Cancer on Facebook: life as it is

Abstract

**Background** Cancer is the generic name for a group of more than 200 diseases that have in common the disordered growth of cells that invade tissues and organs. The incidence of cancer is increasing worldwide, and one of the causes is the longer life expectancy. Nevertheless, between 30% and 50% of cancer cases are preventable and early detection contributes to a better prognosis. So, health communication strategies are essential and in times of web 2.0, Facebook, the world's largest social network in 2017, can be a useful tool to disseminate relevant messages on health promotion, prevention and early detection of cancer.

**Objective** We aim to offer ways for optimizing health communication strategies in the area of cancer on Facebook and to do this; we investigated which aspects of these messages generate greater engagement.

**Methods** We used Quintly, a social media-monitoring tool that allows the monitoring of many social media at the same time, even if the user is not an administrator of such pages. After that, we did a manual content analysis of publications and engagement rate. Finally, we developed software to optimized analysis of Facebook's posts. The tool aims to automate the analysis of any Facebook pages whose main theme is cancer. It is important to highlight that the number of pages that can be inserted for analysis is unlimited.

**Results** To verify what generates greater engagement in this area on Facebook, we studied 16 Brazilian pages whose main theme is cancer. The tool developed for this study allows metrics not publicly available as reach of the post (how many people viewed that post); post clicks (how many clicked to read the full text); post hides (how many people unveiled the page after reading the post, hiding the content of the post or denounced it as spam); likes; shares; comments (these last three are public information); engagement (number of clicks + likes + shares + comments); engagement rate (engagement divided by reach). The tool enables the creation of a ranking of posts, according to each of these metrics. The ranking can be created considering all the postings of the period or it can be filtered according to the category(s) to be analyzed.

**Conclusion** We saw that categories that generated greater engagement in Brazil are not those that have the highest percentage of publications.

**Keywords:** Facebook, health, cancer, content analysis, software

1. Introduction
Cancer is the generic name for a group of more than 200 diseases that have in common the disordered growth of cells that invade tissues and organs [1].

The number of cancer-related deaths worldwide rose from six million in 2000 to 7.6 million in 2007 [2]. In 2012, there were 8.2 million cancer deaths [3]. In 2016, 8.8 million people died due to cancer [4]. Brazil has an incidence rate of 205.5 cases of cancer per 100 thousand inhabitants, ranking tenth in South America and the Caribbean region [3].

With the development of Brazil and the consequent increase in the life expectancy of the population, in the middle of the 20th century, the incidence of cancer increased significantly. The number of older people in Brazil rose from 3 million in 1960 to 7 million in 1975, 20 million in 2008 and 30 million in 2018 [5] - an increase of almost 700% in less than 50 years. It is known that the biggest risk factor for most cancers is simply getting older [6].

Despite this scenario, it is important to reiterate that changes in lifestyle and habits of the population may decrease the likelihood of disease onset. “Only 5–10% of all cancer cases can be attributed to genetic defects, whereas the remaining 90–95% have their roots in the environment and lifestyle. The lifestyle factors include cigarette smoking, diet (fried foods, red meat), alcohol, sun exposure, environmental pollutants, infections, stress, obesity, and physical inactivity” [7]. Prevention is the most cost-effective, long-term strategy for controlling cancer [8].

In addition to the importance of adopting healthy lifestyles for prevention, it is of fundamental importance that there is an increase in early detection in individuals who already have symptoms of the disease. When some types of cancer are diagnosed in its early stages, the chances of treatment success and cure (for at least 5 years after diagnosis) increase dramatically. According to Cancer Research UK [9], some types of cancer can be treated much more easily in this case, for example: bowel cancer; breast cancer; ovarian cancer and lung cancer.

In order for people to be able to detect cancer at the earliest possible stage, it is important for them to be aware of the symptoms of the disease and its risk factors. Not always, however, this kind of knowledge is widely disseminated [10, 11,12,13,14]

**Facebook and health communication**

Facebook is currently the social network with the largest number of active users. In June 2017, Facebook has hit 2 billion monthly active users [15]. Every 60 seconds, 510 thousand comments are posted and 136 thousand photos are sent [16]. Although the pace of user growth has slowed in the last few years, engagement (content publishing and interaction with other users) has
intensified: more than 70% of Facebook users engage at least once a day and 45% do so several times a day [17]. The main forms of interaction are: liking, commenting and sharing publications.

Brazil is the third country with the largest number of Facebook users (123 million), only behind the United States (219 million) and India (213 million) [18]. In this way, it can be said that about 60% of the Brazilian population has an account in this network, considering that currently the population in this country is estimated at 207 million people [19].

There are numerous Facebook pages and groups that address health promotion, which are created around the world and in a wide variety of languages. According to Moorhead [20], the main benefits of using social media for health communication include:

(1) increased interaction with others, (2) more available, shared, and tailored information, (3) increased accessibility and widening access to health information, (4) peer/social/emotional support, (5) public health surveillance, and (6) potential to influence health policy [20].

Some studies have already looked at health pages on Facebook to verify the effectiveness of this communication strategy [21,22,23]. They noted that there is a great deal of user responsiveness to the topics posted on these pages, suggesting that there is still a lot of room for the growth of this type of discussion.

**Facebook and cancer**

We performed a search in Socialbakers, a website that monitors social media and produces a ranking, updated in real time, of Facebook pages with the largest number of followers. The search was performed through an analysis of this ranking. We looked for the page with the word ‘cancer’ in its name, which had the largest number of followers. The Facebook page about cancer with the greater number of followers is ‘Fight Cancer’, with around 632 thousand followers in May 2018. This page is, according to their description, ‘dedicated to all those who lost a loved one to cancer and to those fighting bravely this horrible enemy’ (Fight Cancer, 2018)¹. In second place is ‘Fuck Cancer’ with 340 thousand followers. This is a non-profit organization dedicated to ‘raising awareness, education of early cancer detection, and helping victims along with their families’ (Fuck Cancer, 2018)². In third is Stupid Cancer’, with 338 thousand followers. This is a non-profit organization dedicated to ‘empower, support, and improve health outcomes for the young adult cancer community’

---

¹ [https://www.facebook.com/pg/fight.cancer/about/?ref=page_internal](https://www.facebook.com/pg/fight.cancer/about/?ref=page_internal)

² [https://www.facebook.com/letsfcancer/](https://www.facebook.com/letsfcancer/)
When conducting the same search in Socialbakers for the Brazilian cancer pages, we noticed that the page with the largest number of followers is that of the hospital A.C. Camargo (386 thousand followers in May 2018). In second place is ‘Fundação do Câncer’ [Brazilian Cancer Foundation] page, with 268 thousand followers, and in third place is ‘Hospital do Câncer de Barretos’ [Barretos Cancer Hospital], with 276 thousand followers. The target audience of these pages is also cancer patients, as well as their friends and family. The Brazilian Facebook page with the largest number of followers is that of soccer player Neymar, which gathers more than 60 million users.

There is a gap in the scientific literature regarding the use of Facebook as a platform for disseminating health messages focused on the treatment of cancer or its prevention. One of the papers that deal with this theme [24] studied about 13 thousand comments posted by visiting users in three Brazilian pages whose main theme is cancer. It was observed that there is a strong presence of comments with religious terms, such as: God, faith, Lord, blessed, save and pray. In addition, women made most of the comments, and the positive content of the messages prevailed.

Another article [14] looked at the Facebook page of the National Cancer Institute in the United States to find out which strategies are most effective in engaging the audience. They reviewed the posts and comments made on this page between July 2010 and February 2015, and found the following:

    Audience engagement is associated with the format of cancer-related posts. Specifically, photo posts received significantly more likes, comments and shares than videos, links, and status updates\(^4\). The findings have important implications for how social media can be most effectively used to promote public engagement in important public health issues [14].

Despite the studies presented in this topic, there is still little evidence in the literature on the best ways to engage the public in health communication messages, both in Brazil and abroad. Regarding the area of cancer, the academic analyzes are even scarcer. For this reason, it is possible that Facebook pages of hospitals, non-governmental organizations (NGOs) and

\(^3\) https://www.facebook.com/stupidcancer/

\(^4\) On Facebook, status updates are posts than contain only texts.
informational organizations, among others, are reaching a lower percentage of the audience than their potential. Therefore, the present work intends to:

(1) Offer ways for optimizing health communication strategies in the area of cancer on Facebook, (2) Investigate which aspects of these messages generate greater engagement. On Facebook, this metric is calculated according to the number of reactions, shares, and comments for a post.

2. Method

This is a qualitative-quantitative study [25], with descriptive purpose [26] and that did not start from an *a priori* hypothesis.

**Analysis of 16 Brazilian Facebook pages about cancer**

To verify what generates greater engagement in the area of cancer in Facebook, we studied 16 Brazilian pages whose main theme is cancer.

To choose the pages of this study, we proceeded as follows:

a) In Facebook's internal search, we type the word ‘cancer’ and choose the ‘pages’ option;

b) We disregard pages that are not in Portuguese and those that approach ‘cancer’ as an astrological sign;

c) We selected 11 pages in the following categories: personal pages, newsletters, hospitals / foundations and NGOs;

d) To make this selection, we first consider the pages with the largest number of followers, then we looked at the updates - the page, to enter this survey, should have at least two weekly updates in this four-week cut.

To complete the 16 pages, we selected the three pages that were studied by us [27]. These pages are among those with the highest number of followers in Brazil (in the area of cancer). Our objective in reanalyzing them is to verify if the communication patterns found in that year have changed or have been maintained. Lastly, we also analyzed a Facebook page produced by us whose purpose is to inform the public about prevention and early diagnosis of cancer. This page is called ‘Acubens, museu de câncer’ (in English: ‘Acubens, cancer museum’). It is important to highlight that we did not select pages that specifically addressed prevention or early detection. Our intention is to check how the most popular Brazilian cancer pages deal with this subject.

For this analysis, we used Quintly (quintly.com), a social media-monitoring tool that allows the monitoring of many of these media at the same time, even if the user is not an administrator of such pages. Quintly organizes the publicly available information of all pages (number of followers, likes, comments, and
shares) in charts and tables that show, for example, how many new followers have progressed and the number of posts made in the selected period. This service also provides us with the complete listing of the postings of all selected pages, in a table with date, time, type of post: whether photo (considered as photo any image file), video, event (invitations to events, with the option to accept or decline, created within Facebook), status (text only) or link (posting has some web address redirecting to an external page). Facebook itself provides these post type definitions.

**Content analysis of publications and engagement rate**

The analysis of the texts of the Facebook posts was made based on the methodology proposed by Laurence Bardin [28]. This is a type of inductive analysis [29]. Two researchers performed this analysis independently.

The process of content analysis comprises: (1) Pre-analysis, which is the careful and systematic reading of all publications, in order to identify existing categories in the texts, (2) Categorization, which consists of creating categories so that all the individual posts fit into at least one of them. In this study, the two researchers created their categories independently and, in a second moment, they came together to produce a final listing of categories. When a category was created by one researcher and not by another, discussions were held between them until a final agreement was reached, (3) Interpretation, which consists of studying the data and developing inferences [27].

After this process, the two researchers also created, separately, a list of keywords of each category. Words could not repeat themselves in more than one category. In addition, very generic words that could fit into any category, such as ‘cancer’ and ‘chemotherapy’, were not considered. After each researcher created their lists independently, they gathered to see similarities and differences between their choices. Finally, a unique list was created, based on the consensus between them.

To gain a more holistic view of the categories, we also established the total impact of each of them, which we call ‘engagement rate’. This value considered the three metrics on each page. We calculated the weighted average likes, shares and comments of each post in the 16 pages, giving a weight of 0.05 for likes, 0.2 for shares and 0.75 for comments [27].

The weights created for the calculation of total engagement considered the fact that the type of engagement (liking, commenting, or sharing) follows a hierarchy according to the type of effort employed by the user to undertake it. Liking a post is considered low engagement because, among the three available actions, it is the simplest and quickest to execute. Sharing is considered as medium engagement because in this case the Facebook user identifies with the page in such a way that they want to share the content on their own personal
page in that social network. Finally, we consider comment as high engagement. In this case, the Facebook user needs to reflect on the topic in question, draft a text and state their opinion publicly. Facebook gives more weight to comments than to other interactions [30]. That means that when someone comments on a Facebook page, the likelihood that a friend of this user will receive this content in their news feed is greater than if that same user just likes certain content.

**Elaboration of a system that automates the analyzes**

Our previous analysis of Facebook pages [27] and this second served as the basis for the development of a tool that aims to automate the analysis of any Facebook pages whose main theme is cancer.

The tool developed is software. The system was created in JavaScript® and we could organize the posts of one or more Facebook pages according to metrics, some publicly available, and others provided only to the administrators of the pages. The number of pages that can be inserted for analysis is unlimited. Metrics not publicly available are: reach of the post (how many people viewed that post); post clicks (how many clicked to read the full text); post hides (how many people unveiled the page after reading the post, hiding the content of the post or denounced it as spam); likes; shares; comments (these last three are public information); engagement (number of clicks + likes * 0.05 + shares * 0.2 + comments * 0.75); engagement rate (engagement divided by reach). The tool enables the creation of a ranking of posts, according to each of these metrics. The ranking can be created considering all the postings of the period or it can be filtered according to the category(s) to be analyzed.

In addition, we also created within this software a bank of categories and a dictionary of keywords, which were developed by the researchers in the first phase of this work. This list is editable, that is, at any time we can add or remove categories and words. Our system can only "read" complete words; it does not consider compound words nor the radicals of words. That is, in the keywords list has all the possible variations of the same word, in the singular, in the plural, in the masculine and feminine.

This system is also able to predict what the engagement rate of any text entered by the user will be based on the engagement rates of the previous posts of a given page. That is, if a user-typed text has keywords that have generated high engagement in previous posts, the likelihood of this new post also having high engagement is bigger.

**3. Results and discussion**

The 16 Facebook pages analyzed by us produced a total of 712 posts in the study period. The pages were organized in groups, according to the profile to
which they belong (personal, informative pages, hospitals / foundations and NGOs).

In our previous study [27], we analyzed the publications made by three Brazilian pages about cancer during a period of six months (January to June, 2014) and created eight categories, which are: ‘Testimonies/real life stories’ (when people write about their experiences in relation to cancer, or any real-life story), ‘Solidarity’ (occurring when the post asked people to make some donation, such as blood or hair), ‘Anniversaries’ (when the main subject of the post was the celebration of some important date), ‘Science and health’ (scientific discoveries, academic novelties and advances in treatments), ‘Events’ (when the person in charge of the administration of the page organized or helped to publicize some event), 'Institutional’ (when the institution wrote about itself), ‘Risk factors’ (texts addressing habits that increase the risk of cancer, such as cigarettes) and ‘Beauty’ (texts about makeup, clothes and hairstyles).

By adding new pages to current analysis, we do not feel the need to create new categories. This fact is an indication that, however diverse the authors of the pages are, the spectrum of topics within the theme of cancer remains similar.

When analyzing the performance of the pages according to the groups to which they belong, we have the results presented in table 1.

[INSERT TABLE 1]

We can see that the categories that generated greater engagement are not those that have the highest percentage of publications. In the ‘hospitals / foundations’ group, the ‘Anniversaries’ category generated an engagement of 977.06, almost five times higher than the second one, ‘Science and Health’. However, only 6.03% of the publications of the period deal with such special dates. The category with the most publications in this group is ‘Institutional’ (49.13%). In relation to engagement, however, such a category is in third place.

In relation to the NGO group, once again the pattern repeats itself. The category that generates the greatest engagement is ‘Anniversaries’ (129,13), but only 5.12% of publications fall into this category. In this group of pages, the most frequent category is ‘Solidarity’, with 48.71% of the publications. However, the average engagement is 45.03, almost three times lower than the most successful category and fourth place in the average engagement ranking.

Another category with low presence in the publications is ‘Science and Health’. This category, along with ‘Risk Factors’ is directly related to topics such as cancer prevention, well-being and early diagnosis. In the groups of NGOs and personal pages, nothing was published on the subject. However, in the group of hospitals and foundations, this category ranks second in the average engagement ranking. People looking for pages of hospitals and foundations are
likely to be more interested in the subject than people looking for other cancer page profiles. For this reason, we suggest that page managers of hospitals or Facebook foundations devote more space to this subject.

Furthermore, it is important that the pages care about making their users comment more often, since this is the type of participation that demands greater intellectual effort. Users who comment invest more time in a publication. By requiring more from the reader, this is likely to be the reason why the average number of comments in all categories (except “Testimonies/real life stories”) is less than the average number of shares and reactions.

Some strategies used to generate more comments on Facebook are: asking users questions and responding to all comments [31]. "The more a user interacts with a particular content producer, the greater the chances of that producer appearing in the user news feed" [32]. For the user to interact more, it is important, therefore, that the page encourages such action.

**Facebook analytics software development**

The software developed for this analysis of Facebook posts and its classification into categories has a simple and intuitive interface.

Figure 1. The "posts" tab of the software, which displays the complete list of page posts.

[INSERT FIGURE 1]

Figure 2. "Heat map" tab, which displays the days of the week and times with more engagement on a particular page. The darker the square, the greater the engagement.

[INSERT FIGURE 2]

Figure 3. ‘Word cloud’ tab of the software, showing the most frequent words on a given page (the bigger the word, the higher the frequency).

[INSERT FIGURE 3]

Initially, we included in the tool the eight categories we created, and entered the keywords that correspond to each of them. Afterwards, we analyzed the page "Acubens, cancer museum", a page created by us at Oncobiology Program at Federal University of Rio de Janeiro, during six months, in order to verify if the tool can actually frame the posts in the correct categories. In this period, the page has published 163 posts.

In the first stage of this investigation, two researchers classified all the posts manually. Then, the result of the manual classifications was compared with the classification performed automatically by the software. In this way, the
researchers were able to verify if the tool was able to hit the categories of the posts.

Of the 163 posts, 34 were classified by the tool in the wrong categories. This is an error rate of 20.8% (79.2% accuracy). Table 2 shows the result of the automated analysis performed by the software and also shows the number of errors found according to the category. The errors are based on the comparison between the manual analysis done by the researchers and the analysis performed by the software.

This percentage of errors is considered acceptable. According to the literature [33,34,35,36,37], the final accuracy of multiclass text classification (when texts can be classified into three or more categories) ranges from 46.9% to 83%.

4. Limitations

In relation to the limitations of the software, we can mention the difficulty in choosing the words that will belong to each category, since some of them could belong to more than one. Often we had to make choices based on the evaluative criteria of the researchers. However, other people, with different experiences and distinct ways of writing, could classify words into other categories. Another limitation is the fact that the software does not understand the context and therefore does not capture ironies, jokes, ambiguous words or figurative language.

Despite the limitations, we believe that our software can help many research groups or Facebook page administrators to better understand what their audience wants and what generates more engagement in this social network. Other software features such as "heat map" will also be of great value in this process.

5. Conclusions

Categorizing posts and then calculating engagement rates has revealed that the potential of pages is often underutilized. This is because the categories that generate the greatest engagement are not the most frequent. On the contrary, we have noticed in some cases that the most attractive category for the public is among the least published.

It was interesting to note that there is low number of publications in the "Science and Health" category, but we believe this needs to change. Along with "Risk Factors," these are the categories that most relate to relevant public health issues such as cancer prevention, early diagnosis, and well being. The vast
majority of cancer cases are related to the environment and lifestyle, therefore the importance of talking about prevention and risky attitudes. Personal pages and NGOs during this study did not produce any messages about "Science and Health". NGOs have also not produced any content on "Risk Factors". However, there is a demand for this type of approach. In the group of pages of hospitals of foundations, the category "Science and Health" was the one that generated the second greater average engagement, only behind "Anniversaries". However, only 17.24% of the posts, within this group of pages, fell into this category. Our suggestion is that page managers of hospitals or Facebook foundations give more space to this subject.

The software developed for this research can help cancer page administrators on Facebook analyze their posts more easily. In this way, it will be possible to better understand which categories generate more engagement, so that the content is adapted in order to increase the public interest in the posts.

REFERENCES


