Readability, Relevance and Quality of the Information in Spanish on the Web for Patients With Rheumatoid Arthritis

Jose Dionisio Castillo-Ortiz, a, Jose de Jesus Valdivia-Nuno, a Andrea Ramirez-Gomez, a Heber Garagarza-Mariscal, a Carlos Gallegos-Rios, a Gabriel Flores-Hernandez, a Luis Hernandez-Sanchez, a Victor Brambila-Barba, a Jose Juan Castaneda-Sanchez, a Zalathiel Barajas-Ochoa, a Angel Suarez-Rico, a Jorge Manuel Sanchez-Gonzalez, b Cesar Ramos-Remus a, b

a Unidad de Investigación en Enfermedades Crónicas-Degenerativas, Guadalajara, Mexico
b Vicerrectoría Académica, Universidad Autónoma de Guadalajara, Zapopan, Mexico

A R T I C L E   I N F O
Article history:
Received 8 March 2016
Accepted 22 April 2016
Available online 12 April 2017

Keywords:
Rheumatoid arthritis
Internet
Web
Education

A B S T R A C T
Background: Education is a major health determinant and one of the main independent outcome predictors in rheumatoid arthritis (RA). The use of the Internet by patients has grown exponentially in the last decade.
Objective: To assess the characteristics, legibility and quality of the information available in Spanish in the Internet regarding to rheumatoid arthritis.
Material and methods: The search was performed in Google using the phrase rheumatoid arthritis. Information from the first 30 pages was evaluated according to a pre-established format (relevance, scope, authorship, type of publication and financial objective). The quality and legibility of the pages were assessed using two validated tools, DISCERN and INFLEZ respectively. Data extraction was performed by senior medical students and evaluation was achieved by consensus.
Results: The Google search returned 323 hits but only 63% were considered relevant; 80% of them were information sites (71% discussed exclusively RA, 44% conventional treatment and 12% alternative therapies) and 12.5% had a primary financial interest. 60% of the sites were created by nonprofit organizations and 15% by medical associations. Web sites posted by medical institutions from the United States of America were better positioned in Spanish (Arthritis Foundation 4th position and American College of Rheumatology 10th position) than web sites posted by Spanish speaking countries.
Conclusions: There is a risk of disinformation for patients with RA that use the Internet. We identified a window of opportunity for rheumatology medical institutions from Spanish-speaking countries to have a more prominent societal involvement in the education of their patients with RA.

© 2016 Elsevier España, S.L.U. and Sociedad Española de Reumatología y Colegio Mexicano de Reumatología. All rights reserved.

Legibilidad, relevancia y calidad de la información en español en la Web para pacientes con artritis reumatoide

P R E S U M E N
La educación es un determinant mayor de salud y uno de los predictores independientes de desenlace en artritis reumatoide (AR). El uso del Internet por pacientes ha crecido en forma exponencial en la última década.
Objetivos: Evaluar las características, legibilidad y calidad de la información disponible en Internet en idioma español en relación con la AR.
Material y métodos: Se buscó la frase AR en Google. Se evaluaron las primeras 30 páginas de resultados de acuerdo con un formato diseñado ex profeso (relevancia, autoría, tipo de publicación, enfermedad

Palabras clave:
Artritis reumatoide
Internet
Web
Educación

© 2016 Elsevier España, S.L.U. and Sociedad Española de Reumatología y Colegio Mexicano de Reumatología. All rights reserved.

E-mail address: joseedco006@hotmail.com (J.D. Castillo-Ortiz).

2173-5743/© 2016 Elsevier España, S.L.U. and Sociedad Española de Reumatología y Colegio Mexicano de Reumatología. All rights reserved.
Introduction

Rheumatoid arthritis (RA) is a disease that can be prolonged for decades and has a multidimensional impact, ranging from joint destruction to the development of comorbidities and high direct and indirect costs.\(^1,2\) Formal education of RA patients is of the utmost importance, from the perspective of determinants of health\(^3\) or as a predictor of the outcome.\(^4\) Information is the platform for acquiring knowledge that will be transformed into behavior that will have an influence on the form with which a person faces disease, makes choices concerning caring for his or her health and adheres to treatment.\(^5,6\)

A physician provides the scientific information that a patient may need by means of a process of filtering and dosing.\(^7\) At the present time, what a patient requires of the physician is greater attention, an improvement in communication, sufficient time during visits and not to have to wait too much to be attended to.\(^8\) However, these requirements are frequently not met.\(^9,10\) Although physicians continue to be the primary source of information, the Internet has become an important option for patients in recent years, at least in countries like the United States of America (US). For example, 81% of the adults in the US utilize the Internet and 72% mentioned that they had looked for health information over the past 12 months.\(^11\) In fact, users of the Internet also visit websites that are specific for health professionals; of the millions of searches carried out each year in the web page of the National Library of Medicine, 33% are performed by the general public.\(^12\)

There are more than 40 publications concerning the quality of the information available on the web related to different diseases and at least 15 on RA; all are in English and analyze websites in English. The conclusions in most of these publications focus on whether the information on the Internet can increment the autonomy of patients, but it can also increase their anxiety and be the source of misleading information due to the heterogeneity and low quality of the information;\(^13,14\) moreover, patients can become the victims of persons with commercial interests.\(^15,16\)

However, to the best of our knowledge, there are no publications analyzing the characteristics of the information available in the web about RA in Spanish. This is important if we take into account the fact that Spanish is the second language in numbers of first-language speakers in the world\(^17\) and that the utilization of the Internet in Spanish has also increased; it is used by 88.3% of Spaniards\(^18\) and by 50.2% of Mexicans,\(^19\) which add up to approximately 103.8 million users. If we consider it true that the prevalence of RA is between 0.5% and 1% of the general population, then there would be between 518 thousand and 1 million RA patients as potential users of the Internet only in those 2 countries. This, and because of the chronicity of RA, it can be assumed that there is a large number of potential users of the Internet who speak Spanish as their first language and search for information on their disease. In the present study, we analyzed the quality of the contents, the legibility and the relevance of the websites with information on RA in Spanish.

Material and Methods

This is a cross-sectional study to audit web pages in the Internet to assess the characteristics of the information related to RA in Spanish. We used the methods of a previous publication from Canada because the detailed description of their methodology enabled us to reproduce it.\(^20\) Briefly, we did an extensive search in Spanish, without filters or restrictions, using the phrase “artritis reumatoide” via the Google search engine, as it is the engine most widely used in the world.\(^21\) The search was conducted on a single day and the resulting web pages were stored in Portable Document Format (PDF) for later assessment. All of the pages were evaluated after being visited and the information had been collected from their associated sites. The web pages from the first 30 pages of results were identified and accessed (hits) (10 hits per results page). Two sixth-year medical students evaluated each web page separately; discrepancies and/or controversies were resolved by consensus, together with 2 of the participating rheumatologists (DCO and CRR). We considered a website to be a collection of several related web pages that had a common Internet domain. The evaluation of the pages was performed in accordance with the following domains: relevance, type of publication, scope, contents, authorship and financial objectives, in terms of the definitions provided\(^22,23\):

a. Relevance. Web pages were considered relevant if the content included some clinical aspect of RA—etiology, symptoms, signs, diagnosis, treatment and/or prognosis. If the information did not deal with a clinical aspect but could influence the conduct of the patient with respect to the disease, it was considered of doubtful relevance. Only the relevant or those of doubtful relevance continued in the process of evaluation.

b. Type of publication. The web pages were classified as advertisements, information, links, new articles, part of a publication, attempts to recruit patients for an investigation, chat rooms, case reports and support group posts.

c. Scope. We established whether the web pages dealt only with aspects of RA or discussed other autoimmune diseases, and used “arthritis” as a general term or in reference to several diseases (fibromyalgia, juvenile idiopathic arthritis, scleroderma, etc.).

d. Contents. The web pages were categorized according to their contents: the inclusion of general information (clinical or epidemiological data on RA) or the mention of conventional therapy or some type of alternative therapy.

e. Authorship. The information from the websites was classified as being from a lucrative industry or a for-profit organization.
a nonprofit organization, individuals with no clear affiliation and medical institutions (universities, hospitals, medical associations).

f. Financial objectives. The websites were cataloged depending on whether they contained explicit information on products being marketed, requested donations or required the payment of a membership fee to access. They were classified as being direct or indirect product marketing, as requiring a membership fee, as requesting donations or none of the above.

Aside from the procedures described in the reference study,20 others were added to assess the “content quality” domains of the web pages, the legibility and possible variations in the results depending on the Internet Protocol (IP) address utilized. The content quality was evaluated using the DISCERN tool.22 DISCERN is an instrument designed to rate the quality of the content of a publication on the Internet, but cannot be utilized to evaluate the scientific quality or accuracy of the evidence on which the publication is based. DISCERN consists of 15 specific questions plus a question that measures the general quality. Each of the first 15 questions represents in itself a specific quality criterion; the rating is 1 (if the response is “no”), 2–4 (up to a certain point) or 5 (if the response is a definite “yes”). As our study did not evaluate information on specific treatments and each of the 15 questions represents a separate quality criterion, questions referring to treatments were excluded and we utilized only the following items: 1 (The objectives are clear?); 2 (Those objectives are achieved?); 4 (Is there a clear declaration of any sources of information used to compile the data, aside from the author or the product?); 6 (Is it objective and non-biased?); and 7 (Is data provided on additional sources of support and information?); question 16 was also included (an intuitive summary of the responses to the 15 previous questions/a subjective evaluation of the general quality).

Legibility was evaluated using the INFLESZ program, which measures the syntactic complexity of a written text, whether it be a complete file or a fragment of a work.21 The score obtained with the INFLESZ scale ranges from 0 to 100. A text is considered to be very difficult to read if the score is from 0 to 40 (scientific or university publications), somewhat difficult to read with a score of 40.1 to 55 (high school level, specialized journals, etc.); normal if the score is from 55.1 to 65 (newspapers or sports news), quite easy with a score of 65.1 to 80 (grade school level) and very easy if the score is over 80.1 (grade school level, comic books, novels, etc.). For the analysis of legibility, we excluded figures, tables, footnotes, links and other pages and references.

We analyzed web pages published by hospitals, institutions and societies or medical associations (medical institutions) separately, using the procedures described above.

To evaluate the possible variation in the results depending on the IP address, we did the same search utilizing 3 different IP addresses: from Guadalajara, Mexico (189.169.108.15), from San Andrés, Colombia (104.224.35.115) and from Madrid, Spain (185.25.87.186).24

### Statistical Analysis

Dichotomized variables were expressed as proportions. Chi-squared test was used to compare the general characteristics, quality and legibility of the web pages in each group according to the order in which they appeared, taking the pages from medical associations, societies and institutions together with the rest of the relevant pages. A P value <.05 was considered to indicate statistical significance. The analysis was done with the STATA (version 12.0) software package.

### Results

We found 323 web pages, 121 (38%) of which were not evaluated (10 because they were in another language and 111 because we considered them irrelevant). Table 1 shows the type of publication, the characteristics of the disease being dealt with and the content of the 202 relevant web pages. In all, 80% of the pages provided general information exclusively on RA; 13% of the pages included an alternative therapy, either as the main topic or in association with conventional therapy. We identified 186 websites, nearly all of which had been created by nonprofit institutions, although only 15% had been posted by a medical institution (Table 2).

The evaluation of the content quality showed that 61% of the web pages clearly declared their objectives, but only 48% fulfilled them. Around a third of the pages clearly provided the sources of the information utilized, only 22% offered unbiased information and more than half failed to name additional sources of information. The general quality of a fourth of the pages was considered to be high (Table 3). With respect to the legibility, more than half of the pages were very difficult (9%) or somewhat difficult (52%) to read, whereas a fourth of the pages had a normal level of legibility in accordance with the INFLESZ SCALE (Table 4).

Table 5 shows the position of the web pages in Spanish belonging to medical institutions, according to the place in which they appeared, the content quality (DISCERN) and legibility (INFLESZ). Among the first 300 pages that appeared in Google, only 26 (8.6%) were created by medical institutions: 17 (65%) from the US, 7 (27%) from Spain, one from Mexico and one from Venezuela. Among the first 10 websites, 3 were from medical institutions; they included the Arthritis Foundation (US) in 4th place, the Spanish Society of...
Table 3
Evaluation of the Quality of the 202 Web Pages.

<table>
<thead>
<tr>
<th>Quality as defined by the question</th>
<th>No n (%)</th>
<th>Partially n (%)</th>
<th>Yes n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The objectives are clear?</td>
<td>59 (29)</td>
<td>20 (10)</td>
<td>123 (61)</td>
</tr>
<tr>
<td>2. Are the objectives fulfilled?</td>
<td>65 (32)</td>
<td>41 (20)</td>
<td>96 (48)</td>
</tr>
<tr>
<td>4. Is there a clear declaration of any sources of information used for the publication, aside from the author?</td>
<td>114 (57)</td>
<td>17 (8)</td>
<td>71 (35)</td>
</tr>
<tr>
<td>6. The information is clear and unbiased?</td>
<td>102 (50)</td>
<td>56 (28)</td>
<td>44 (22)</td>
</tr>
<tr>
<td>7. Is data provided on additional sources of support and information?</td>
<td>116 (57)</td>
<td>14 (7)</td>
<td>72 (36)</td>
</tr>
<tr>
<td>General quality of the publication</td>
<td>Low</td>
<td>Moderate</td>
<td>High</td>
</tr>
<tr>
<td>16. Based on all of the responses, what is the general quality of the publication as a source of information concerning the choice of treatment?</td>
<td>67 (33)</td>
<td>85 (42)</td>
<td>50 (25)</td>
</tr>
</tbody>
</table>

Table 4
Legibility of the 202 Relevant Web Pages (INFLESZ).

<table>
<thead>
<tr>
<th>INFLESZ scale</th>
<th>Web pages n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–40</td>
<td>18 (9)</td>
</tr>
<tr>
<td>40.1–55</td>
<td>105 (52)</td>
</tr>
<tr>
<td>55.1–65</td>
<td>53 (26)</td>
</tr>
<tr>
<td>65.1–80</td>
<td>16 (8)</td>
</tr>
<tr>
<td>&gt;80.1</td>
<td>10 (5)</td>
</tr>
</tbody>
</table>

0–40: very difficult to read; 40.1–55: somewhat difficult to read; 55.1–65: normal level of legibility; 65.1–80: quite easy to read; >80.1: very easy to read.

Rheumatology in 8th and 9th positions, and the American College of Rheumatology in 10th place. The web page of the Mexican College of Rheumatology appeared in 118th place, but was classified as irrelevant, and was excluded from the analysis. The only Mexican page included in the analysis was in 31st and 122nd place. The content quality was rated good in 11 pages (42%) and poor in 10 (38%). Concerning the evaluation of legibility, 8 (31%) were quite easy or easy to read, 9 (35%) were normal and 9 (35%) were somewhat difficult or very difficult to read.

Table 5
Characteristics of the Web Pages of Medical Institutions.

<table>
<thead>
<tr>
<th>Web page</th>
<th>Position in Google search</th>
<th>DISCERN(^a)</th>
<th>INFLESZ(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arthritis Foundation (US)</td>
<td>4</td>
<td>4</td>
<td>74.3</td>
</tr>
<tr>
<td>Spanish Society of Rheumatology (SP)</td>
<td>8, 9</td>
<td>5, 1</td>
<td>57.35</td>
</tr>
<tr>
<td>American College of Rheumatology (US)</td>
<td>10</td>
<td>4</td>
<td>71.9</td>
</tr>
<tr>
<td>Clínica DÁM (private medical specialties in Madrid) (SP)</td>
<td>12</td>
<td>4</td>
<td>60.9</td>
</tr>
<tr>
<td>Centers of Disease Control and Prevention (US)</td>
<td>18</td>
<td>4</td>
<td>61.1</td>
</tr>
<tr>
<td>Medical Center for Rheumatic Diseases – ARTRICENTER (MEX)</td>
<td>31, 122</td>
<td>5</td>
<td>61.1</td>
</tr>
<tr>
<td>Cleveland Clinic (US)</td>
<td>39</td>
<td>1</td>
<td>62.7</td>
</tr>
<tr>
<td>University of Chicago Kids Hospital (US)</td>
<td>44</td>
<td>3</td>
<td>66.0</td>
</tr>
<tr>
<td>American Academy of Orthopaedic Surgeons (US)</td>
<td>52</td>
<td>2</td>
<td>69.7</td>
</tr>
<tr>
<td>Clínica Universidad de Navarra (SP)</td>
<td>56</td>
<td>5</td>
<td>48.7</td>
</tr>
<tr>
<td>University of California San Francisco Medical Center (US)</td>
<td>69</td>
<td>1</td>
<td>55.1</td>
</tr>
<tr>
<td>Universidad de las Islas Baleares (SP)</td>
<td>105</td>
<td>1</td>
<td>44.8</td>
</tr>
<tr>
<td>Mayo Clinic (US)</td>
<td>111</td>
<td>3</td>
<td>54.7</td>
</tr>
<tr>
<td>New York Presbyterian Hospital (US)</td>
<td>166</td>
<td>3</td>
<td>57.9</td>
</tr>
<tr>
<td>Medical Center for Rehabilitation Premium-Madrid (SP)</td>
<td>175</td>
<td>3</td>
<td>44.4</td>
</tr>
<tr>
<td>Hospital Universitario Marqués de Valdecilla, Santander (SP)</td>
<td>180</td>
<td>5</td>
<td>42.9</td>
</tr>
<tr>
<td>Venezuelan Society of Rheumatology</td>
<td>198</td>
<td>2</td>
<td>56.4</td>
</tr>
<tr>
<td>Austin Heart (US)</td>
<td>217</td>
<td>1</td>
<td>48.9</td>
</tr>
<tr>
<td>Miami University (US)</td>
<td>218</td>
<td>3</td>
<td>54.9</td>
</tr>
<tr>
<td>National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH (US)</td>
<td>238</td>
<td>4</td>
<td>62.6</td>
</tr>
<tr>
<td>The Medical Center of Plane (US)</td>
<td>247</td>
<td>1</td>
<td>48.1</td>
</tr>
<tr>
<td>Children's Healthcare of Atlanta (US)</td>
<td>248</td>
<td>1</td>
<td>55.6</td>
</tr>
<tr>
<td>Yale Medical Group (US)</td>
<td>256</td>
<td>5</td>
<td>59.2</td>
</tr>
<tr>
<td>Wesley Medical Center (US)</td>
<td>265</td>
<td>1</td>
<td>54.6</td>
</tr>
<tr>
<td>The University of Iowa (US)</td>
<td>269</td>
<td>4</td>
<td>48.5</td>
</tr>
<tr>
<td>Hospital General Universitario de Elda Virgen de la Salud (SP)</td>
<td>279</td>
<td>2</td>
<td>40.5</td>
</tr>
</tbody>
</table>

MEX, Mexico; NIH, National Institutes of Health; SP, Spain; US, United States of America.

\(^a\) General quality using DISCERN – 1: low quality; 2–3: moderate quality; 4–5: high quality.

\(^b\) Legibility score using the INFLESZ scale – 0–40: very difficult to read; 40.1–55: somewhat difficult to read; 55.1–65: normal level of legibility; 65.1–80: quite easy to read; >80.1: very easy to read.

Discussion

The importance of this cross-sectional study designed to audit web pages is based on 4 premises: a) there is an upward trend in the utilization of the Internet as a source of health information; b) the information obtained via Internet can influence decision making; c) there has not been an earlier study similar to this one; and d) the high number of potential users (between 500 thousand and a million RA patients only in Spain and Mexico).

The relevance (specificity) of the web pages obtained via Google is low (62%), and is similar using different search engines and
The majority of the relevant pages are websites providing information specifically on RA. Regarding treatment, nearly half of the relevant pages refer to conventional therapy, and 13% to alternative therapies. Although half of the web pages are posted by nonprofit organizations, only 15% were created by universities, hospitals, associations or medical societies. These results could vary, however, if another search word or strategies and filters were used.

The content quality of the web pages is irregular; the majority express their objectives, but less than half of them fulfill them, and they do not state the source of the information (references); the quality is rated good in only 1 page in every 4. However, we should bear in mind that DISCERN evaluates the quality of a publication in terms of its content, but not the scientific quality or accuracy of the evidence on which the message it provides is based. On the other hand, there are 2 features that could result in a better or a poorer rating of the relevant pages. The first is the exclusion of several of the DISCERN questions related to the specific evaluation of treatment; the second is the fact that the quality was evaluated by sixth-year medical students. In this respect, we consider that the view of medical students may be “closer” to that of the patients than that of rheumatologists, whose evaluation could be contaminated with assessment bias (scientific validity). Therefore, the rates reported here should be considered subjective and taken from the perspective of the evaluators.

Legibility is a key aspect in preparing any document for patients; it is the whole of the typographical and linguistic characteristics of the written text that make it easy to read and understand. Our results show that the majority of the texts are directed at persons who have attended school for over 12 years; more than half are very difficult or somewhat difficult to read. Difficulties in terms of legibility have also been reported for web-based texts in Spanish concerning other chronic rheumatic diseases. In agreement with our results and with the average formal education in Mexico (11.9 years), more than half of the pages on RA will not be understandable to an average Mexican. The same phenomenon occurs in English. Thus, it has been recommended that the texts have a legibility equivalent to that obtained in 6 years of schooling. It is commendable that universities, hospitals, societies and medical associations (medical institutions) have a predominant position in the large volume of information on RA on the Internet. The position of these institutions can be evaluated, at least partly, considering the potential users, the place they occupy in the pages that appear with the search engine, the content quality and legibility of the text. The number of potential users (RA patients) is high. On the other hand, the place in which a website appears during the search for information is related to user traffic. In a recent study, the percentage of traffic received at the first position in Google was measured in the US and Canada; the website in the first position contributed 33% of the traffic, compared with 18% for the second position. The cumulative loss in traffic when the sites that appear on the first page in Google are compared with those of the second is even more notable; the sites listed in the first 10 results in Google generated 92% of all of the traffic in an average search, and when we changed to the second page in Google, the traffic fell by 95%. In our study, we found 26 medical institutions that posted their web page in Spanish (within the first 300): 17 (65%) from the US, 7 (27%) from Spain, one from Venezuela and one from Mexico. The Arthritis Foundation was in 4th place, with good content and a legibility that was “quite easy to read”; the Spanish Society of Rheumatology occupied places 8 and 9 (repeated), with good content in its first appearance and “normal” legibility; and the American College of Rheumatology was in 10th place, with good content and “quite easy” to read. It is notable that the only Mexican web page which occupied the 31st and 122nd places (repeated), with good content and “normal” legibility, has no rheumatologists among its medical staff. These results correspond to a certain point of time and the position in Google can vary slightly, depending on the IP address. However, it is striking that medical institutions in which Spanish is not the official language are in a better position with respect to Spanish-speaking users than the medical institutions of countries in which the predominant mother tongue is Spanish.

In short, RA patients can find a vast amount of information on their disease, although it is irregular in terms of relevance, content quality and legibility. The medical institutions, in general, are scattered with respect to their position; nevertheless, 3 are within the first 10 positions, although 2 are from the US. There is an area of opportunity so that medical institutions in countries in which Spanish is primarily the mother tongue can take the lead to provide useful information, with quality, that is easy to read and in harmony with the sociocultural characteristics of their patients.

Ethical Disclosures

Protection of human and animal subjects. The authors declare that no experiments were performed on humans or animals for this study.

Confidentiality of data. The authors declare that no patient data appear in this article.

Right to privacy and informed consent. The authors declare that no patient data appear in this article.

Funding

The authors received no external funding for performing this study.

Conflicts of Interest

The authors declare they have no conflicts of interest.

References