

## Assessment of the contents related to screening on Portuguese language websites providing information on breast and prostate cancer

Avaliação dos conteúdos em websites sobre rastreamento do câncer da mama e da próstata em língua portuguesa

Evaluación de contenidos en páginas web de lengua portuguesa sobre rastreo del cáncer de mama y próstata

Daniel Ferreira <sup>1,2</sup>  
 Helena Carreira <sup>1,2</sup>  
 Susana Silva <sup>1,2</sup>  
 Nuno Lunet <sup>1,2</sup>

### Abstract

*The objective of this study was to assess the quality of the contents related to screening in a sample of websites providing information on breast and prostate cancer in the Portuguese language. The first 200 results of each cancer-specific Google search were considered. The accuracy of the screening contents was defined in accordance with the state of the art, and its readability was assessed. Most websites mentioned mammography as a method for breast cancer screening (80%), although only 28% referred to it as the only recommended method. Almost all websites mentioned PSA evaluation as a possible screening test, but correct information regarding its effectiveness was given in less than 10%. For both breast and prostate cancer screening contents, the potential for overdiagnosis and false positive results was seldom addressed, and the median readability index was approximately 70. There is ample margin for improving the quality of websites providing information on breast and prostate cancer in Portuguese.*

*Breast Neoplasms; Prostatic Neoplasms; Internet*

### Resumo

*O objetivo deste estudo foi avaliar a qualidade dos conteúdos sobre rastreamento, numa amostra de websites com informação sobre o câncer da próstata e/ou mama, em língua portuguesa. Consideraram-se os primeiros 200 resultados de cada busca no Google. A adequação dos conteúdos sobre rastreo foi definida de acordo com a melhor evidência científica disponível e se avaliou a sua legibilidade. Cerca de 80% dos websites referiu a mamografia como um método para rastreamento do câncer da mama, mas apenas 28% a referiram como o único método recomendado. Quase todos os websites referiram a pesquisa de PSA como um possível teste de rastreo, mas somente 10% apresentaram informação correta relativamente à efetividade dessa forma de rastreo. Para os conteúdos de ambos os cânceres, o potencial para sobrediagnóstico e para um resultado falso-positivo raramente foi mencionado, e a mediana do índice de legibilidade foi de aproximadamente 70. Existe uma larga margem para melhorar a qualidade dos websites com informações sobre câncer da mama e da próstata.*

*Neoplasias da Mama; Neoplasias da Próstata; Internet*

<sup>1</sup> Faculdade de Medicina, Universidade do Porto, Porto, Portugal.

<sup>2</sup> Instituto de Saúde Pública, Universidade do Porto, Porto, Portugal.

#### Correspondence

N. Lunet  
 Departamento de Epidemiologia Clínica, Medicina Preditiva e Saúde Pública, Faculdade de Medicina, Universidade do Porto.

Al. Prof. Hernâni Monteiro  
 4200, 319, Porto, Portugal.  
 nlunet@med.up.pt

## Introduction

The use of the Internet has increased over the last years, mainly because it is easily accessible and allows gathering information from different sources<sup>1,2</sup>. It has become one of the most important sources of both general and health-related information, and its potential to influence individual health behaviours emphasizes the importance of monitoring the quality of health contents available on websites<sup>3,4,5</sup>.

Although there are different guidelines to assess the formal quality of these sources of information<sup>6,7</sup>, as well as tools to assess the readability of the contents, there are no instruments to evaluate the accuracy of the information on specific topics<sup>8,9</sup>. Such an evaluation needs to be conducted case by case, taking into account the best available evidence on each health topic and the local health policies<sup>10</sup>.

Information related to oncological diseases corresponds to an important proportion of the Internet searches on health issues<sup>11</sup>, and breast and prostate cancer patients are the ones who use the Internet more frequently to search for information related to their disease<sup>12</sup>. Specifically breast and prostate cancers are leading causes of oncological morbidity and are among the malignancies with the highest relative survival, which leads to the seeking of information on these topics by the general population, and by patients and their families in particular<sup>11,13,14,15</sup>. Furthermore, breast and prostate cancers have specificities regarding the potential for control through secondary prevention, and a large proportion of women and men participate in screening activities, even though population-based screening is recommended only for breast cancer. Thus, these oncological diseases may constitute a good model for designing a framework for website quality assessment that may be extended to other conditions for which screening is recommended or effectively conducted, regardless of the available evidence on its effectiveness.

We aimed to replicate an Internet search conducted by a lay-person and to assess the quality of the contents on breast and prostate cancer screening in the websites that provide information on breast and prostate cancer in Portuguese.

## Material and methods

### Selection of the websites for analysis

We searched the World Wide Web to identify Portuguese language web pages that addressed breast or prostate cancers, on the 16<sup>th</sup> and 15<sup>th</sup> of

September 2011, respectively, using the Google search engine (<http://www.google.com>), with the expressions “*cancro da mama*” and “*cancro da prostate*”, respectively. We saved the first 200 results from each search for further analysis, including information on the URL (Uniform Resource Locator) of each web page, and registered its rank in the search. The websites were initially screened to assess eligibility, by applying the following exclusion criteria: inaccessible websites due to non-functioning URL; websites not providing information in Portuguese; repeated websites (corresponding to different web pages from the same website); websites providing information on breast or prostate cancer only in the format of downloadable files (e.g. slideshows, portable document files), or only through audio or videos (e.g. YouTube videos); scientific articles (whether or not located in medical websites); blogs or forums; general encyclopedias; websites only providing information about female breast or prostate cancers in the form of news; websites with no specific information on female breast or prostate cancers (e.g. advertising only, male breast cancer).

To identify the contents related to breast or prostate cancers in the eligible websites we proceeded as follows: when the URL corresponded to a website's main page, we searched the whole site; when the URL corresponded to a web page other than the website's main page, we navigated towards the latter, and then a more comprehensive screening of the website was conducted for identification of all relevant pages.

### General characterization of the websites

The general characterization of the websites was accomplished using information depicted in any of their pages. One investigator (D.F.) gathered data on the following variables: the website's main subject; country of origin; intended audience; media used to convey the cancer-specific information; profit motive of the owner of the website.

The websites were classified regarding the predominant relation of its contents to health (health related/not only health related), to cancer (cancer related/not only cancer related) and to breast or prostate specific disease (breast or prostate cancer specific/not specific for breast or prostate cancer).

The websites were identified as registered in Portugal, Brazil or another country. This information was assessed through the domain (“.pt” for Portugal, “.br” for Brazil). For other domains, the contact information of the website was consulted. The other origins of websites included

African countries where Portuguese is the official language, or other non Portuguese-speaking countries.

The intended audience was classified as general population, patients, health professionals or media. Since the websites could target more than one population group, these categories were not mutually exclusive. We searched the website "disclaimer"/"about" item to obtain this information. When it was not specified, we carried out a search to assess whether the area of activity of the institution that owned the website could be associated with a specific population or population group; if not, the website was considered to target the general population.

The media used by the websites to convey the cancer-specific information (display of contents) were classified in six mutually exclusive categories: text only; text and figures; text and video; text and charts; text and audio; other.

The affiliation of the websites was primarily defined as public or private. Among the private institutions, we distinguished the organizations responsible for the websites based on whether it was a profit making organization or not, and grouped them as for-profit (e.g. health care providers, pharmaceutical industries, individual subjects) or non-profit (e.g. non-governmental organizations). The websites were classified according to the profit intent, considering public and non-governmental organizations as non-profit, and private institutions as for-profit.

#### **Analysis of the contents related to screening of breast and prostate cancer**

##### **• Specific contents on cancer screening**

We analyzed the contents of the websites on this topic, namely regarding the existence of specific information on cancer screening and their accuracy. We selected topics that covered the different methods for screening and its effectiveness, the potential harms of screening, the recommended periodicity, the eligibility for screening and instructions on how to proceed to be screened.

The criteria to assess the accuracy of the information and its adequacy to the Portuguese setting were defined in accordance with the evidence summarized by the U.S. Preventive Services Task Force (USPSTF)<sup>16,17</sup>, the European Union Advisory Committee on Cancer Prevention<sup>18</sup> and the local policy for cancer screening<sup>19</sup>. In Portugal, there is a screening program for breast cancer, which differs slightly from the U.S. Preventive Task Force recommendations and the EU Advisory Committee on Cancer

Prevention, especially regarding the age from which women should start their regular biennial mammograms (45-69 years)<sup>19</sup>.

From each website we selected the information about screening for further analysis. The specific items searched, as well as the message considered the most appropriate to convey to the general population are presented in Figures 1 and 2, for breast and prostate cancer. For each item three options were possible: does not mention the subject; mentions the subject but the information is incorrect or incomplete; mentions the subject and the information is correct.

##### **• Readability**

To assess the readability of the contents on cancer screening, in the websites providing information on breast cancer we selected the text from the sections related to symptoms, diagnosis, types of cancer and screening, while in the websites providing information on prostate cancer we selected information related to screening, cancer detection and diagnosis. These sections were systematically selected in all websites to ensure comparability.

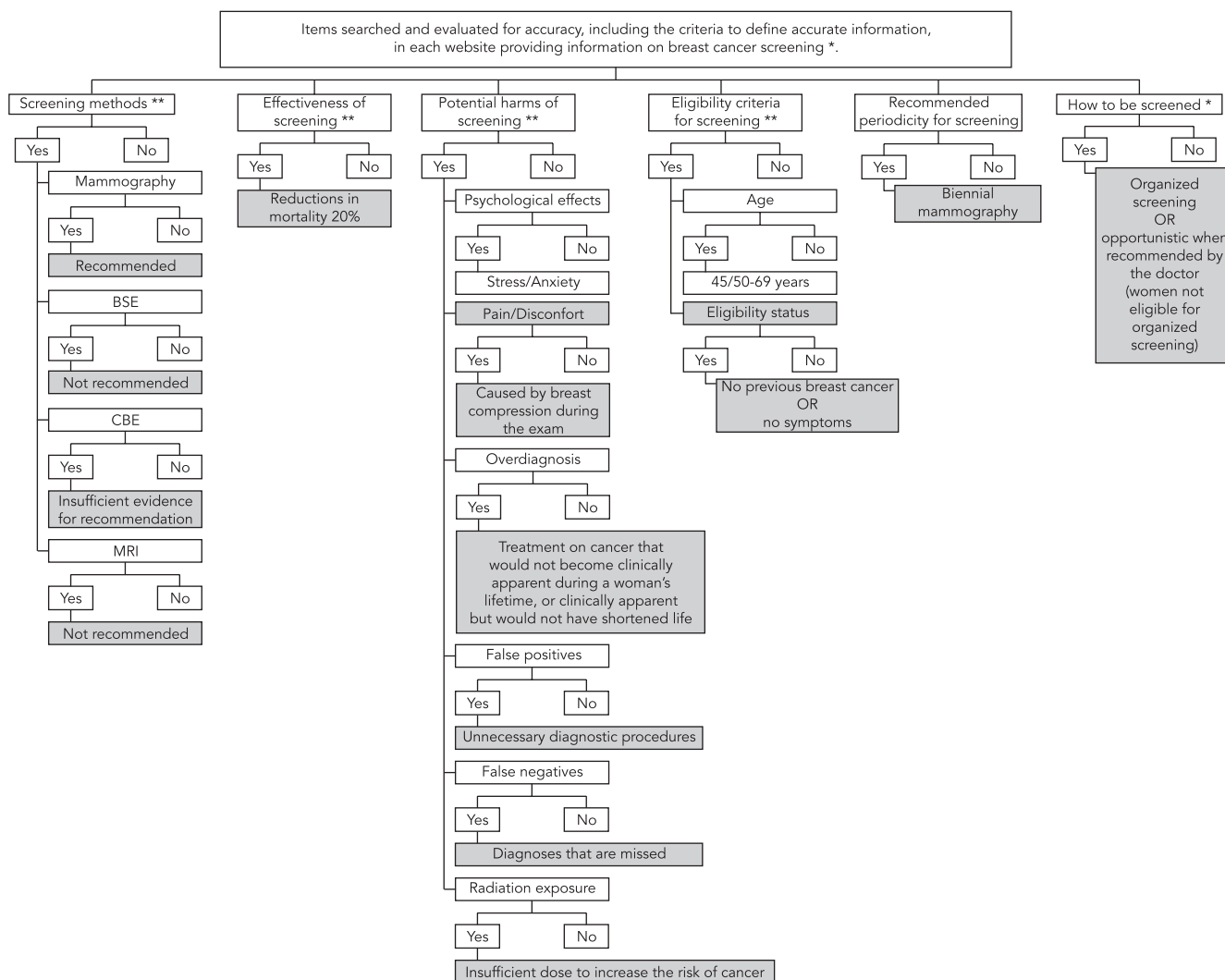
We used the Fernandez-Huerta index to determine the readability of the contents. This index is computed as  $[206.84 - 0.6 * (\text{average number of syllables per word}) - 1.02 * (\text{average number of words per sentence})]$ ; the results range from 0 to 100, representing the worst level (very difficult to read) and the better level of readability, respectively. To estimate the number of words and of syllables per word, we extracted the information to a Microsoft Office Word (Microsoft Corp., USA) document and analyzed the text using the software TextMeter (<http://www.lazarusbrasil.org/textmeter.php>, Brazil), which is an application of text statistics only for the Portuguese language. This software counts the number of words and sentences, and also provides an algorithm for counting syllables.

##### **Data analysis**

The results are presented as the proportion of websites depicting each one of the characteristics assessed, for the whole sample and by cancer type (breast vs. prostate cancers) and by website rank in each of the searches (first 30 URL vs. remaining results). This cut-off was selected because individuals who search on the Internet tend to navigate until the third page of results<sup>20</sup>. The contents on screening were further analyzed by country of origin of the websites and the profit motive. The proportions were compared with the  $\chi^2$  or the Fisher exact test, as appropriate.

Figure 1

Procedure followed in the analysis of information on breast cancer screening.



BSE: breast self-examination; CBE: clinical breast exam; MRI: magnetic resonance imaging.

\* Accuracy of information defined according to U.S. Preventive Services Task Force <sup>16</sup>, Advisory Committee on Cancer Prevention <sup>18</sup> and Coordenação Nacional Para as Doenças Oncológicas, Alto Comissariado da Saúde, Ministério da Saúde <sup>19</sup>. The shadowed cells represent the information considered correct;

\*\* The items addressing similar subjects were grouped together and it was considered that the topic was not mentioned (when none of the items within the topic were mentioned), mentioned with incorrect information (when this applied to at least one of the items with no correct information being provided in each of them), or that the topic was mentioned correctly (when this applies to at least one of the items within the same topic).

The results regarding the readability index were compared between breast and prostate cancer websites and, for each of them, according to the websites' characteristics using the Kruskal-Wallis test.

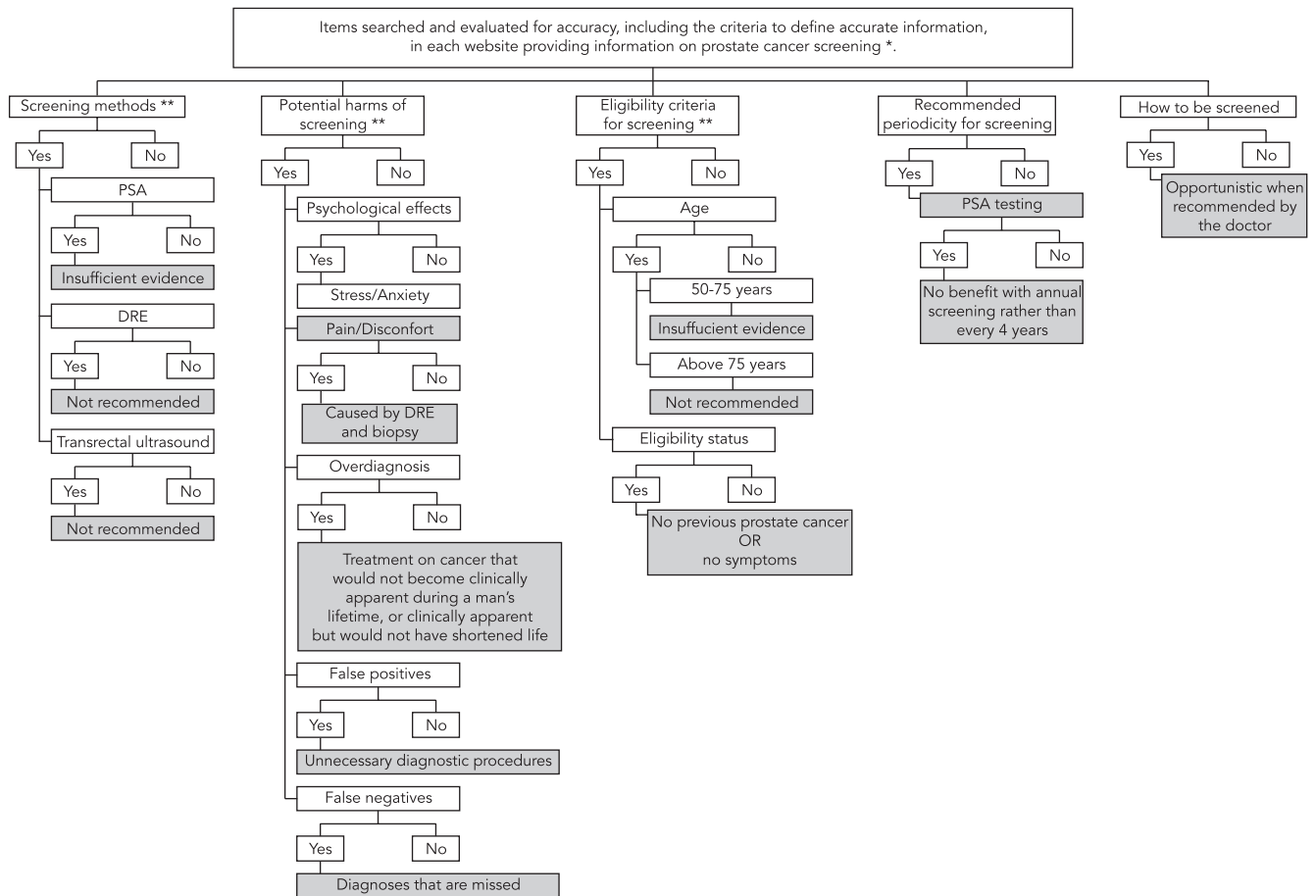
## Results

### Websites selected for analysis

In the first 200 results retrieved by each cancer-specific Google search, 47 websites addressing issues related with breast cancer and 67 websites

Figure 2

Procedure followed in the analysis of information on prostate cancer screening.



DRE: digital rectal examination; PSA: prostate specific antigen.

\* Accuracy of information defined according to U.S. Preventive Services Task Force 16, Advisory Committee on Cancer Prevention 18 and Coordenação Nacional Para as Doenças Oncológicas, Alto Comissariado da Saúde, Ministério da Saúde 19. The shadowed cells represent the considered correct justification;

\*\* The items addressing similar subjects were grouped together and it was considered that the topic was not mentioned (when none of the items within the topic were mentioned), mentioned with incorrect information (when this applied to at least one of the items with no correct information being provided in each of them), or that the topic was mentioned correctly (when this applies to at least one of the items within the same topic).

with prostate cancer information fulfilled the eligibility criteria (Figure 3). Among the former, 35 websites (74%) covered issues related with breast cancer screening and 43 websites of the latter (64%) provided specific information on prostate cancer screening.

#### General characteristics of the websites

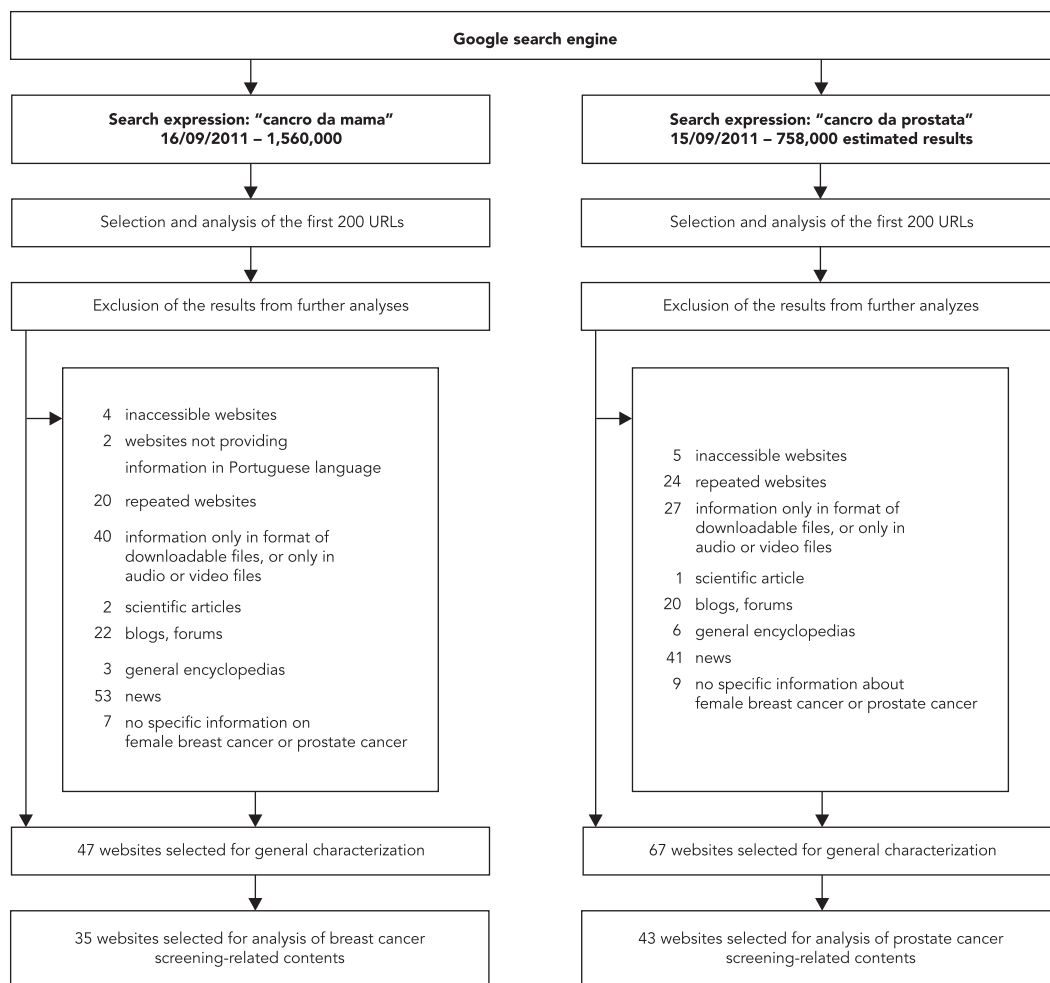
Seven out of 10 websites providing information on breast and prostate cancers were health-related, and the proportion was higher among those

appearing in the first thirty results of the search (86.2% vs. 67.1%,  $p = 0.048$ ). Approximately 20% and 10% of the websites exclusively covered issues related with cancer or specifically breast/prostate cancer, respectively; these appeared in the first pages of the search 4 and 9 times more frequently, respectively (Table 1).

Nearly half of the websites were from Portugal, and it was more likely to find a Portuguese website in the first 30 results (79.3% vs. 43.5%,  $p = 0.004$ ). The websites appearing in the first three pages of results were more frequently aimed

Figure 3

Selection of the Internet search results for breast cancer and prostate cancer.



at cancer patients (44.8% vs. 3.5%,  $p < 0.001$ ) and less often at the general population (75.9% vs. 96.5%,  $p = 0.001$ ). Approximately three-quarters of the websites provided the information only in the format of text; video and audio were seldom used. Approximately 15% of the websites were from non-profit organizations, and appeared more frequently in the first 30 results (31% vs. 9.4%,  $p = 0.005$ ).

There were no statistically significant differences in the characteristics of the websites according to the cancer addressed. However, those providing information on breast cancer tended to target the general population less often (85.1% vs. 95.5%,  $p = 0.053$ ) and those on prostate can-

cer were more frequently from Brazil (35.8% vs. 21.3%,  $p = 0.123$ ).

#### **Contents related to screening of breast and prostate cancer**

##### **• Accuracy of the contents on breast cancer screening**

Most websites mentioned mammography as a method for breast cancer screening (80%), although only 28% mentioned it correctly as the only recommended method for screening, and sound quantitative estimates of the effectiveness were provided in only 14%. The breast self-exam

Table 1

General characteristics of the websites selected for analysis.

	Websites providing health information on cancer						
	All	Cancer		Order of appearance in search			
	n (%) *	Breast n (%)	Prostate n (%)	p-value	≤ 30 n (%)	> 30 n (%)	p-value
Health-specific website							
No	32 (28.1)	13 (27.7)	19 (28.4)	0.935	4 (13.8)	28 (32.9)	0.048
Yes	82 (71.9)	34 (72.3)	48 (71.6)		25 (86.2)	57 (67.1)	
Cancer-specific website							
No	92 (80.7)	38 (80.8)	54 (80.6)	0.973	16 (55.2)	76 (89.4)	< 0.001
Yes	22 (19.3)	9 (19.2)	13 (19.4)		13 (44.8)	9 (10.6)	
Cancer-specific (covering exclusively breast or prostate cancer)							
No	102 (89.5)	42 (89.4)	60 (89.6)	0.974	20 (69.0)	82 (96.5)	< 0.001
Yes	12 (10.5)	5 (10.6)	7 (10.4)		9 (31.0)	3 (3.5)	
Country of origin							
Portugal	60 (52.6)	30 (63.8)	30 (44.8)	0.123	23 (79.3)	37 (43.5)	0.004
Brazil	34 (29.8)	10 (21.3)	24 (35.8)		4 (13.8)	30 (35.3)	
Other	20 (17.5)	7 (14.9)	13 (19.4)		2 (6.9)	18 (21.2)	
Intended audience **							
General population	104 (91.2)	40 (85.1)	64 (95.5)	0.053	22 (75.9)	82 (96.5)	0.001
Cancer patients	16 (14.1)	9 (19.2)	7 (10.4)	0.188	13 (44.8)	3 (3.5)	< 0.001
Media professionals	2 (1.8)	1 (2.1)	1 (1.5)	> 0.999	2 (6.9)	0 (0.0)	0.063
Health professionals	9 (7.9)	5 (10.6)	4 (6.0)	0.363	2 (6.9)	7 (8.2)	0.817
Display of contents **							
Text only	83 (72.8)	37 (78.7)	46 (68.7)	0.234	21 (72.4)	62 (72.9)	0.956
Text and figures	27 (23.7)	8 (17.1)	19 (28.4)	0.161	6 (20.7)	21 (24.7)	0.660
Text and videos	4 (3.5)	2 (4.3)	2 (3.0)	0.717	2 (6.9)	2 (2.4)	0.251
Text and charts	8 (7.1)	4 (8.5)	4 (6.0)	0.601	2 (6.9)	6 (7.1)	0.976
Text and audio	0 (0.0)	0 (0.0)	0 (0.0)	-	0 (0.0)	0 (0.0)	-
Profit intent							
Profit	97 (85.1)	39 (83.0)	58 (86.6)	0.596	20 (69.0)	77 (90.6)	0.005
Non-profit	17 (14.9)	8 (17.0)	9 (13.4)		9 (31.0)	8 (9.4)	

\* Within each variable the sum of the proportions may not be 100% due to rounding;

\*\* Categories not mutually exclusive.

and the clinical breast exam were mentioned almost as often as the mammography, but the information provided was usually incorrect. The potential for overdiagnosis, false positive and false negative results was addressed in a very low percentage of the websites, and most of the times the information was not correct. The information that the dose of radiation exposure in mammography testing is insufficient to increase the risk of cancer was correctly mentioned in 14.3% of the websites. Approximately one-quarter of the websites gave correct information about the eligible ages for screening, but the fact that screen-

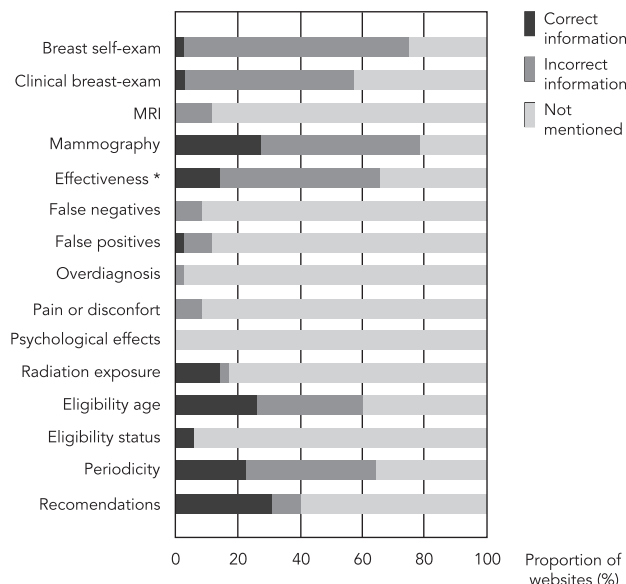
ing applies only to asymptomatic subjects was seldom addressed. The adequate periodicity for screenings was mentioned in 22.2% of the websites and the recommendations on how to perform a screening test were correct in 31.4% of the websites (Figure 4a).

The websites appearing on the first 30 results tend to have better information about screening harms (30% vs. 4%,  $p = 0.014$ ), and about periodicity of screening (70% vs. 12%,  $p = 0.004$ ). Websites owned by a non-profit organization tended to provide information more frequently on how to proceed to be screened (83.3% vs. 20.7%,

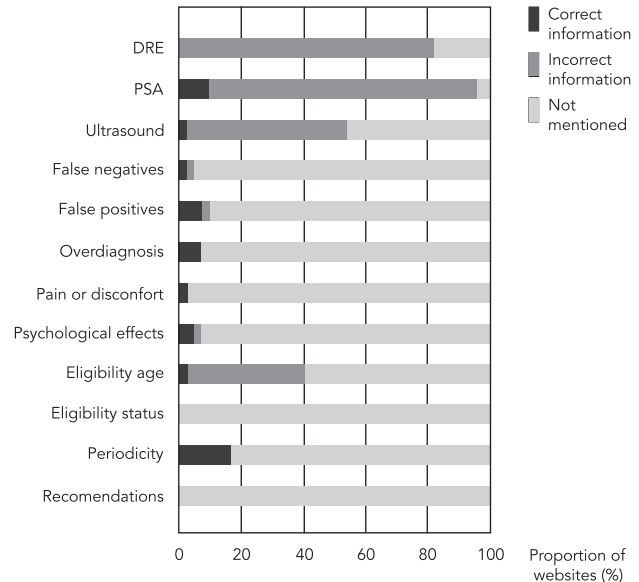
Figure 4

Quality of the contents on breast and prostate cancer screening.

## 4a) Breast cancer



## 4b) Prostate cancer



DRE: digital rectal examination; MRI: magnetic resonance imaging; PSA: prostate-specific antigen.

\* There is no evidence about the effectiveness of screening of the prostate cancer.

$p = 0.011$ ), and to mention correctly the potential harms of screening (50% vs. 3.5%,  $p = 0.019$ ) (Table 2).

- **Accuracy of the contents on prostate cancer screening**

The evaluation of the prostate specific antigen (PSA) was mentioned as a possible screening test in nearly all websites, but information regarding the insufficient evidence of its effectiveness was given in less than 10%. The most frequently referred harm of screening with the correct information was the potential for overdiagnosis and false positive (both 6.9%). None of the websites mentioned that screening targets asymptomatic subjects, and the age-groups potentially eligible for screening were addressed by 39.5%, most of the times incorrectly. The periodicity of the screening was mentioned in less than a fifth of the websites and never with the correct information. None of the websites provided information on how to proceed to be screened (Figure 4b).

No significant differences were found in the analysis of contents on prostate cancer screen-

ing, according to the order of appearance in the search, country of origin or profit intent of the websites' affiliation, except for the less frequent reference to the potential harms of screening among the first 30 results (Table 3).

- **Readability**

The median readability index values were not significantly different between the websites providing information on breast and prostate cancer (73.1 vs. 69.7,  $p = 0.144$ ). The readability of the contents related to breast cancer screening was lower on Portuguese websites (median: 70.2 vs. 75.7,  $p = 0.036$ ) and on for profit websites (68.7 vs. 73.7,  $p = 0.035$ ). Readability of content related to prostate cancer screening did not vary meaningfully by order of appearance of the website, country of origin or profit motive (Figure 5).

## Discussion and conclusion

Most of the websites that addressed breast or prostate cancer provided information on cancer



Table 2

Quality of the contents on breast cancer screening according to websites' order of appearance and country of origin and profit intent of the websites' affiliation.

	Contents on breast cancer screening								
	Order of appearance in search			Country of origin			Profit intent of the websites' affiliation		
	≤ 30	> 30	p-value	Portugal	Other	p-value	Non-profit (Public or NGO)	For profit (Private)	p-value
	n (%)	n (%)		n (%)	n (%)		n (%)	n (%)	
Screening methods									
Not mentioned	0 (0.0)	5 (20.0)	0.458	4 (16.7)	1 (9.1)	0.788	0 (0.0)	5 (17.2)	0.208
Mentioned/Incorrect	4 (40.0)	9 (36.0)		8 (33.3)	5 (45.4)		1 (16.7)	12 (41.4)	
Mentioned/Correct	6 (60.0)	11 (44.0)		12 (50.0)	5 (45.4)		5 (83.3)	12 (41.4)	
Effectiveness									
Not mentioned	2 (20.0)	10 (40.0)	0.588	8 (33.3)	4 (36.4)	> 0.999	1 (16.7)	11 (37.9)	0.382
Mentioned/Incorrect	6 (60.0)	12 (48.0)		12 (50.0)	6 (54.6)		5 (83.3)	13 (44.8)	
Mentioned/Correct	2 (20.0)	3 (12.0)		4 (16.7)	1 (9.1)		0 (0.0)	5 (17.2)	
Harms									
Not mentioned	5 (50.0)	23 (92.0)	0.014	18 (75.0)	10 (90.9)	0.442	3 (50.0)	25 (86.2)	0.019
Mentioned/Incorrect	2 (20.0)	1 (4.0)		2 (8.3)	1 (9.1)		0 (0.0)	3 (10.3)	
Mentioned/Correct	3 (30.0)	1 (4.0)		4 (16.7)	0 (0.0)		3 (50.0)	1 (3.5)	
Eligibility									
Not mentioned	2 (20.0)	12 (48.0)	0.224	10 (41.7)	4 (36.4)	< 0.001	2 (33.3)	12 (41.4)	0.391
Mentioned/Incorrect	4 (40.0)	8 (32.0)		8 (33.3)	4 (36.4)		1 (16.7)	11 (37.9)	
Mentioned/Correct	4 (40.0)	5 (20.0)		6 (25.0)	3 (27.3)		3 (50.0)	6 (20.7)	
Periodicity									
Not mentioned	2 (20.0)	14 (56.0)	0.004	13 (54.2)	3 (27.3)	0.148	2 (33.3)	14 (48.3)	0.063
Mentioned/Incorrect	1 (10.0)	8 (32.0)		4 (16.7)	5 (45.4)		0 (0.0)	9 (31.0)	
Mentioned/Correct	7 (70.0)	3 (12.0)		7 (29.2)	3 (27.3)		4 (66.7)	6 (20.7)	
Access on screening									
Not mentioned	5 (50.0)	16 (64.0)	0.322	12 (50.0)	9 (81.8)	0.155	1 (16.7)	20 (69.0)	0.011
Mentioned/Incorrect	0 (0.0)	3 (12.0)		2 (8.3)	1 (9.1)		0 (0.0)	3 (10.3)	
Mentioned/Correct	5 (50.0)	6 (24.0)		10 (41.7)	1 (9.1)		5 (83.3)	6 (20.7)	

NGO: Non Governmental Organization.

screening, though it was often incomplete or inaccurate. It is noteworthy that the possible harms of the screening were frequently overlooked. Despite the poor overall quality of the contents, the websites obtained good scores on readability.

In the present study we described the assessment of the quality of the websites' contents with the necessary detail to ensure the transparency of the process. It provides a framework for analysis that can be used by other researchers and for the monitoring of the quality of the health information provided in the internet. However, it has limitations that need to be addressed.

The number of websites selected for analysis was relatively small, as we were attempting to replicate searches conducted by a layper-

son looking for general information on breast or prostate cancer. The small sample is probably an unavoidable limitation, given the need to use relatively simple and unspecific search terms and the expectation that most people are not willing to filter through a large number of websites to obtain the information they require<sup>20,21</sup>. Nonetheless, this study is one of the largest conducted on this issue, as other similar works selected 30<sup>3,22</sup>, 50<sup>23</sup>, or 100<sup>4</sup> results.

Similarly, only Google was selected because it is the most popular search engine among the Portuguese speaking population<sup>15</sup>. Although the use of other search engines could yield a different sample of eligible websites, the internal validity of our study is not compromised by this

Table 3

Quality of the contents on prostate cancer screening according to websites' order of appearance and country of origin and profit intent of the websites' affiliation.

	Contents on prostate cancer screening								
	Order of appearance in the search			Country of origin			Profit intent of the websites' affiliation		
	≤ 30	> 30	p-value	Portugal	Other	p-value	Non-profit (Public or NGO)	For profit (Private)	p-value
	n (%)	n (%)		n (%)	n (%)		n (%)	n (%)	
Screening methods									
Not mentioned	0 (0.0)	2 (6.9)	0.811	1 (5.0)	1 (4.4)	0.801	0 (0.0)	2 (5.3)	> 0.999
Mentioned/Incorrect	13 (92.9)	24 (82.8)		18 (90.0)	19 (82.6)		5 (100)	32 (84.2)	
Mentioned/Correct	1 (7.1)	3 (10.3)		1 (5.0)	3 (13.0)		0 (0.0)	4 (10.5)	
Harms									
Not mentioned	11 (78.6)	24 (82.8)	< 0.001	15 (75.0)	20 (87.0)	0.869	5 (100)	30 (79.0)	> 0.999
Mentioned/Incorrect	1 (7.1)	1 (3.4)		1 (5.0)	1 (4.4)		0 (0.0)	2 (5.3)	
Mentioned/Correct	2 (14.3)	4 (13.8)		4 (20.0)	2 (8.7)		0 (0.0)	6 (15.8)	
Eligibility									
Not mentioned	10 (71.4)	16 (55.2)	0.666	13 (65.0)	13 (56.5)	0.871	3 (60.0)	23 (60.5)	> 0.999
Mentioned/Incorrect	4 (28.6)	12 (41.4)		7 (35.0)	9 (39.1)		2 (40.0)	14 (36.8)	
Mentioned/Correct	0 (0.0)	1 (3.4)		0 (0.0)	1 (4.4)		0 (0.0)	1 (2.6)	
Periodicity									
Not mentioned	13 (92.9)	23 (79.3)	0.396	17 (85.0)	19 (82.6)	> 0.999	5 (100.0)	31 (81.6)	0.572
Mentioned/Incorrect	1 (7.1)	6 (20.7)		3 (15.0)	4 (17.4)		0 (0.0)	7 (18.4)	
Mentioned/Correct	0 (0.0)	0 (0.0)		0 (0.0)	0 (0.0)		0 (0.0)	0 (0.0)	
Access on screening									
Not mentioned	14 (100.0)	29 (100.0)	-	20 (100.0)	23 (100.0)	-	34 (100.0)	9 (100.0)	-
Mentioned/Incorrect	0 (0.0)	0 (0.0)		0 (0.0)	0 (0.0)		0 (0.0)	0 (0.0)	
Mentioned/Correct	0 (0.0)	0 (0.0)		0 (0.0)	0 (0.0)		0 (0.0)	0 (0.0)	

NGO: Non Governmental Organization.

methodological option. The same reasoning applies to the fact that our search was conducted on one day for each type of cancer, and the websites that would be identified at other moments could be different <sup>24</sup>.

Another limitation of our study is the data collection from the websites made by only one investigator. However, the procedures for the evaluation of the websites were standardized and based on criteria defined *a priori*, to make the assessment as replicable as possible. Furthermore, a second investigator was involved in the discussion of the evaluation of the websites whenever their characteristics did not match entirely the predefined framework of assessment.

Our study evaluated the quality of the contents specifically related with screening. Other investigations on the overall quality of the contents of websites assessed a wider range of aspects, according to the specific subject <sup>3,4,25,26,27</sup>.

Therefore, it is not possible to compare directly the quality of the websites providing information on breast or prostate cancers screening with previous investigations, though the quality of contents available on the Internet related with health issues has been considered poor <sup>27</sup>.

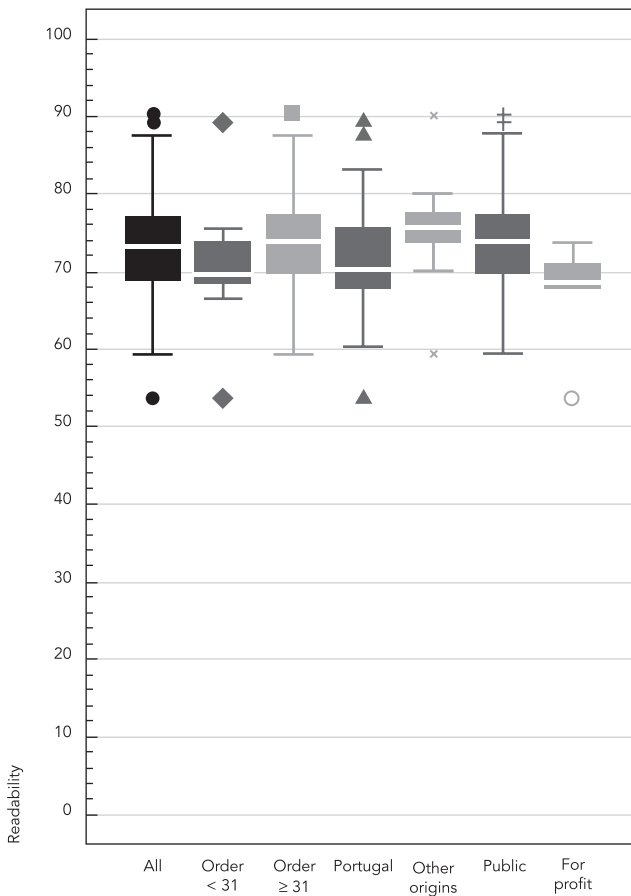
As in previous studies, the results tend to demonstrate that websites appearing in the first 30 results (with lower page rank) tend to provide more reliable information <sup>28</sup>, which can be explained by a higher specificity of the websites for breast or prostate cancer issues (as shown by our results). Also, better websites tend to be more linked or referred by other websites, which increases their importance and as a consequence decreases their page rank, placing them in the first places of the search results.

The information on screening tends to be better on the female breast cancer websites than on those related to prostate cancer, namely

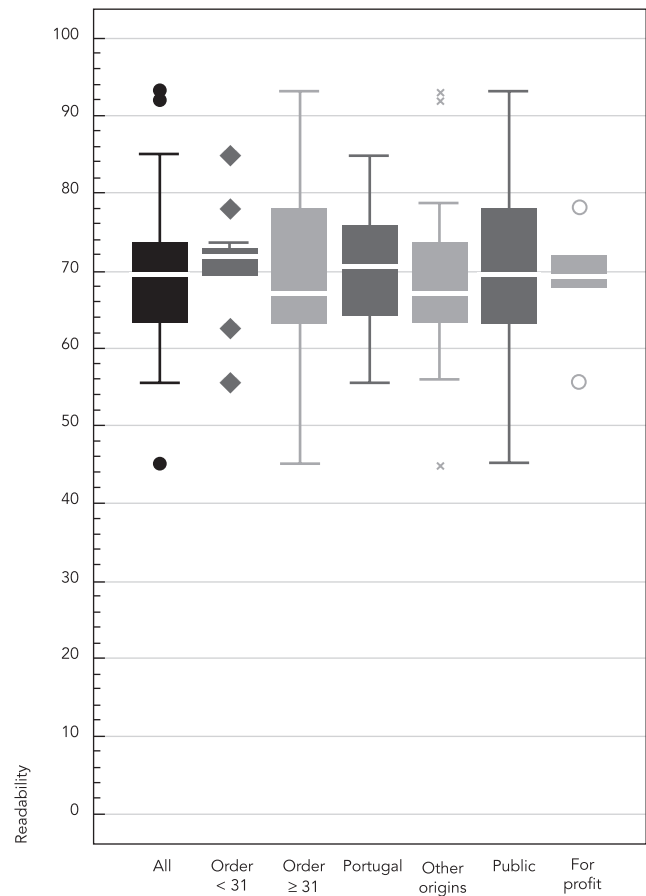
Figure 5

Readability of the contents on breast (n = 35) and prostate cancer screening (n = 43) by order of appearance in the search, country of origin and profit intent of the websites.

5a) Breast cancer



5b) Prostate cancer



regarding the screening methods. In the setting of our study, we hypothesized that this fact must be explained by the existence of organized screening for breast cancer<sup>19</sup>, while no similar screening strategy is recommended or available for prostate cancer, as its effectiveness remains controversial and overdiagnosis is a major public health concern<sup>17</sup>. Also, the websites whose country of origin was Portugal tended to provide better information on screening, which can be explained by the fact that we assessed the correctness of information according to the recommendations/guidelines followed in Portugal. This shows that, although this general frame-

work for evaluation of the quality of the website's contents may be used in any other Portuguese speaking-country, the results obtained will be setting-specific. Moreover, the assessment of websites' contents in other Portuguese-speaking countries needs to account for the specificities of the Portuguese language in each setting. For instance, in Brazil the term used to refer to cancer is different from the one used in Portugal (Brazil: *câncer*; Portugal: *cancro*). Therefore, if the Brazilian form was used the search would retrieve different websites, which illustrates the need to conduct setting specific surveys of the quality of health information available in the Internet.

In spite of the aforementioned expected differences across settings, our results are in accordance with what would be expected in most contexts<sup>27,29,30</sup>. Particularly, the results related with the profit motive of the websites and their affiliated organizations are less likely to be locale-specific. Internet users may be expected to find the information provided by websites from public or nongovernmental organizations more reliable than the ones associated with for profit organizations, as commercial interests may be responsible for incomplete or incorrect information on these websites<sup>31</sup>.

The harms of screening were also seldom addressed. This is of particular relevance for prostate cancer screening, whose potential benefits are not considered to outweigh the deleterious effects that may be associated with it<sup>11</sup>. The absence of this information was particularly notorious in the for profit websites. At a population level, this may contribute to a larger number of subjects undergoing screening without having the necessary knowledge for a well-informed decision.

Our study only focused on contents related with screening, which targets asymptomatic subjects in eligible ages. Notwithstanding, subjects already presenting signs and symptoms that require medical attention may search the internet to obtain more information. The impact of the information on screening over these subjects is difficult to ascertain, and the assessment of the accuracy of the websites' contents directed to these conditions was not the aim of our study.

Readability refers to the facility with which a text is read<sup>32</sup>, being an important aspect of the quality of a website's content<sup>33</sup>. We assessed the readability of the content on screening us-

ing the Fernandez-Huerta index, which was created to assess texts written in Spanish. Although it has not yet been validated in the Portuguese language, the Spanish and Portuguese languages share the same Latin basis, and this tool has been used to assess the quality of Brazilian governmental websites<sup>33</sup>. We considered that the websites presented a good level of readability, as 70 has been accepted to correspond to a good level of readability, in Portuguese texts, when using the Fernandez-Huerta index<sup>33</sup>. Further work is needed to establish the correspondence between the score attributed to the website and the education level needed to understand the information (according to the Portuguese curricula)<sup>34</sup>. To the best of our knowledge, there are no similar investigations on health related aspects that aimed to assess readability in websites in Portuguese, which precludes a more in depth discussion of our results.

The present study demonstrates that the quality of the contents on breast and cancer screening in Portuguese is far from good, warranting continuous monitoring and educational and regulatory actions to ensure that the general population and the patients are not "exposed" to misleading information on the Internet. Regulation of the websites and education by information providers have been considered to improve the general quality of the websites<sup>9,35</sup>.

In conclusion, there is a large margin for improving the quality of Portuguese language websites providing information on breast and prostate cancer. This study provides a framework for the standardized assessment of the quality of the contents of websites providing information on breast or prostate cancer, which may be used for the monitoring of the quality of the health information provided in the Internet.

## Resumen

*El objetivo fue evaluar la calidad de los contenidos en lengua portuguesa sobre rastreo en una muestra de páginas web con información sobre el cáncer de próstata y/o mama. Se consideraron los primeros 200 resultados de cada búsqueda en Google. La adecuación de los contenidos sobre rastreo se definió de acuerdo con la mejor evidencia científica disponible y se evaluó su legibilidad. Cerca de un 80% de las páginas web se refirieron a la mamografía como un método para el rastreo del cáncer de mama, sin embargo, solamente un 28% la mencionaron como el único método recomendado. Casi todas las páginas web señalaron el examen de Antígeno Prostático Total (APT/PSA en inglés) como un posible test de rastreo, pero solamente un 10% presentó información correcta respecto a la efectividad de esta forma de rastreo. En lo referente a los contenidos de ambos cánceres, el potencial para un sobrediagnóstico y un resultado falso positivo raramente fue mencionado, y la mediana del índice de legibilidad fue de aproximadamente 70. Existe un ancho margen para mejorar la calidad de las páginas web con información sobre cáncer de mama y de próstata.*

*Neoplasias de la Mama; Neoplasias de la Próstata; Internet*

## Contributors

D. Ferreira contributed in the study design, was the main person responsible for the acquisition of data, collaborated in its analysis and interpretation, and wrote the first draft of the manuscript. H. Carreira collaborated in the data analysis and revision of the article. S. Silva and N. Lunet participated in the design of the study, and reviewed the article for important intellectual content.

## Acknowledgments

This study was financed by the Foundation for Science and Technology (Portuguese Ministry of Education and Science) within the project *Health Information of Portuguese Population: Knowledge and Perceived Quality and Accessibility of Health Information Sources* (HMSP-IISE/SAU-ICT/0004/2009).

## References

- Helft PR, Hlubocky F, Daugherty CK. American oncologists' views of internet use by cancer patients: a mail survey of American Society of Clinical Oncology members. *J Clin Oncol* 2003; 21: 942-7.
- Shepperd S, Charnock D. Against internet exceptionalism. *BMJ* 2002; 324:556-7.
- Bichakjian CK, Schwartz JL, Wang TS, Hall JM, Johnson TM, Biermann JS. Melanoma information on the Internet: often incomplete--a public health opportunity? *J Clin Oncol* 2002; 20:134-41.
- Biermann JS, Golladay GJ, Greenfield ML, Baker LH. Evaluation of cancer information on the Internet. *Cancer* 1999; 86:381-90.
- Anderson J, McKemmish S, Manaszewicz R. Quality criteria models used to evaluate health websites. In: *Proceeding of the 10th Asia Pacific Special, Health and Law Librarians' Conference*. <http://conferences.alia.org.au/shllc2003/papers/043.pdf> (accessed on 18/May/2012).
- Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. *J Epidemiol Community Health* 1999; 53:105-11.
- Health on the Net Foundation. The commitment to reliable health and medical information on the Internet. <http://www.hon.ch/HONcode/Patients/Visitor/visitor.html> (accessed on 01/May/2012).
- Frické M, Fallis D, Jones M, Luszko GM. Consumer health information on the Internet about carpal tunnel syndrome: indicators of accuracy. *Am J Med* 2005; 118:168-74.
- Ferreira L. A Internet como fonte de informação sobre saúde: um levantamento de percepções dos médicos portugueses [Masters Thesis]. Lisboa: Instituto Superior de Ciências do Trabalho e da Empresa; 2006.

10. Kunst H, Groot D, Latthe PM, Latthe M, Khan KS. Accuracy of information on apparently credible websites: survey of five common health topics. *BMJ* 2002; 324:581.
11. U.S. Preventive Services Task Force. Screening for prostate cancer. <http://www.uspreventiveservicestaskforce.org/prostatecancerscreening.htm> (accessed on 12/May/2012).
12. Nagler RH, Gray SW, Romantan A, Kelly BJ, DeMichele A, Armstrong K, et al. Differences in information seeking among breast, prostate, and colorectal cancer patients: results from a population-based survey. *Patient Educ Couns* 2010; 81 Suppl:S54-62.
13. Sociedade da Informação e do Conhecimento. Inquérito à utilização das tecnologias da informação e da comunicação pelas famílias. Lisboa: Instituto Nacional de Estatística; 2008. (Destaque: Informação à Comunicação Social).
14. Sociedade da Informação e do Conhecimento. Inquérito à utilização de tecnologias da informação e da comunicação pelas famílias: indivíduos dos 10 aos 15 anos. Lisboa: Instituto Nacional de Estatística; 2008. (Destaque: Informação à Comunicação Social).
15. Graphic, Visualization, and Usability Center. Gvu's 10th "www" user survey. [http://www.cc.gatech.edu/gvu/user\\_surveys/survey-1998-10/](http://www.cc.gatech.edu/gvu/user_surveys/survey-1998-10/) (accessed on 05/May/2012).
16. U.S. Preventive Services Task Force. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2009; 151:716-26.
17. Moyer VA; U.S. Preventive Services Task Force. Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2012; 157:120-34.
18. Advisory Committee on Cancer Prevention. Recommendations on cancer screening in the European Union. Vienna: Advisory Committee on Cancer Prevention; 1999.
19. Coordenação Nacional Para as Doenças Oncológicas, Alto Comissariado da Saúde, Ministério da Saúde. Plano Nacional de Prevenção e Controle das Doenças Oncológicas 2007/2010. Lisboa: Coordenação Nacional Para as Doenças Oncológicas, Alto Comissariado da Saúde, Ministério da Saúde; 2009.
20. Jenkins VL, Fallowfield L, Saul J. Information needs of patients with cancer: results from a large study in UK cancer centres. *Br J Cancer* 2001; 84:48-51.
21. Jansen B, Spink A. An analysis of document viewing patterns of web search engine users. Pittsburgh: School of Information Sciences, University of Pittsburgh; 2002.
22. Hargrave DR, Hargrave UA, Bouffet E. Quality of health information on the Internet in pediatric neuro-oncology. *Neuro Oncol* 2006; 8:175-82.
23. Hellawell GO, Turner KJ, Le Monnier KJ, Brewster SE. Urology and the Internet: an evaluation of internet use by urology patients and of information available on urological topics. *BJU Int* 2000; 86:191-4.
24. Suarez-Almazor ME, Kendall CJ, Dorgan M. Surfing the Net-information on the World Wide Web for persons with arthritis: patient empowerment or patient deceit? *J Rheumatol* 2001; 28:185-91.
25. Black PC, Penson DF. Prostate cancer on the Internet-information or misinformation? *J Urol* 2006; 175:1836-42.
26. Nilsson-Ihrfelt E, Fjällskog ML, Blomqvist C, Ahlgren J, Edlund P, Hansen J, et al. Breast cancer on the Internet: the quality of Swedish breast cancer websites. *Breast* 2004; 13:376-82.
27. Abbott VP. Web page quality: can we measure it and what do we find? A report of exploratory findings. *J Public Health Med* 2000; 22:191-7.
28. Meric F, Bernstam EV, Mirza NQ, Hunt KK, Ames FC, Ross MI, et al. Breast cancer in the World Wide Web: cross sectional survey of quality of information and popularity of websites. *BMJ* 2002; 324:577-81.
29. Quinn EM, Corrigan MA, McHugh SM, Murphy D, O'Mullane J, Hill AD, et al. Breast cancer information on the internet: analysis of accessibility and accuracy. *Breast* 2012; 21:514-7.
30. Wang L, Wang J, Wang M, Li Y, Liang Y, Xu D. Using Internet search engines to obtain medical information: a comparative study. *J Med Internet Res* 2012; 14:e74.
31. Mahé E, Qattini S, Beauchet A, Saiag P. Web-based resources for sun protection information: a French-language evaluation. *Eur J Cancer* 2009; 45:2160-7.
32. Jones K, Thibault J. Readability analysis: assessing and enhancing the readability of holocaust historical documents. Providing accessibility for high school students. Washington DC: Western Washington University; 2000.
33. Barboza E, Nunes E. A inteligibilidade dos websites governamentais brasileiros e o acesso para usuários com baixo nível de escolaridade. *Inclusão Social* 2007; 2:19-33.
34. Flesch R. How to write plain English. Chapter 2: let's start with the formula. <http://pages.stern.nyu.edu/~wstarbuc/Writing/Flesch.htm> (accessed on 11/Mar/2011).
35. Purcell G, Wilson P, Delamothe T. The quality of health information on the Internet: as for any other medium it varies widely: regulation is not the answer. *BMJ* 2002; 324:557-8.

---

Submitted on 12/Oct/2012

Final version resubmitted on 11/Jan/2013

Approved on 03/Jun/2013