

Web page quality: can we measure it and what do we find? A report of exploratory findings

Vincent P. Abbott

Abstract

Background The aim of this study was to report exploratory findings from an attempt to quantify the quality of a sample of World Wide Web (WWW) pages relating to MMR vaccine that a typical user might locate.

Method Forty pages obtained from a search of the WWW using two search engines and the search expression 'mmr vaccine' were analysed using a standard proforma. The proforma looked at the information the pages contained in terms of three categories: content, authorship and aesthetics. The information from each category was then quantified into a summary statistic, and receiver operating characteristic (ROC) curves were generated using a 'gold standard' of quality derived from the published literature. Optimal cut-off points for each of the three sections were calculated that best discriminated 'good' from 'bad' pages. Pages were also assessed as to whether they were pro- or anti-vaccination.

Results For this sample, the combined contents and authorship score, with a cut-off of five, was a good discriminator, having 88 per cent sensitivity and 92 per cent specificity. Aesthetics was not a good discriminator. In the sample, 32.5 per cent of pages were pro-vaccination; 42.5 per cent were anti-vaccination and 25 per cent were neutral. The relative risk of being of poor quality if anti-vaccination was 3.3 (95 per cent confidence interval 1.8, 6.1).

Conclusion The sample of Web pages did contain some quality information on MMR vaccine. It also contained a great deal of misleading, inaccurate data. The proforma, combined with a knowledge of the literature, may help to distinguish between the two.

Keywords: World Wide Web, quality, ROC curves, MMR vaccine

Introduction

The World Wide Web (WWW) has, according to Silberg *et al.*,¹ vast areas of incomplete, misleading or inaccurate information. Silberg *et al.* and other researchers^{2–6} have expressed great concern over the quality of medical information on the Web. This becomes an important public health problem when the Web is used to access medical information. The WWW is increasingly being accessed by the general public on a wide range of topics,⁷ including medical ones, yet the content of the pages is largely unregulated. An increasing number of individuals have their own Web page and may put whatever

information they wish onto it.^{7,8} The enquirer, looking for information on a particular topic, will be presented with many possibly relevant pages and no way of deciding whether the information is reliable. This makes it impossible to make an informed decision on the basis of the information obtained. If the information on the Web, for example about MMR vaccine, is wrong, misleading or unbalanced, then it may result in people refusing vaccination on the basis of this information. There is evidence that this is happening⁹ and that vaccine coverage does fall as a result of adverse publicity.^{10,11}

In an attempt to overcome this problem, some researchers have laid down guidelines as to what a 'good' Web page should contain.^{1,12–18} These include stating the authorship of the page with credentials, giving references, openly stating any sponsorship or advertising, and stating the date and time of posting and updating of the information. Others consider aesthetics and a balanced reporting of the evidence about a treatment also to be important. This field has recently been reviewed by Kim *et al.*¹⁹

There are, in my opinion, two problems with this guidelines approach. First, even if most of the above criteria are satisfied, it still does not guarantee that the information contained in the page is correct. It does not necessarily equate with accuracy. Second, it is not clear how many of the criteria have to be satisfied for a page to be deemed good quality. Nor is it always clear whether some criteria are more important than others. In other words, the assessment of the page tends to be more qualitative than quantitative.

This study attempts to follow Wyatt's⁶ suggestion that 'evaluators should first identify a subset of web sites which typical users do locate and then assess the quality of these sites'. To do this the study attempts to develop a test for measuring aspects of Web page quality using a standard proforma. The result is a number – a 'quality score' – for a Web page. The work then proceeds to compare this score against a 'gold standard' to establish an ideal cut-off value that best distinguishes 'good' quality pages from 'bad' ones. A not unreasonable assumption is made that a factually correct and balanced Web page is a good quality page. But equally

Department of Public Health, North West Lancashire Health Authority, Wesham Park Hospital, Wesham PR4 3AL.

Vincent P. Abbott, Specialist Registrar in Public Health Medicine

important is the converse corollary of this assumption, which is that no matter whatever else may be good about a page, if it is inaccurate or unbalanced then it is poor quality. Accuracy and balance are deemed to be indicators of quality.

Method

The Web pages sample

Web search method

The WWW was searched, over a period of days, using two search engines, AltaVista²⁰ and Lycos,²¹ and the search expression 'mmr vaccine'. These search engines were chosen because, although they both aim for 100 per cent coverage of the Net, they index the information in slightly different ways.^{22,23} Although no search engine can ever provide 100 per cent coverage, Lycos claimed 91 per cent coverage in February 1996.²² Hierarchical search engines, such as Yahoo!,²⁴ were not used for two reasons. First, a trial search carried out by myself for pages on MMR vaccine found no pages. This may have been due to my own poor searching technique but if I, a fairly experienced Web user, could not find any pages, then a member of the general public would be no more likely to find any either. Second, the sites included in their directories are those sites that have been registered by the page creator.²² Consequently, only those sites that are aggressively managed are included. MMR vaccine was chosen because it was, at the time, very much in the public eye following a recent publication suggesting a link with autism.²⁵

Inclusion and exclusion criteria

To pilot the proforma, the first 20 relevant pages from each search (a total of 40 pages) were downloaded onto disk. As Web pages can change very rapidly²⁶ it was these downloaded copies that were subsequently analysed. No attempt was made to go back to the original Web page address. A page was defined as relevant if it was a Web site and contained information on the MMR vaccine. This excluded discussion groups, message boards, pages that only mentioned the phrase MMR vaccine but gave no information about it and pages that were lists of links to other pages. Links were followed only if information in the page or the link suggested that the linked page would meet the inclusion criteria. Any such linked pages were included for analysis. A random sample was not taken because search engines attempt to order pages in terms of relevance to the searcher; the more closely matched are nearer the top of the list. The mechanics of ranking have been discussed elsewhere.²²

The gold standard

To be able to make a judgement on the factual accuracy of a page, a literature search was undertaken to determine the current state of knowledge with regard to the MMR vaccine in terms of its effectiveness, efficacy, uses, side-effects, contra-indication and dangers. This was carried out in four ways:

searching Medline, following up relevant references in identified papers, consulting the 'Green Book'²⁷ and by discussion with a local Consultant in Communicable Disease Control. Each Web page was judged against this current literature to determine factual accuracy. All Web pages were also assessed for balance: stating side-effects and contra-indications along with advantages. If a page was both factually accurate and balanced then it was classified as 'good'; otherwise as 'bad'. This classification was used as a 'gold standard' against which to compare the proforma score.

The proforma (see Appendix)

The layout

Each Web page had a proforma completed for it. The proforma, modified from Pealer and Dorman,¹² consists of six sections. The first part records general information about the page. Of the remaining five sections, section 1 looks at the content, section 2 the authorship and section 3 aspects of the aesthetics of the page. Section 4, completed last, is a judgement based on the information contained in the page as to whether the page is pro-vaccination, anti-vaccination or neutral.

Method of use

With the exception of question 1.4: 'Flesch Reading Ease Score' (FRES) and 3.5: 'Minimal page layering', all questions in sections 1–3 are answered by a 'yes' (scoring 1) or 'no' (scoring zero). Question 1.4 was answered by loading the page into Corel[®] WordPerfect[®] version 8 and using the program to calculate the FRES. This is then converted to a score as follows: FRES 0–50 scores as zero; 51–60 as one; 61–90 as two; 91–100 as one. Question 3.5 scores one if no links are followed to access the page, and zero otherwise. The results for each section are then added, resulting in three scores: one each for content, authorship and aesthetics.

The analysis

I compared the results from the proforma with my gold standard and generated a series of 2 × 2 tables of the general format of Table 1 for all possible scores for each of the three sections: content, authorship and aesthetics.

The resultant sensitivities and specificities were used to generate receiver operating characteristic (ROC) curves (Fig. 1). From these curves the best cut-off values for each of the three sections of the proforma were determined. The best cut-off value is that which best distinguished 'good' pages from 'bad' ones with the fewest false positives and false negatives. If any of the three sections appeared, independently, to be good discriminators then their scores were added together and further ROC curves produced (Fig. 2).

Receiver operating characteristic (ROC) curves

The theory of ROC curves has been well documented

Table 1 General 2 × 2 table for calculating the sensitivity and specificity of all possible scores for each of the three sections (content, authorship and aesthetics) of the proforma

	Gold standard		Total
	Number of 'good' pages	Number of 'bad' pages	
Proforma section score > <i>n</i>	<i>a</i>	<i>c</i>	<i>a + c</i>
Proforma section score ≤ <i>n</i>	<i>b</i>	<i>d</i>	<i>b + d</i>
Sensitivity (%) =	$a/(a + b) \times 100\%$	$d/(c + d) \times 100\%$	= Specificity (%)

a, number of pages judged 'good' that scored > *n*; *b*, number of pages judged 'good' that scored ≤ *n*; *c*, number of pages judged 'bad' that scored > *n*; *d*, number of pages judged 'bad' that scored ≤ *n*; *n*, score produced by the proforma for each of the sections 1–3.

elsewhere.^{28–34} They can be used to assess how well a test can discriminate between a population having a particular characteristic and another lacking it. To generate the ROC curve, one plots the sensitivity on the ordinate (y-axis) against one minus the specificity on the abscissa (x-axis) for all possible values of the test. A test with no discriminatory power will follow the *y = x* line and have an area under the graph of 50 per cent of the total possible area. A test with perfect discriminatory power will rise up the *y*-axis to 100 per cent and then run parallel to the *x*-axis and have an area under the graph of 100 per cent of the total. The greater the area under the line, the better the discriminatory power of the test. The best cut-off

value for the test is the value that maximizes the vertical distance between the test line and the *y = x* line.

Results

General

Of the 40 pages analysed, 32.5 per cent were pro-vaccination, 42.5 per cent were anti-vaccination and 25 per cent were not obviously either. Only 40 per cent were judged 'good' by the criteria of agreeing with current literature and being balanced. Hence 60 per cent of the sample were inaccurate or unbalanced. Only 15 per cent of pages had references.

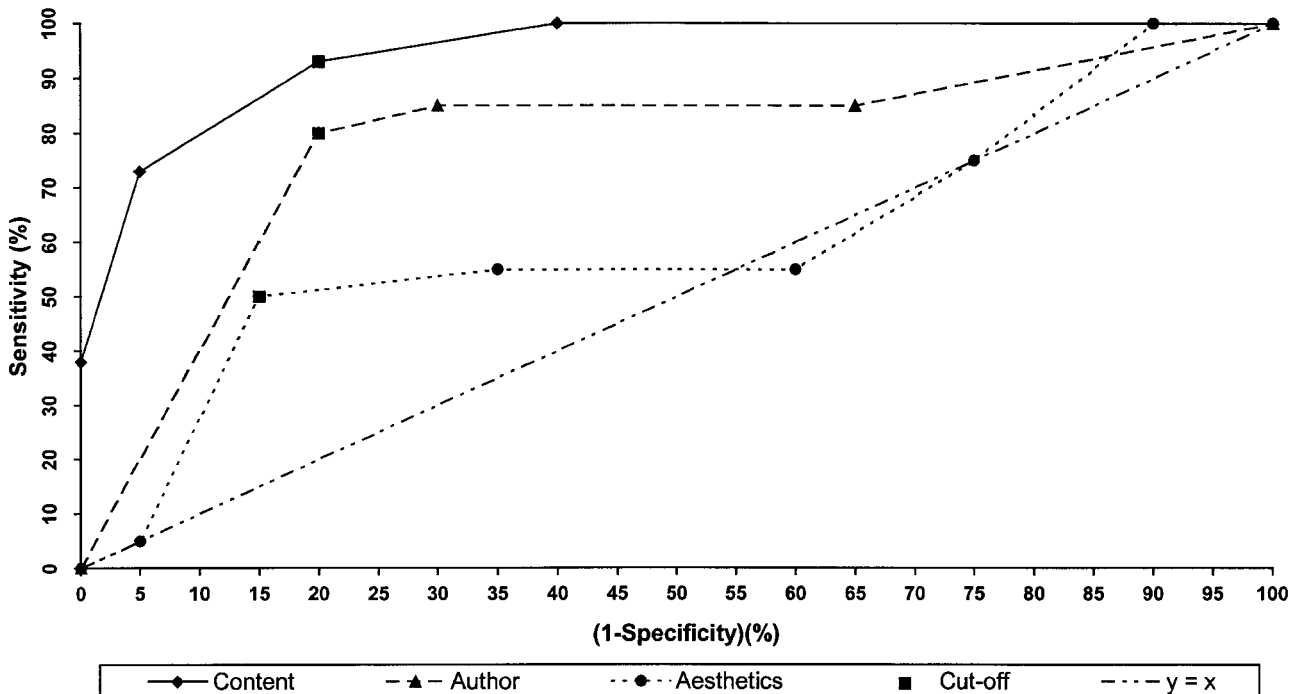


Figure 1 ROC curves for content, authorship and aesthetics and *y = x* line representing a non-discriminatory test for comparison.

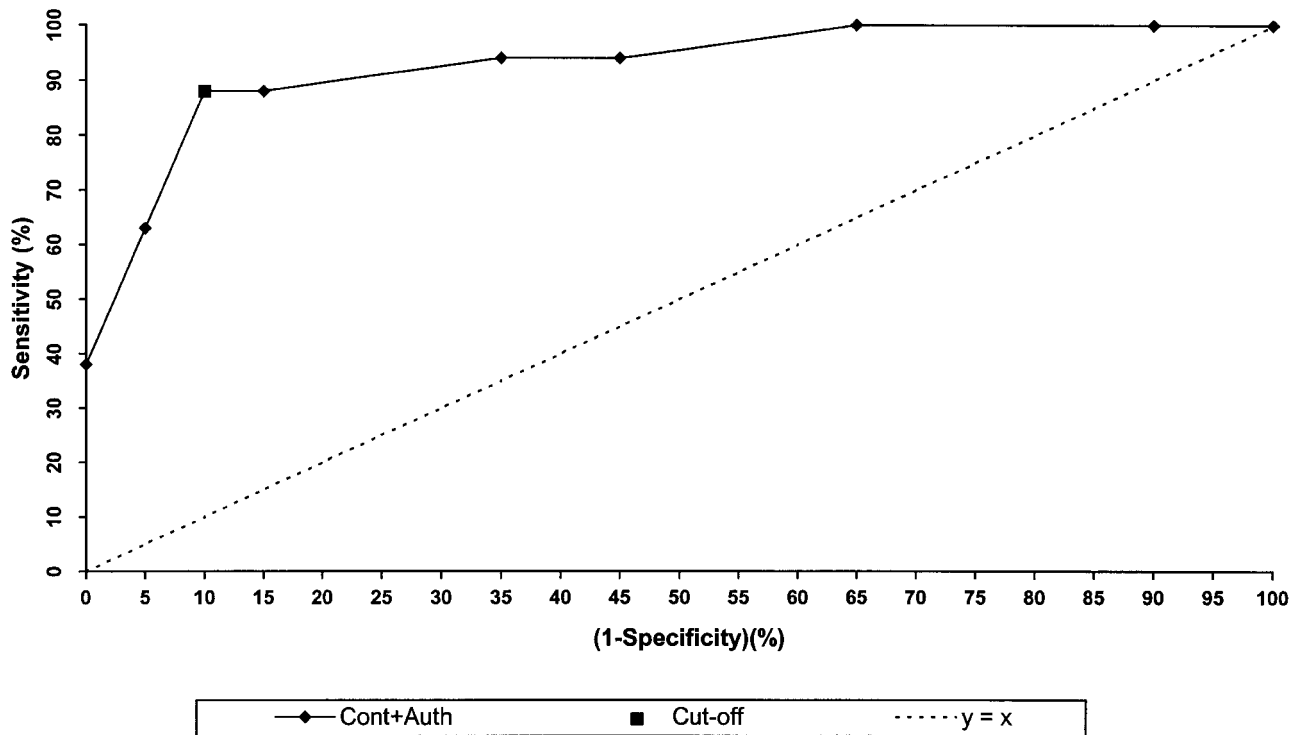


Figure 2 ROC curve for content + author score with $y = x$ line for comparison.

Like other researchers, I found instances of pages giving inaccurate medical information. One page stated that MMR vaccine caused type I diabetes.³⁵ There is no evidence that this is so. Many pages made statements suggesting that vaccines do not work at all or even increase the likelihood of becoming infected by the disease.^{36–39}

From the proforma

For each of the three sections (content, authorship and aesthetics) the respective sensitivities and specificities for all possible scores were tabulated and sensitivity was plotted against one minus specificity. The resultant ROC curves are shown in Fig. 1.

This figure shows that the contents score is a very good discriminator between 'good' and 'bad' pages, with an area under the graph of 95 per cent of the total [95 per cent confidence interval (CI) 90 per cent, 100 per cent]. The best cut-off value for the content score is two. Scoring greater than two indicates a 'good' page with a sensitivity of 94 per cent and specificity of 79 per cent. The authorship score had lower discriminant ability with an area under the graph of 79.7 per cent (95 per cent CI 65 per cent, 95 per cent). Here again the best cut-off is two, with a sensitivity of 81 per cent and a specificity of 79 per cent. Aesthetics is a poor discriminator, the area under the graph being 60.4 per cent (95 per cent CI 41.3 per cent, 80.6 per cent). This is not statistically significantly different from 50 per cent, the value for a test with no discriminatory ability.

As both section 1 (contents) and section 2 (authorship) seemed, independently, to be good discriminators their results were added together and the results used to generate a new ROC curve (Fig. 2).

In terms of ability to discriminate, the combined scores were only slightly less good than content alone, with an area under the graph of 93 per cent (95 per cent CI 85 per cent, 100 per cent). The best cut-off value was five with sensitivity of 88 per cent and specificity of 92 per cent.

Relative risk

Interestingly, in the sample studied, the relative risk of being an inaccurate or unbalanced page, if the page took an anti-vaccination stance, was 3.3 (95 per cent CI 1.8, 6.1) (Table 2). Being anti-vaccination seemed, in this sample, to be a useful marker for inaccurate or unbalanced information.

Discussion

In line with other researchers' findings,^{1–6} the amount of medical misinformation in the sample is worrying, with 60 per cent of pages giving inaccurate or unbalanced information on MMR vaccine in my sample. If this applies generally then the scope for being medically misinformed is considerable.

Although the area under the graph for the content score alone is larger than using the combined content + authorship score, this was not statistically significant. The content score has better sensitivity and worse specificity than the combined

Table 2 Data for the relative risk (RR) calculation for being a poor quality page if also being anti-vaccination

	Number of pages inaccurate or unbalanced	Number of pages accurate and balanced	Total
Number anti-vaccination	17	0	17
Number pro- or neutral	7	16	23
Total	24	16	40

RR = 3.3 (95 per cent CI 1.8, 6.1).

score. Using the combined score trades a slight decrease in sensitivity for a larger increase in specificity. In my opinion, a high specificity is more important, as it is vital to correctly classify inaccurate pages at the expense of misclassifying some accurate ones. As there are so many pages to choose from, the content + authorship score appears better. Measuring the various aesthetic indices is clearly not a good way of discriminating and should not be used when judging pages. Also, the number of links is not a fair way of judging pages. How many links are followed is a function of the search engines' indexing strategy.

Because of the need to have a knowledge of the literature about the subject under investigation to use the proforma properly, its usefulness as a general, objective tool for the non-expert is a little limited.

A new and unexpected finding in this sample is that a page is over three times more likely to be poor quality if it is anti-vaccination.

Just having references, even to peer-reviewed journal papers, does not necessarily correlate with quality. This conclusion was drawn from two observations. First, only 15 per cent of the sample had any references. Few were to peer-reviewed journals. Second, at least one page³⁸ was fully referenced, citing many papers from peer-reviewed journals. Those references, however, either did not support the author's claims or they had been selectively quoted out of context, thus giving a completely misleading view of the reference's message. As a consequence of this, I did not use whether a page was referenced or not as part of the gold standard criteria.

Implications and recommendations for policy and public health

No UK site was discovered in the first 80 pages from the Lycos search and only one⁴⁰ (at number 169) in the first 200 pages from the AltaVista search. This was a press release from the Northern Ireland Office. There were no pages found from either search from a UK medical or public health site. It is hardly surprising, therefore, that the public may be misled, as the UK professionals who can speak authoritatively on a subject and correct misinformation are just not on the Net at the right time. Those with the professional public health and medical expertise must start to have a greater presence on the Web and respond

rapidly (possibly within 24 h) and critically to important medical issues in clear, readable, non-technical language. This should include both national and local issues.

Conclusion

The Web does contain some accurate balanced information on MMR vaccine; it also contains a lot of misleading inaccurate information. The proforma may help to distinguish between the two but perhaps only if the user has a reasonable knowledge of the subject in the first place. Thus, the WWW should not be considered as a reliable source of information on subjects about which one knows little. This is especially true for medical information, as, if my sample is representative, much of what a typical user may find will be inaccurate or biased.

Acknowledgements

This study was conducted as part of an MPH degree at the Nuffield Institute for Health, at the University of Leeds. I would especially like to thank Dr Roberta Marshall and Dr Viv Peiris, Consultants in Communicable Disease Control, and Kevin McKeown for their valuable advice and help at critical stages in the project.

References

- 1 Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling and assuring the quality of medical information on the internet. Caveant lector et viewer – let the reader and viewer beware. *JAMA* 1997; **277**(15): 1244–1245.
- 2 Bower H. Internet sees growth of unverified health claims. *Br Med J* 1996; **313**: 381.
- 3 Impicciatore P, Pandolfi 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. *Br Med J* 1997; **314**: 1875–1879.
- 4 Lindsay B. Information on the Internet: consumer guides to health care. *Br J Hlth Care Manage* 1996; **2**(8): 445–447.
- 5 Marra CA, Carleton BC, Lynd LD, *et al.* Drug and poison information resource on the Internet, part 2: identification and evaluation. *Pharmacotherapy* 1996; **16**(5): 806–818.

- 6 Wyatt JC. Measuring quality and impact of the World Wide Web. *Br Med J* 1997; **314**: 1879–1881.
- 7 Georgia Tech Research Corporation. GVU's 8th WWW user survey. Accessed 4 July 1998. http://www.cc.gatech.edu/gvu/user_surveys/
- 8 Anon. The web of information inequality (editorial). *Lancet* 1997; **349**: 1781.
- 9 Selway J. Medical practitioners need to give more than reassurance (letters). *Br Med J* 1998; **316**: 1824.
- 10 Anon. MMR vaccine coverage falls after adverse publicity. *Comm Dis Rep CDR Wkly* 1998; **8**: 5.
- 11 Gangerosa EJ, Galazka AM, Wolfe CR, *et al.* Impact of anti-vaccine movements on pertussis control: the untold story. *Lancet* 1998; **351**: 356–361.
- 12 Pealer LN, Dorman SM. Evaluating health-related Web sites. *J School Hlth* 1997; **67**(6): 232–235.
- 13 Scholz A. Evaluating World Wide Web information. Accessed 23 March 1998. <http://thorplus.lib.purdue.edu/research/classes/gs175/3gs175/evaluation.html>
- 14 Jadad AR, Gagliardi A. Rating health information on the Internet: navigating to knowledge or to Babel. *JAMA* 1998; **279**(8): 611–614.
- 15 Health on the Net Foundation. Health on the Net foundation code of conduct for medical Web sites. Accessed 3 December 1997. http://www.hon.ch/HONcode/conduct_t.html
- 16 Day A. A model for monitoring Web site effectiveness. *Internet Res Electron Networking Appl Policy* 1997; **7**(2): 109–115.
- 17 British Healthcare Internet Association. Quality standards for medical publishing on the web. Accessed 31 March 1999. http://www.bhia.org/reference/documents/recommend_webquality.htm
- 18 Health Information Technology Institute, Mitretek Systems. Criteria for assessing the quality of health information on the internet. Accessed 31 March 1999. <http://www.mitretek.org/hiti/showcase/documents/criteria.html>
- 19 Kim P, Eng TR, Deering MJ, Maxfield A. Published criteria for evaluating health related web sites: review. *Br Med J* 1999; **318**: 647–649.
- 20 AltaVista. Accessed April 1998. <http://www.altavista.digital.com/>
- 21 Lycos. Accessed April 1998. <http://www.lycos.co.uk/>
- 22 Branscomb HE. *Casting your net. A student's guide to research on the Internet.* Boston, MA: Allyn and Bacon, 1998.
- 23 Kiley R. *Medical information on the Internet: a guide for health professionals.* London: Churchill Livingstone, 1996.
- 24 Yahoo! Accessed April 1998. <http://www.yahoo.com/>
- 25 Wakefield AJ, Murch SH, Anthony A, *et al.* Ileal–lymphoid–nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *Lancet* 1998; **351**: 637–641.
- 26 Pallen M. Guide to the internet: the world wide web. *Br Med J* 1995; **311**: 1552–1556.
- 27 Salisbury DM, Begg NT. *Immunisation against infectious disease. Edward Jenner bicentenary edition.* London: HMSO, 1996.
- 28 Beck JR, Shultz EK. The use of relative operating characteristic (ROC) curves in test performance evaluation. *Arch Pathol Lab Med* 1986; **110**: 13–20.
- 29 Goodenough DJ, Rossmann K, Lusted LB. Radiographic applications of receiver operating characteristic (ROC) curves. *Radiology* 1974; **110**: 89–95.
- 30 Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983; **148**: 839–843.
- 31 Shultz EK. Multivariate receiver-operating characteristic curve analysis: prostate cancer screening as an example. *Clin Chem* 1995; **41**(8B): 1248–1255.
- 32 Swets JA. ROC analysis applied to the evaluation of medical imaging techniques. *Invest Radiol* 1979; **14**(2): 109–121.
- 33 Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental tool in clinical medicine. *Clin Chem* 1993; **39**(4): 561–577.
- 34 Bowling A. *Research methods in health: investigating health and health services.* Buckingham: Open University Press, 1997.
- 35 Anon. Case reports to NVIC of diabetes following MMR shots. Accessed 23 April 1998. <http://www.909shot.com/dicasee.htm>
- 36 Read M. The vaccine controversy. Accessed 23 April 1998. <http://www.marysherbs.com/vaccine.htm>
- 37 Neustaedter R. About vaccine choices. Accessed 23 April 1998. <http://www.healthy.net/library/articles/neustaedter/aboutvac.htm>
- 38 International Chiropractic Pediatric Association. Vaccinations. Accessed 23 April 1998. <http://www.4icpa.org/Vaccinations/mmr.htm>
- 39 Anon. Welcome to PAVE (People Advocating Vaccine Education). Accessed 23 April 1998. <http://www.unidial.com/~metroprint/pave2.html>
- 40 Northern Ireland information service. Accessed 30 April 1998. <http://alexandra14.nio.gov.uk/press.htm>

Accepted on 1 October 1999

Appendix: health-related Web site evaluation form

Title:

Author/body:

URL:

1.0	Content	Y	N	
1.1	Content accurately reflects the most recent research			
1.2	Content is balanced			
1.3	Are all claims referenced/all quotes attributable			
1.3.1	If references provided are there functional HyperText links to them			
1.3.2	If references, are they from quality peer-reviewed journals/quotes from acknowledged authorities/bodies			
1.4	Reading level: Flesch Reading Ease Score (FRES) (100 = very readable)			
1.5	Date page created/info posted/updated is stated			
1.6	Overall content rating: (total)			
2.0	Authorship	Y	N	
2.1	Author/body responsible for publishing/for original information (if quoting) is identified			
2.2	Author/body of page/original information is recognized authority			
2.2.1	If not then the author's/body's credentials are listed			
2.3	Author/body responsible for page/original information is qualified to publish this document/news report/original information			
2.4	Overall authorship rating: (total)			
3.0	Page aesthetics	Y	N	
3.1	Document has a distinguishable header, body, footer			
3.2	Major headings and subheadings easily identifiable			
3.3	Additional relevant links			
3.4	Use of relevant graphics			
3.5	Minimal page layering (no links followed)			
3.6	Page does not contain outside advertisements			
3.6.1	If it does, then they are not placed within the body of the text			
3.7	Other features present: e.g. sound, frames, text-only, interactive, etc.			
3.8	e-mail/contact details of author/publisher/body is/are given			
3.9	other contact information for more information are given			
3.10	Overall aesthetics rating: (total)			
4.0	Page is (P)ro (A)nti or (N)eutral/unknown re MMR vaccine	P	A	N
5.0	Comments:			

Modified from Pealer and Dorman.¹²