**OBJECTIVE** Most individuals would like to maintain the privacy of their medical information on the World Wide Web (WWW). In response, commercial interests and other sites post privacy policies that are designed to inform users of how their information will be used. However, it is not known if these statements are comprehensible to most WWW users. The purpose of this study was to determine the reading level of privacy statements on Internet health Web sites and to determine whether these statements can inform users of their rights.

**STUDY DESIGN** This was a descriptive study. Eighty Internet health sites were examined and the readability of their privacy policies was determined. The selected sample included the top 25 Internet health sites as well as other sites that a user might encounter while researching a common problem such as high blood pressure. Sixty percent of the sites were commercial (.com), 17.5% were organizations (.org), 8.8% were from the United Kingdom (.uk), 3.8% were United States governmental (.gov), and 2.5% were educational (.edu).

**OUTCOMES MEASURED** The readability level of the privacy policies was calculated using the Flesch, the Fry, and the SMOG readability levels.

**RESULTS** Of the 80 Internet health Web sites studied, 30% (including 23% of the commercial Web sites) had no privacy policy posted. The average readability level of the remaining sites required 2 years of college level education to comprehend, and no Web site had a privacy policy that was comprehensible by most English-speaking individuals in the United States.

**CONCLUSIONS** The privacy policies of health Web sites are not easily understood by most individuals in the United States and do not serve to inform users of their rights. Possible remedies include rewriting policies to make them comprehensible and protecting online health information by using legal statutes or standardized insignias indicating compliance with a set of privacy standards (eg, “Health on the Net” [HON] http://www.hon.ch).

**KEY WORDS** Readability levels; Internet; privacy.
These guidelines are designed to “protect the privacy of individually identifiable health information.” Although Internet use does not generate a formal medical record, online profiling allows the collection of detailed medical information about a user’s diagnoses, medications, etc, which essentially creates “individually identifiable health information” when associated with their names.

One proposed solution to maintaining Internet privacy has been the voluntary posting of privacy statements. These statements serve to inform users of the privacy policies of the Web site, such as what user information is collected and with whom this information will be shared. Three recent studies have shown that the readability level of much of the patient information on the Internet is beyond the comprehension of many individuals in the United States. For voluntary privacy statements to be useful, they need to be written at a level understood by most individuals using the Internet. The purpose of this study was to determine the readability level of privacy statements on Internet health Web sites.

METHODS
A total of 80 Internet health Web sites were examined in May 2001 to see if they included privacy policies. To emulate the way a consumer might find information on the Internet, 55 of the sites were selected by entering search terms for common conditions into a widely used Internet search engine (http://www.google.com). The terms searched for were “high blood pressure,” “fever,” “cough,” and “wellness.” The other 25 Web sites analyzed represent the most commonly visited health information Web sites on the Internet. For Web sites identified by Google, the Web pages represented by the top 10 results for each term were viewed, and any links on those pages were followed until 55 health information Web sites were identified. We did not limit the Web sites to only those identified by the search engines because in many cases, users will follow the links on a page identified by the search engine. Links that led to medical school lectures, nonhuman diseases, online journal articles aimed at health care professionals, or the contents of e-mail or listserv summaries were omitted.

The Web sites were then examined by 1 researcher (J.J.W.) to see if they included privacy statements. This review included looking at the page that the link led to, the Web site’s homepage, any Web site policies, “about us” type of material, etc. Links to privacy foundations such as Health on the Net were considered to represent privacy statements. If a privacy statement was found, the text was copied into Microsoft Word 98 for Macintosh (Microsoft Corporation, Redmond, WA) and the documents’ Flesch Reading Ease score was generated using the built-in software in Microsoft Word 98 (Macintosh). The same privacy statements were printed and then analyzed using 2 hand-calculation methods: the Fry Readability formula and the SMOG method (simplified measure of gobbledygook). If a Web site was found to have no privacy statement, this was confirmed by a second researcher (M.G.). Web sites on which no privacy policy was found after 2 searches were not contacted to determine if a policy exists. It is not likely that end users would contact a site to determine the existence of a privacy policy. Thus, there would be the de facto absence of a privacy policy with respect to the end user. Ten percent of the readability levels were analyzed by a second researcher (D.M.D.) and found to have accurate coding. All data were entered into Microsoft Excel spreadsheet and analyzed using the built-in statistical formulas. We did not seek to evaluate the content of the privacy statements themselves.

Approximately 40 different readability formulas are available for use, all of which will give a reasonably accurate grade level (generally plus or minus 1 grade level with 68% confidence: an acceptable standard in the field). Most of these formulas rely on number of syllables in a word and sentence length to judge readability. In general, readability is easier when there are a low number of words per sen-

<table>
<thead>
<tr>
<th>Top level domain</th>
<th>All Web sites examined (n = 80)</th>
<th>Web sites with a privacy policy (n = 54)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commercial (.com)</td>
<td>48 (60%)</td>
<td>37 (68.5%)</td>
</tr>
<tr>
<td>Organizations (.org)</td>
<td>14 (17.5%)</td>
<td>8 (14.8%)</td>
</tr>
<tr>
<td>United Kingdom (.uk)</td>
<td>7 (8.8%)</td>
<td>5 (9.2%)</td>
</tr>
<tr>
<td>Government (.gov)</td>
<td>3 (3.8%)</td>
<td>5 (9.2%)</td>
</tr>
<tr>
<td>Network (.net)</td>
<td>2 (2.5%)</td>
<td>0 (0.0%)</td>
</tr>
<tr>
<td>Educational (.edu)</td>
<td>2 (2.5%)</td>
<td>1 (1.8%)</td>
</tr>
<tr>
<td>Web site did not exist*</td>
<td>2 (2.5%)</td>
<td>0 (0.0%)</td>
</tr>
<tr>
<td>South Africa (.za)</td>
<td>1 (1.2%)</td>
<td>1 (1.8%)</td>
</tr>
<tr>
<td>Numerical (no domain listed)</td>
<td>1 (1.2%)</td>
<td>0 (0.0%)</td>
</tr>
<tr>
<td>Total</td>
<td>80 (100%)</td>
<td>54 (100%)</td>
</tr>
</tbody>
</table>

*Although the list of the top 25 health-related WWW sites was current, 2 of the sites had gone out of business and were no longer available.
tence, characters per word, and syllables per word, and a low percentage of passive sentences. Illustrations and tables also improve the readability. The Flesch Reading Ease score is one of the most widely used and validated systems for scoring readability. It is the standard used by the insurance industry for consumer documents and contracts. Documents scoring 70 or above are described as “easy” and are written at the grade school level. A score of 60 to 70 is described as “standard” and written at approximately the high school level. Scores below 60 are described as “fairly difficult,” “difficult,” or “very difficult” as the score decreases.

The Fry formula is a hand-calculation method that is recommended by experts in the field for use with patient education materials. The SMOG is another hand-calculation method commonly used for evaluation of health information. Results from the Fry formula and SMOG methods are expressed as standard United States grade levels.

RESULTS
Of the 80 health Web sites in the initial sample, 2 had recently closed down and an additional 24 (30%) (including 23% of the commercial Web sites) did not have a privacy statement. Thus, the privacy statements of 54 Web sites were analyzed using the 3 readability formulas. The Web site domains for the 80 Web sites as well as the 54 with privacy statements are shown in the Table. The average Flesch Reading Ease score of privacy statements was 39 or “difficult.” This level is similar to reading a corporate annual report. The Flesch Reading Ease score range was 24.4 to 54.2. This range is described as “very difficult” (eg, similar to reading legal contracts) to “fairly difficult” (eg, similar to reading novels such as the Henry James novel, The Ambassadors). The Fry formula had a readability level equal to 14.6 or 14th grade (range, 10–17). The overall SMOG readability level was 14.7 (range, 11–19).

DISCUSSION
The goal of this study was to determine the readability level of privacy statements on Internet health Web sites. Privacy statements are meant to protect the individual, a particularly important endeavor when dealing with medical information because of its potentially sensitive nature. Only 70% of the health Web sites examined in this study even contained a privacy statement. The readability levels of the privacy statements found were considered difficult and would require approximately 2 years of university training to be comprehensible. In contrast, only 60% of diabetic patients in 1 study could understand information written at the 6th grade reading level. Most studies find that only a minority of patients can comprehend information written at a 9th grade reading level. Overall, for 90% of adults to be able to read and comprehend written materials, the materials should be written at less than the 8th grade reading level. Although current Internet users may be a more educated group than the general population, this trend is changing. A recent study found that 21% of individuals with less than a high school education have Internet access, as do 43% of high school graduates. Because an individual's actual reading level is usually 2 to 5 grades below the grade completed, many of these individuals will not be able to comprehend a Web site's privacy policy and thus may not be protected or understand their options pertaining to protection.

Many patients are beginning to use the Internet as a source of information and many are unaware of their every move is being tracked. Physicians can help to protect patients by making them aware of this problem and suggesting that patients take steps to protect their information.

A number of possible solutions can be applied to this problem. Rewriting privacy policies to be comprehensible to most Internet users should be the initial step. Resources are available to Web site developers that help to maximize the readability of information. More widespread use of the “Health on the Net” (HON, http://www.hon.ch) privacy criteria would reassure individuals that their data are secure. Consideration should also be given to extending the same protection to medical information gleaned from the Internet by Web sites and advertising companies as is given to the formal medical record because, in effect, they can reflect the same type of information. For example, on several sites, users can enter medications they are taking and diagnoses (http://www.drkoop.com and others).

The first limitation of this study is that the Web sites studied represent a minority of the health Web sites on the Internet. However, the study set includes the top 25 visited health Web sites and did seek many different types of Web sites for evaluation. Also, none of the Web sites had a policy that was written at a level comprehensible to most of the English-speaking people in the United States. It is unlikely that including more Web sites would substantially change the outcome. Finally, given the large (and unknown) number of health Web sites on the Internet, it is not practical to obtain a large enough sample to be representative of all health Web sites.

The second problem is readability levels themselves. Readability calculations have been criticized because they depend mainly on sentence length and...
the number of syllables per word. If an individual is familiar with the sentence terminology, it may be possible to interpret the information correctly using context clues. However, most individuals do not have a working knowledge of legal terminology (e.g., “indemnify”) and thus may not be able to use these clues to determine the meaning of a sentence.

**CONCLUSIONS**

Thirty percent of health Web sites do not have privacy statements, and those that are posted are beyond the reading level of most adults. For this reason, current privacy statements do not function to adequately inform users of a Web site’s privacy policy and do not protect users’ privacy rights.

**REFERENCES**