that more work should be made to explore the potential of computers to determine indirect quality indicators by means of automatic (mathematical) methods. Current research suggests that "web surfing" follows strong mathematical patterns,[24] and work in the new discipline of

Table 2 Comparison of quality control in traditional publishing and in present and possible future quality control on internet

	Traditional scholarly publishing	Internet publishing	
		Present (upstream filtering)	Possible future (downstream filtering)
What is rated or evaluated?	Scholarly articles	Mostly websites	In principle all information, including websites, multimedia items, news groups, emails
Structured description of work provided by author or third party	Structured abstracts preceding journal article, written by author or editor	Electronic labels within web pages may contain descriptive information of authors. Review services summarise contents of selected websites	Electronic labels within web pages (or sent with any information) contain descriptive information provided by author or by third party labelling service in standard computer readable format
Body in charge of quality control	Scientific journals	Review services (publishers, professional organisations, medical associations)	Third party labelling services (may be publishers, professional organisations, medical associations, etc)
Quality criteria	Checklists for peer reviewers	Criteria and requirements set by review services. Rating instruments are poorly developed[17]	Categories laid down in PICS rating system. Set of validated core categories should be defined and used by every rating service. Other categories could be developed by the individual rating services
Quality evaluation	Peer review by journal. Experts evaluate contents of article and decide on publication	Mainly by employed or volunteer reviewers working for review services. Little participation by readers	Review by many labelling services. Organisations, individuals, or computer programs rate contents of information and provide PICS labels. Expert readers participate in review process
Role of readers	Readers decide which journal to subscribe to (according to interests and trust in the quality of the journal). Readers further select on basis of abstracts and editorials which articles to read	Readers decide which review services to check for evaluative information. Finding this information can be difficult, and checks have to be done manually and sequentially	Readers decide once which labelling services to subscribe to (according to interests and trust in labelling service). Rating labels are checked automatically by client software. Readers further configure local filtering software so that only material meeting personal interests and quality requirements gets through. Readers may, if expert in subject, participate in collaborative rating effort
Indirect (automatic) quality indicators	Citation analysis	Hardly in use	Number, structure, or dynamics of hyperlink references; number of hits or users; number of labels requested; other automatic measures

"cybermetrics" has indicated promising methods for measuring the impact of websites—distinguishing low quality websites from high quality sites by analysis of user behaviour, user pattern, complexity of the website, etc (box 3). Of course, the specificity of such indicators is low (a popular website with many users may still harm with unreliable information), but they are sensitive and, once the methods are established and validated, easy to obtain.

> **Box 3: Possible indirect quality indicators suitable for automatic selection by software**
>
> *Web citations*—A "webcite index," analogous to the *Science Citation Index*,[25] could be compiled from the absolute number of hyperlinks to a certain website or new hyperlinks established over a period of time, etc (see http://webcite.net)
>
> *Number of visitors a day (determined by an independent party)*—This is analogous to the circulation in traditional publishing. It may be particularly valid if not all visitors are counted but only those from a certain (expert) user group: for example, calculation of the medical internet addresses visited most often by staff and students of a university hospital. If different departments around the world with common interests regularly exchanged this information for analysis, the user base would be huge and valuable information could be extracted
>
> *User behaviour*—Innovative indicators, which have no analogous counterparts in traditional publishing, may be based on user behaviour, such as number of hits per website, time spent visiting a web page, etc. More research is required to determine the relation of these rather unspecific indicators with quality. These indicators may be more helpful for "webmasters" rather than for third parties

Conclusion and call for action

While suggestions for an agreed formal international standard for medical publications on the internet, enforced by appropriate peer or government organisations,[26] are probably not realistic, there should at least be a core standard for labelling health related information. In our proposed collaboration for critical appraisal of medical information on the internet,[22] organisations, associations, societies, institutions, and individuals interested in reviewing, assessing, and compiling medical information will be invited to join the discussion.

The internet—a decentralised medium by nature—not only allows access to information distributed on various computers but also allows a distributed management of quality with decentralised quality control and evaluation. Filtering techniques and infrastructures such as PICS may help to overcome the present oligarchic approach of a few review services attempting to rate all the information of the internet towards a truly distributed, democratic, collaborative rating.

1 Dearlove OR, Sharples A, Stone C. Internet is useful for information on rare conditions [letter]. *BMJ* 1997;315:491.
2 De Bousingen DD. Austrian cancer patient's parents sentenced [news]. *Lancet* 1996;348:1440.
3 Nigel G. German "quack healer" arrested [news]. *Lancet* 1997;349:1679.
4 De Bousingen DD. German "quack healer" sentenced [news]. *Lancet* 1997;350:874.
5 International Organization for Standardization, Technical Committee ISO/TC 176. *ISO 8402: Quality management and quality assurance—Vocabulary*. 2nd ed. Geneva: International Organization for Standardization, 1994. (1994-04-01.)
6 Weisbord SD, Soule JB, Kimmel PL. Brief report: poison on line—acute renal failure caused by oil of wormwood purchased through the internet. *N Engl J Med* 1997;337:825.

7 Impicciatore P, Pandolfini C, Casella N, Bonati 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;314:1875-81.

8 Culver JD, Gerr F, Frumkin H. Medical information on the internet: a study of an electronic bulletin board. *J Gen Intern Med* 1997;12:466-70.

9 Hernandez-Borges AA, Pareras LG, Jimenez A. Comparative analysis of pediatric mailing lists on the internet. *Pediatrics* 1997;100:E8.

10 Eysenbach G, Diepgen TL. Responses to unsolicited patient e-mail requests for medical advice on the world wide web. *JAMA* 1998;280:1333-5.

11 Coiera E. The internet's challenge to health care provision—a free market in information will conflict with a controlled market in health care [editorial]. *BMJ* 1996;312:3-4.

12 Wyatt JC. Commentary: measuring quality and impact of the world wide web. *BMJ* 1997;314:1879-81.

13 Donabedian A. *Explorations in quality assessment and monitoring.* Vol 1. *The definition of quality and approaches to its assessment.* Ann Arbor, MI: Health Administration Press, 1980.

14 Silberg M, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the internet. Caveat lector et viewer—let the reader and viewer beware. *JAMA* 1997;277:1244-5.

15 Jadad AR, Gagliardi A. Rating health information on the internet—navigating to knowledge or to Babel? *JAMA* 1998;279:611-4.

16 McNab A, Anagnostelis B, Cooke A. Never mind the quality, check the badge-width! *Ariadne* 1997 May 18;9. (http://www.ariadne.ac.uk/issue9/quality-ratings.)

17 Ambre J, Guard R, Perveiler FM, Renner J, Rippen H. *White paper: criteria for assessing the quality of health information on the internet (working draft, edited 14 Oct 1997).* http://www.mitretek.org/hiti/showcase/documents/criteria.html [6 Apr 1998].

18 Krauskopf T, Miller J, Resnick P, Treese GW. Label syntax and communication protocols. *World Wide Web J* 1996;1(4):45-69. (http://w3.org/PICS/labels.html [28 Aug 1997].)

19 Miller J, Resnick P, Singer D. Rating services and rating systems (and their machine-readable descriptions). *World Wide Web J* 1996;1(4):23-43. (http://w3.org/PICS/services.html [28 Aug 1997].)

20 Resnick P, Miller J. PICS: internet access controls without censorship. *Commun ACM* 1996;39(10):87-93. (http://www.w3.org/PICS/iacwcv2.htm [28 Aug 1997].)

21 Resnick P. Filtering information on the internet. *Sci Am* 1997;3:106-8.

22 Collaboration for Critical Appraisal of Medical Information on the Internet. med-PICS electronic rating vocabulary for medical websites. http://medpics.org [28 Aug 1997].

23 Appleyard RJ, Malet G. A proposal for using metadata encoding techniques for health care information indexing on the WWW. *American Medical Informatics Association annual fall symposium, Nashville, 25-27 Oct 1997.* Philadelphia: Hanley and Belfus, 1997.

24 Huberman BA, Pirolli PLT, Pitkow JE, Lukose RM. Strong regularities in world wide web surfing. *Science* 1998;280:95-7.

25 Garfield E. Significant journals of science. *Nature* 1976;264:609-15.

26 Malet G. Medical matrix code of conduct beta version 1.0. http://206.1.96.60/forums/Thread.cfm?CFApp=3&Thread_ID=33&mc=2 [28 Aug 1997].

(Accepted 16 July 1998)

Hallmarks for quality of information

J A Muir Gray

NHS Executive Anglia and Oxford, Department of Health Institute of Health Sciences, Oxford OX3 7LF

J A Muir Gray, *director of research and development*

graym@rdd-phru.cam.ac.uk

The Goldsmiths' Company was founded in London in 1327 and has flourished for over 650 years. It never traded gold but specialised in the assay of gold and other precious metals. The Goldsmiths' Company has flourished because it has been an independent assay service, measuring the quality of gold and stamping the gold with a hallmark to indicate to the public the purity of the metal with an explicit system of measurement (the word "carat" derives from the Arabic for the carob bean, for the beans of the carob are of uniform size and can be used as standard weights).

Knowledge hallmarks are needed to perform the function of gold hallmarks, and the Cochrane logo has already become a knowledge hallmark, clearly defining the quality of knowledge because readers can look at the *Cochrane Collaboration Handbook* and see the methods used to produce and appraise the Cochrane Reviews. Journal titles have been another hallmark, but the dependability and credibility of that hallmark is fading as doubts increase about the rigour of the assay method called peer review and evidence shows that even in prestigious journals the assay procedure is flawed and unreliable. Worryingly, all the flaws in the assay procedure seem to overemphasise the strength of the positive effect of new interventions and treatments, with a significant increase in the positive effect of the treatment resulting from poor trial design (table 1) and biased reporting (table 2).

This is a problem in the paper world and will be even more of a problem in the electronic world, in part because electronic journals are so easy to create. Every time information on the world wide web has been critically reviewed or assayed, the quality has been shown to be very variable. Even more worrying, it is hard, and sometimes impossible, to assess the quality of a website because the necessary evidence is not present.

When the printing press was invented, there was concern that the printed word would give undue credibility to an idea or proposition. The same applied to the world wide web when it started, although people now have a healthier scepticism for anything on the web because of the rapid growth of electronic junk. However, the web is an important means of communication, and will become increasingly important when it becomes available on digital television. Already tools have been developed to monitor the quality of healthcare information: DISCERN and the National Centre for Information Quality are examples of initiatives taken to help the public appraise the quality of information provided to them. What is needed, however, is a common standard based on the intellectual equivalent of carob beans, with an Honourable Company of Healthcare Knowledgesmiths to run the assay procedure in an independent and disinterested way so that people can not only distinguish gold from a base metal but also know whether they are reading 24 carat or 18 carat knowledge.

Table 1 Effects of poor design of controlled trials on estimates of treatment effects (trials with poor evidence of randomisation compared with trials with adequate randomisation, data from Schultz et al[1])

Design fault	Exaggeration of odds ratio
Inadequate method of treatment allocation	Larger by 41%
Unclear method of treatment allocation	Larger by 30%
Trials not double blind	Larger by 17%

Table 2 Sources of positive bias in the reporting of controlled trials (data from Gray[2])

Source of bias	Cause
Submission bias	Research workers are more strongly motivated to complete, and submit for publication, positive results
Publication bias	Editors are more likely to publish positive studies
Methodological bias	Methodological errors such as flawed randomisation produce positive biases

1 Schultz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408-12.

2 Gray JAM. *Evidence-based healthcare.* London: Churchill Livingstone, 1997.

Quality on the internet

Maurizio Bonati, Piero Impicciatore, Chiara Pandolfini

Interest in searching the world wide web for health related information continues to increase, increasing the need for internet resources to be accountable to doctors and the public.[1] Function, structure, and content of a website are the main aspects used to evaluate material on the internet.[2] Although we have not yet developed reliable methods for evaluating the effects (the impact) of such material on clinical practice or on a user's behaviour, improved technology today allows for the control of function and structure of a website.

Eysenbach and Diepgen propose the use of a promising automatic "downstream filtering" system of metadata based on PICS technology. This uses a rating vocabulary that contains descriptive and evaluative categories based on rating instruments already available for evaluating health information on the internet. The authors suggest that assessing quality of information depends not only on evaluating its reliability but also on the provision of context; a valid idea in that it resembles the traditional system of submitting and publishing scientific articles. Thus, providing descriptive tags (metadata) for context and content—like supplying keywords for articles submitted for publication—would allow more accurate searches by web browsers.

The problem lies in assigning tags for reliability of information. Guidelines for every aspect of health care do not exist, so each "rater" in the authors' proposed collaboration for critical appraisal of medical information on the internet would assign his or her own values. The benefits of having many raters need to be weighed against the possibility of having unqualified or uninformed medical workers (and lay people) judge web information incorrectly. Who would then check the raters? It has, after all, been found that doctors are also sources of incorrect, outdated information on the internet.[3]

Thus far, more attention has been paid to presentation and reliability than to the accuracy of the content material.[4] To determine the accuracy of medical information on the internet we need to compare it with the best evidence.[2] The evidence based methodology and the Cochrane Collaboration are two useful examples of critical appraisal that should also characterise future evaluations of websites. In the meantime, interaction and feedback may be markers of high quality for websites: allowing a user to submit comments or questions demonstrates a serious intention by the authors to both improve the information supplied by them and to become respectable sources of health information in the long run.

This is a just a starting point for the demystification of medicine and the development of real partnerships between all parties concerned. We must find ways of producing, validating, and diffusing appropriate information in a manner that involves users (consumers) in order to guarantee a non-authoritarian practice, access for all to healthcare information, and high quality information on the internet.

Laboratory for Mother and Child Health, Istituto di Ricerche Farmacologiche "Mario Negri," Via Eritrea 62, 20157 Milan, Italy

Maurizio Bonati, *head*

Piero Impicciatore, *senior research fellow*

Chiara Pandolfini, *research fellow*

Correspondence to: Dr Bonati mother_child@ irfmn.mnegri.it

1 Morris TA, Guard RJ, Marine SA, Schick L, Haag D, Tsipis G, et al. Approaching equity in consumer health information delivery: NetWellness. *J Am Med Inf Assoc* 1997;4:6-13.
2 Wyatt JC. Commentary: measuring quality and impact of the world wide web. *BMJ* 1997;314:1879-81.
3 Impicciatore P, Pandolfini C, Casella N, Bonati 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;314:1875-81.
4 Jadad AR, Gagliardi A. Rating health information on the internet. Navigating to knowledge or to Babel? *JAMA* 1998;279:611-4.

Assuring quality and relevance of internet information in the real world

Subbiah Arunachalam

Interest in how new information technology can be used to improve health is growing steadily. Telemedicine is making it possible to erase geographical constraints on the provision of health care. However, the information revolution is not a worldwide phenomenon: in India today there are fewer than two main telephone lines per 100 people. Even in Western countries such as the United States there is a wide disparity in terms of access to telephones and computers between poor communities—inner city populations, blacks, and Hispanics—and the suburban elite.

Access to technology is only a part of the problem. There are three aspects to provision of information: collection, distribution and dissemination, and authentication and quality control. While the internet is good at the first two, the information it provides is not thought to be very dependable or reliable. Eysenbach and Diepgen address this problem with regard to medical information and suggest "distributed quality management" as a possible solution. They argue their case—that questions of both relevance and reliability can be tackled by a common measure—very well. In particular, their proposal that both "top down" and "bottom up" approaches involving peer review by a large body of people should be used is attractive and could be cost effective.

There are good examples of achieving quality assurance through a combination of centralised and decentralised approaches in other specialties. The United Nations Environment Programme has the maESTro (Managing Environmentally Sound Technologies) program, which operates from Japan and which verifies with the developer of the technology as well as cross checks with databases (http://www. unep.or.jp/ietc/ESTdir/maestro/introduction.html). In

M S Swaminathan Research Foundation, Taramani Third Cross Street, Chennai 600 113, India

Subbiah Arunachalam, *distinguished fellow*

subbiah_a@ hotmail.com

physics the e-Print Archive, based in Los Alamos, works well. Usually, if someone wants to comment on a preprint, he or she directs it to the author, but some do forward their comments to the archive, thus making it available to the worldwide audience.

Unlike in physics and technological information services, in medicine a whole range of people, and not only experts, take part in the information exchange, both inputting and searching. Well known sites such as those of the *BMJ*, *JAMA*, and *Human Genome News* (http://www.ornl.gov/TechResources/Human_Genome/publicat/hgn/vgn3/01eyes.html) are dependable, but what about all the material in usenet groups, listservs, and email messages? In this respect medicine is closer to astrology than to the hard sciences—hence the need for assuring quality. We should encourage doctors and biomedical researchers, as well as institutions, to comment on what they see on the internet. Also, agencies such as Magellan and Starting Point (web search engines that also evaluate websites) perform the function of third party evaluators. Ultimately, the reliability of the meta-analysis approach (gaining new insights by amalgamating existing data from different sources) would depend on the weights we give to differ-

ent constituents in the distributed, democratic, and collaborative process of rating suggested by Eysenbach and Diepgen. Another problem, not just with medical information but with any information, is the cost of standardisation of vocabulary, evaluation procedures, etc. Who will pay?

Certain new developments in searching the world wide web, such as the "hyperlink induced topic search" developed by Jon Kleinberg of Cornell University and being evaluated by IBM and Digital (now Compaq) for implementation, can help to reduce the time taken to find relevant medical information in an internet search (see http://www.almaden.ibm.com/cs/k53/clever.html). This is similar to the citation links in journal literature that form the basis of the *Science Citation Index*. But it is still unclear whether the system for the internet will be as powerful as the citation indexes in clustering related material through cognitive links.

Finally, we live in the real world, and there can be no ideal solution to our problems. Every time we find a way to overcome a problem, those that create the problem do things to make our solutions inadequate. But scepticism should not hold us back from looking for ways to make the internet the ultimate source of easily accessible and reliable information.

A patient who changed my practice?
The new Holy Grail

E was an 87 year old man with chronic bronchitis and some deafness. He presented yet again with increasing dyspnoea. I listened to his chest for the expected wheeze. "Haven't really got much better Doc, started the steroids like you said and doubled my inhalers." Then I noted in the background the irregular, rapid pulse. I congratulated myself for my clinical acumen in spotting his atrial fibrillation.

For some time I had considered with unease the emerging evidence in favour of anticoagulation for older patients with atrial fibrillation. Audit within the practice showed only 60% of patients on any form of antithrombotic medication and most of these on aspirin.

I attend a typically didactic lecture from a group of teaching hospital cardiologists presenting the case for anticoagulation with impressive statistics and apparent scientific rigour. At the end I voice my worries. Is the evidence from hospital based studies or have they recruited patients directly from general practice? Should I go back to my practice and anticoagulate all those elderly people who have never been anywhere near a hospital doctor? "Good question," conceded the lecturer, "most studies are hospital based, but the hearts of your patients and mine will not be significantly different."

I am concerned he does not appreciate the increased frailty, multiple disorders, polypharmacy, tendency to fall, and mild confusion so prevalent in elderly patients managed in primary care and the subsequent impact on the safety of anticoagulation. Many of these patients are muddled about straightforward doses of digoxin, never mind warfarin with varying doses on different days of the week.

So I speak to my local consultant geriatrician for advice. "I try to find a reason not to," when faced with anticoagulation in the very elderly patient with atrial fibrillation, she says.

I am pleased when E walks into the surgery one Monday a few weeks later. I felt that I had done well and not fallen into that common general practice trap of assuming recurrently presenting symptoms were due to his usual previously known condition. He looks grave as he hands me the letter from the

emergency eye unit. Retinal artery embolus—"They say I should have been on blood thinners Doc."

Evidence based medicine is the new Holy Grail. But is the evidence really rock solid? If not, what makes us change our practice and reject well tried and trusted methods?

Have I a unique perspective and knowledge of my patients in primary care, or has a doctor who has "fallen off the ladder" of hospital medicine no right to interpret the views of the experts when incorporating new evidence into my practice?

This anecdotal experience has undoubtedly lowered my threshold for anticoagulation of these patients far more than the evidence base, about which I had some doubts. Is my practice now evidence based? Perhaps more realistically it is guilt driven.

So now I know what I'll do when faced with this again, or do I? Today E presented with haematuria. Last week an elderly woman who I knew well died of a cerebral haemorrhage. She had presented to me with atrial fibrillation and an arterial embolus in a lower limb some four years ago. Ironically, she was put on aspirin by a registrar in vascular surgery; this was changed to warfarin only after I suggested this management option to the consultant. In a final twist of fate she collapsed on the way to her anticoagulant clinic.

But, of course, soon I will no longer need to ponder these issues; the National Institute for Clinical Excellence will do it for me.

Mark Gradwell, *general practitioner, Stockport*

We welcome articles of up to 600 words on topics such as *A memorable patient, A paper that changed my practice, My most unfortunate mistake*, or any other piece conveying instruction, pathos, or humour. If possible the article should be supplied on a disk. Permission is needed from the patient or a relative if an identifiable patient is referred to, We also welcome contributions for "Endpieces," consisting of quotations of up to 80 words (but most are considerably shorter) from any source, ancient or modern, which have appealed to the reader.