Whose Cancer? Visualizing the Distribution of Mentions to Cancer Sites on Instagram

Varela-Rodríguez, Miguel[a] and Vicente-Mariño, Miguel[b]

ABSTRACT: This article presents a quantitative analysis of mentions to cancer on Instagram. Using thousands of images with cancer-related hashtags, we build several visualizations to capture their distribution. Source images are clustered by their visual traits and by the incidence, prevalence, and mortality of the cancer site they refer to. Our goal is three-fold: to provide a quantitative basis for future research on the representation of cancer online; to offer an interpretation of the sources of the imbalanced representation of the different cancer sites; and to motivate a debate on how that representation may affect patients and families.

Keywords: *health communication; public understanding of science and technology; visual communication; social media and cancer*

[a]*Department of Sociology and Social Work, University of Valladolid, Valladolid, Spain.* [*miguel.varela@uva.es*](mailto:miguel.varela@uva.es). [*https://www.linkedin.com/in/miguel-varela/*](https://www.linkedin.com/in/miguel-varela/); [b]*Department of Sociology and Social Work, University of Valladolid, Valladolid, Spain.* [*miguel.vicente@uva.es*](mailto:miguel.vicente@uva.es). [*http://sociologiaytrabajosocial.sitios.uva.es/?q=node/58*](http://sociologiaytrabajosocial.sitios.uva.es/?q=node/58)

Corresponding author: Miguel Varela-Rodríguez. Department of Sociology and Social Work, Faculty of Commerce, University of Valladolid, Plaza Campus Universitario, 1, 47011 – Valladolid, SPAIN. miguel.varela@uva.es. https://www.linkedin.com/in/miguel-varela/

Introduction: why cancer, why social media, and why cancer on Instagram?

The internet is now the second main source of information for cancer patients in the US, only following the information provided directly by doctors (Blanch-Hartigan & Viswanath, 2015). For patients, social media can serve as the gate to a community of support and information, and as a tool for self-expression (Braun et al., 2019; Chou et al., 2011; Crannell et al., 2016; Noar et al., 2018). In particular, images have been found to be an important vehicle for a patient-owned discourse of illness (Pardo, 2019).

Visual social media are today amongst the fastest-growing and most-used globally, especially for younger audiences (Pew Research Center, 2019). While online photographs and other images are increasingly used by cancer patients to share and obtain information (Struck et al., 2018), existing research has mainly focused on text and speech analysis.

This paper aims to visualize the distribution of images on Instagram that make explicit reference to cancer on Instagram, one of the leading visual social media. It is framed within a broader project on “The Shape of Illness”, which seeks to analyze the representation of cancer in visual social media, how it shapes social narratives, and the emotional impact it has on patients and their social circles.

This first stage addresses the following two questions:

1. What is the distribution of mentions to each cancer site on Instagram images?
2. How is the number of images mentioning a cancer site on Instagram related to its incidence, prevalence, and mortality rates?

The data visualizations and the correlations presented in the Method and Results sections will allow us to respond to both questions.

The function of social media for cancer patients and researchers

A growing body of work highlights the functions that online communities and social media can have for cancer patients. In general, these studies are based on the understanding that patients have a strong desire and need for information after diagnostic (Hawkins et al., 2008, p. 10). But networks like social media go beyond information-sharing: they also provide a means to interact with other patients personally, build a community around shared interests, and receive (and provide) emotional support (Moorhead et al., 2013). This search for togetherness is an important function of online networks (see Firth et al., 2019; Ridings & Gefen, 2004) that has positive results for many patients (Attai et al., 2015).

On the other hand, the fast-paced nature of social media and the “attraction mechanism” that determines their visibility and promotion through algorithms (Firth et al., 2019, p. 120) leads some images to triumph over others. The “winners”—that is, the images that capture our attention our achieve the highest number of “likes”—typically show faces (Bakhshi et al., 2014), are aesthetically pleasing, look professional, and inspire positive emotions (Tifentale & Manovich, 2018). As users and companies follow the “rules” to achieve a likeable image on Instagram, a standardized visual discourse begins to emerge. This process has been visible on mass media for several decades with the widespread use of the survivorship discourse around breast cancer, which has also been identified online. The use of such standardized narratives has been identified online and found to affect the emotional well-being of patients negatively (Banerjee et al., 2018; Pertl et al., 2014).

An early exploration of the nexus between social media and cancer is the work of Chou et al. (2011). Theirs is an account of the cancer experience of young YouTube users. Among their findings, they note that “the success of cancer communication efforts depends largely on creating emotional engagement with message content” (Chou et al., 2011, p. 7). Similarly, by Gibson et al. (2016) insist on the importance of storytelling to understand the experience of younger patients and adapt informational and emotional support. Their findings are consistent with general approaches to community-building in social media: emotions and personal contact build lasting relationships.

The importance of social media for cancer patients is discussed also by Sugawara et al. (2012), who highlight the potential of Twitter as a connector. The authors argue that “Twitter could be a valuable medium for sharing information among cancer patients” (2012, p. 5), particularly as users share daily messages about their treatment and life with cancer more generally, helping normalize reactions to treatment and providing patients with a safe space.

From a clinical perspective, the last ten years have seen a surge of research into the functions of social media in cancer treatment. Zaid et.al. (2014) discuss how such platforms can help identify patients and accelerate surveying processes; Attai et al. (2015) remark that social media is “fertile territory” for cancer research and patient education (2015, p. 3); Bottorff et al. (2014) explore the use of social media in tobacco prevention campaigns; Bravo and Hoffman-Goetz (2016) assess mentions to prostate cancer in social media, which Carneiro and Dizon (2019) expand on. Banerjee et al. (2018) put emphasis on the online space as a source of support for patients of melanoma; Lee-won et al. (2017) assess the impact of virality measures on screening intention, while Noar et al. (2018) review how a single, viral photograph uploaded by a skin cancer patient resulted in a peak of Google searches for this type of cancer. Taylor and Plagiari (2019) scanned Facebook and Twitter for discussions touching on lung cancer. Like them, close to 100 academic papers were devoted to the topic by 2015 (Koskan et al., 2014; Moorhead et al., 2013), focusing mostly on case studies, and speech and text analysis.

Despite the growing wealth of work, systematic reviews suggest that there is room for improvement. Relying on top posts and small samples often means that studies focus on the most visible cancer sites and neglect sites with comparable rates of incidence but less visibility online (Crannell et al., 2016, p. 539; Döbrössy et al., 2020, p. 12; Han et al., 2018; Moorhead et al., 2013, p. 5). In doing so, they are subject to the effect of the attraction mechanism. A deeper review of the representation of each cancer site online can help fill the gap and examine how online mentions are distributed and why. Understanding the social roots of the representation of cancer in social media

Some of the papers reviewed for this work attempt a quantitative comparison of the presence of different cancer sites on social media and note the overrepresentation of breast cancer as compared to sites with similar incidence or mortality rates. Further, Sugawara et al. “found it interesting that the cancer prevalence of [their] influential users and the general population are so dissimilar” (2012, p. 5), signaling a distribution of cancer sites online that does not match that of their epidemiology. Crannell et al. speak to this imbalance, too, while specifying that the most tweeted cancers are breast, lung, prostate and colorectal (Crannell et al., 2016, p. 538).[1]

In a 2018 study on Instagram and Twitter traffic related to cancer, Vraga et al. (2018) observe that the campaign for Movember outperforms breast cancer campaigns only in the month of November while breast cancer dominates discussions the rest of the year. They link this seeming seasonality to the decades-long development of breast cancer awareness raising campaigns and their inclusive nature, which help give them visibility during the year, while the male-focused, physical and actionable nature of Movember gathers a great degree of attention during the month of November, especially in visual networks like Instagram.[2] Importantly, Vraga et al. concede that "[w]hile the Movember campaign generates traffic in November, Movember is not encouraging communication about prostate cancer" (Vraga et al., 2018, p. 8), similarly to findings by Bravo & Hoffman-Goetz (2016).

Common explanations to the unequal visibility of different cancer sites are based on epidemiology and social media demographics. Breast cancer has a lower average age of incidence and is one of the most prevalent cancer sites globally, which helps its presence online (Sugawara et al., 2012, p. 5). A similar interpretation can be seen in Crannell et al., who note that "the fact that the breast cancer was the toptweeted cancer was not surprising, considering breast cancer is one of the most prevalent cancer types" (Crannell et al., 2016, p. 538). This argument stands when looking at the next top-tweeted cancer sites in Crannel et al., with high prevalence rates in the United States, but does not align with our findings. Through a correlation test, we will show that the volume of content produced discussing other cancer sites with high prevalence, like lung or prostate, is not significantly comparable to that of breast cancer.

---

1 Findings from our study for Instagram do not reveal the same distribution for Instagram mentions, as can be seen in the results section.

2 While it started as a movement to give visibility to prostate cancer, the Movember movement encourages men to grow a moustache or a beard as a show of support to issues related to men's mental and physical health more generally. See www.movember.com for more information.

Another interpretation offered by Sugawara et al. and in Hartigan & Visnawath (2015) is that social media activity is related to well-being and the length of treatment. Breast cancer is typically accompanied by longer-term treatment and higher survival rates, allowing patients the strength and the time to post their experiences on social media (Sugawara et al., 2012, p. 6). Breast cancer also presents generally higher rates of prevalence and lower rates of mortality than other types of cancer that evolve quickly and fatally; consequently, the assumption is that breast cancer patients have more of an opportunity to use social media during treatment. Were that the case, it should be possible to observe a clear correlation between prevalence and online presence for a cancer site, a conclusion that does not emerge from our results either.

We will argue that a statistically significant correlation between online presence and incidence, prevalence or mortality cannot be concluded on Instagram from the data obtained.

While results from previous studies shed light on specific groups and cancer sites, the approach to social media and cancer would benefit from a general understanding of the volume of mentions to cancer online, their visibility, and their distribution. Doing so would help enable a better understanding of the impact that cancer-related communications can have on patients, family, and friends, before defining possible interventions. Indirect exposure to cancer-related content (through social media campaigns, events, or popular culture) may create expectations in future patients and relatives, which are important to account for in health education (Fishman et al., 2010). While there is an opportunity to use existing online infrastructures to build communities of patients and offer continuous care, it is paramount to analyze the distribution of cancer online and question the perceptions that it may lead to.

Materials and Method

Attempting a study of “cancer” as a global keyword on Instagram would yield too many results for analysis. Instead, this study mixes manual and automated data extraction to obtain a sample of photographs that can help sketch a global image of cancer in English and Spanish-speaking Instagram.

Identification of profiles for keyword extraction

A first step covers the identification of relevant profiles, used to reveal the hashtags most used when posting images related to cancer. These first few profiles include both organizational and individual accounts.

The geographic reach of this initial search is limited to Spain, where the Spanish Society Against Cancer (AECC, *Asociación Española Contra el Cáncer*) leads civil society efforts on cancer research and support to patients.[3] The profile was observed for a week in April 2019, giving Instagram's Explore enough time to suggest similar profiles and images for the researchers to follow.

Posts suggested by the Explore page were mainly focused on breast cancer, a site that has great presence online (Sugawara et al., 2012). To allow for other sites to be included in the sample, additional accounts were introduced manually.

The final sample for observation is made of 49 different profiles, with accounts posting in Spanish, English, and Portuguese (see table 1). The handles of individual accounts were not registered to protect the privacy of their owners and appear as "N/A" in Table 1. All accounts were reviewed manually for relevance and only public accounts were followed.

Table 1. Initial sample of 49 profiles from which cancer hashtags were obtained

Instagram Handle	Topic	Type
acancerprostata	Prostate Cancer	Support
aeacap	Lung Cancer	Support
aecc_es	All	Support
agradecidaypoderosa	All	Patient
cancer_de_pancreas	Pancreatic Cancer	Support
cancermetanoia	Breast Cancer	Patient
cr_uk	All	Support
donatupelodonatupelo	All	Support
ejercicio_fisico_oncologico	All	Support

3 Spanish social media served as a starting point, expanding the search globally in subsequent phases.

fundacionaladina	All	Support
grupagata	Breast Cancer	Support
hayguerrerosparato	All	Patient
macmillancancer	All	Support
mugronets	Breast Cancer	Support
N/A	All	Patient
N/A	Breast Cancer	Patient
N/A	Breast Cancer	Support
N/A	Breast Cancer	Patient
N/A	Breast Cancer	Patient
N/A	Breast Cancer	Patient
N/A	Breast Cancer	Patient
N/A	Breast Cancer	Patient
N/A	Breast Cancer	Patient
N/A	Breast Cancer	Patient
N/A	Lymphoma	Patient
N/A	Lymphoma	Patient
N/A	Breast Cancer	Patient
N/A	Breast Cancer	Patient
N/A	Breast Cancer	Patient
N/A	Breast Cancer	Patient
N/A	Lung Cancer	Patient
N/A	Breast Cancer	Patient
N/A	Breast Cancer	Patient
N/A	Breast Cancer	Patient
N/A	Breast Cancer	Patient
N/A	Lymphoma	Patient
N/A	Breast Cancer	Patient
notetapesdotcom	Breast Cancer	Support
oncolliga	All	Support
oncowellness	All	Support
psicooncologia_para_todos	All	Support
quimioencolombia	All	Support
savethemama	Breast Cancer	Support
standup2canceruk	All	Support
tatuajesolidario	Breast Cancer	Support
thecure_forcancer	All	Support
unoentrecienmil	Leukemia	Support
venci_el_cancer	All	Patient
wecanbeheroeses	Breast Cancer	Support

Note: Names of patients' accounts have been removed to protect their privacy.

Identification of keywords

Over the course of two weeks in April 2019, the sample profiles returned a series of phrases commonly used when mentioning cancer in online posts. The list included specific mentions to cancer sites as well as phrases or slogans (such as “fuck cancer”. “cancer sucks” or “breast cancer awareness”). The resulting list includes 9 different sites and 11 phrases:

A. Cancer Sites:

(1) *Lung (#LungCancer)*
(2) *Breast (#BreastCancer)*
(3) *Pancreas (#PancreaticCancer)*
(4) *Leukemia*[4] *(#Leukemia)*
(5) *Prostate (#ProstateCancer)*
(6) *Colorectum (#ColorectalCancer)*
(7) *Colon (#ColonCancer, merged with #ColorectalCancer in the analysis)*
(8) *Stomach (#StomachCancer)*
(9) *Uterus (cervix and corpus, commonly referred to online as #WombCancer)*

B. Cancer Keywords:

(1) *Cancer Research (#CancerResearch)*
(2) *Contra El Cancer* (*#ContraElCancer*, “against cancer”, in Spanish)
(3) *Fuck Cancer (#FuckCancer)*
(4) *Cancer Survivor (#CancerSurvivor)*
(5) *Cancer Sucks (#CancerSucks)*
(6) *Movember (#Movember)*
(7) *Breast Cancer Awareness (#BreastCancerAwareness)*
(8) *Breast Cancer Survivor (#BreastCancerSurvivor)*
(9) *Lung Cancer Awareness (#LungCancerAwareness)*
(10) *Pancreatic Cancer Awareness (#PancreaticCancerAwareness)*
(11) *Cancer Free (#CancerFree)*

Extraction of data for descriptive visualization

Manual searches were conducted for each of the 20 hashtags in the list between 1

4 The American spelling for leukaemia was selected as it returned a higher number of hits. Some of these types of cancer returned virtually no results, but the list was based on metrics of prevalence and mortality for each type of cancer, to test whether high prevalence translated into high presence in social media.

and 20 December 2019, and images downloaded manually from the Explore page using screen scrapers available in the Google Chrome Store (see Varela-Rodríguez & Vicente-Mariño, 2021, 2020). For each search, only public images were collected, with a minimum of 300 images per cancer site and phrase (Table 2).

Table 2. Total images collected in the sample per site and phrase.

Hashtag	Number of Images Collected
Cancer Site	
Breast Cancer	1,484
Colon Cancer	909
Colorectal Cancer	1,004
Leukemia	1,042
Lung Cancer	585
Pancreatic Cancer	331
Prostate Cancer	1,128
Stomach Cancer	1,241
Womb Cancer	327
Cancer-related phrase	
BreastCancerAwareness	1,558
BreastCancerSurvivor	738
CancerFree	1,338
CancerResearch	1,036
CancerSucks	772
CancerSurvivor	1,037
ContraElCancer	581
FuckCancer	1,398
LungCancerAwareness	773
Movember	1,278
PancreaticCancerAwareness	904
TOTAL	19,464

In total, 19,464 images were collected, renamed, cleared of metadata, and stored in an offline hard drive. Only the images and their associated keywords were stored, removing any information about the user profile. At no point were the original images released to the public. No additional metadata (such as location) was collected, as it is not made available by Instagram.

Content review and visualization

This first dataset allows for a visual representation of results using polar graphs, following visualization methodologies close to those developed by Lev Manovich and the Software Studies Initiative (Manovich, 2011).

Our method differs significantly from Manovich's approach to computational analysis in cultural studies. In AI Aesthetics, Manovich advocates for the need to "learn to see cultures in more detail, without immediately looking for, and noticing, only types, structures or patterns" (Manovich 2018, 384). Our study does make use of structures and categories to narrow down the search for images and allow for descriptive analysis that sheds light on the different types of cancer represented online. In future phases, however, it is expected that categorizations will lose importance, especially as the research looks at emotions and social discourse.

Once stored and labelled, all images were put through the image-processing software developed by the Software Studies Initiative[5] to extract quantitative information about their hue, brightness, and saturation. These data help identify clusters of color that can reveal a visual identity for certain hashtags or cancer sites.

The results are reported using descriptive tables, distribution plots, and a series of polar graphs that collect all the images and arrange them according to different indicators. Two types of polar graphs feature in the results to answer the two research questions:

1. *What is the distribution of mentions to each cancer site on social media?* A polar graph visualizes all the images in the sample. Distance from the center of the graph indicates the dominant hue of the image: mostly red images along the inner edge, mostly blue and violet images around the middle of the radius, and mostly pink images along the outer edge. Each radius corresponds to the hashtag that returned the image, which is indicated in text. A second polar graph shows the same images, but adjusted to the total number of images

5 Mainly ImageMeasure, bundled in the ImagePlot pack provided at http://lab.softwarestudies.com/p/imageplot.html

returned by each hashtag. Distance from the center in this second graph indicates the total amount of hits returned by the search for each hashtag.

2. *How is the number of images mentioning a cancer site related to its incidence, prevalence, and mortality rates?* Values for cancer sites on prevalence, incidence and mortality are superposed on the previous graphs, allowing for an instant visualization of the tags that are most prominent on Instagram and how they compare to their rates. In addition, distribution plots show data relevant to the correlation tests.

The original montages online can be zoomed-in to a point where the viewer can ascertain the presence of individuals, animals, or hospital equipment, but where it is not possible to identify the people in them. For print purposes, a zoomed-in extract is provided.

Statistical analysis

While the profiles observed only returned nine cancer sites for image collection, numerical data was collected also for 31 different sites listed in the Cancer Dictionary of the Global Cancer Observatory (International Agency for Research on Cancer, 2021):

Bladder	Kidney	Mesothelioma	Pancreas	Uterus (Cervix or Corpus)
Brain	Larynx	Myeloma	Penis	Vagina
Breast	Leukemia	Nasopharynx	Prostate	Vulva
Colorectum	Lip-oral	Non-Hodgkin's Lymphoma	Sarcoma	
Esophagus	Liver	Nonmelanoma	Stomach	
Hodgkin's Lymphoma	Lung	Oropharynx	Testis	
Hypopharynx	Melanoma	Ovary	Thyroid	

Each of the 31 sites was manually searched for on Instagram on 22 January 2020, collecting the number of results returned for each (variable Hits). In addition, data on prevalence, incidence and mortality was collected from the Global Cancer Observatory (World Health Organization, 2020). Since no further data was collected on the images, it is not possible at this stage to categorize the images by location or other demographic indicators.

A correlation test was performed for the variables Hits, Incidence, Prevalence and Mortality on SPSS, returning their Pearson correlation coefficient as well as distribution plots. Correlations were tested including all cancer sites at first and then excluding the case of breast cancer, as it appeared to behave as an outlier in the data.

Results

1. What is the distribution of mentions to each cancer site on social media?

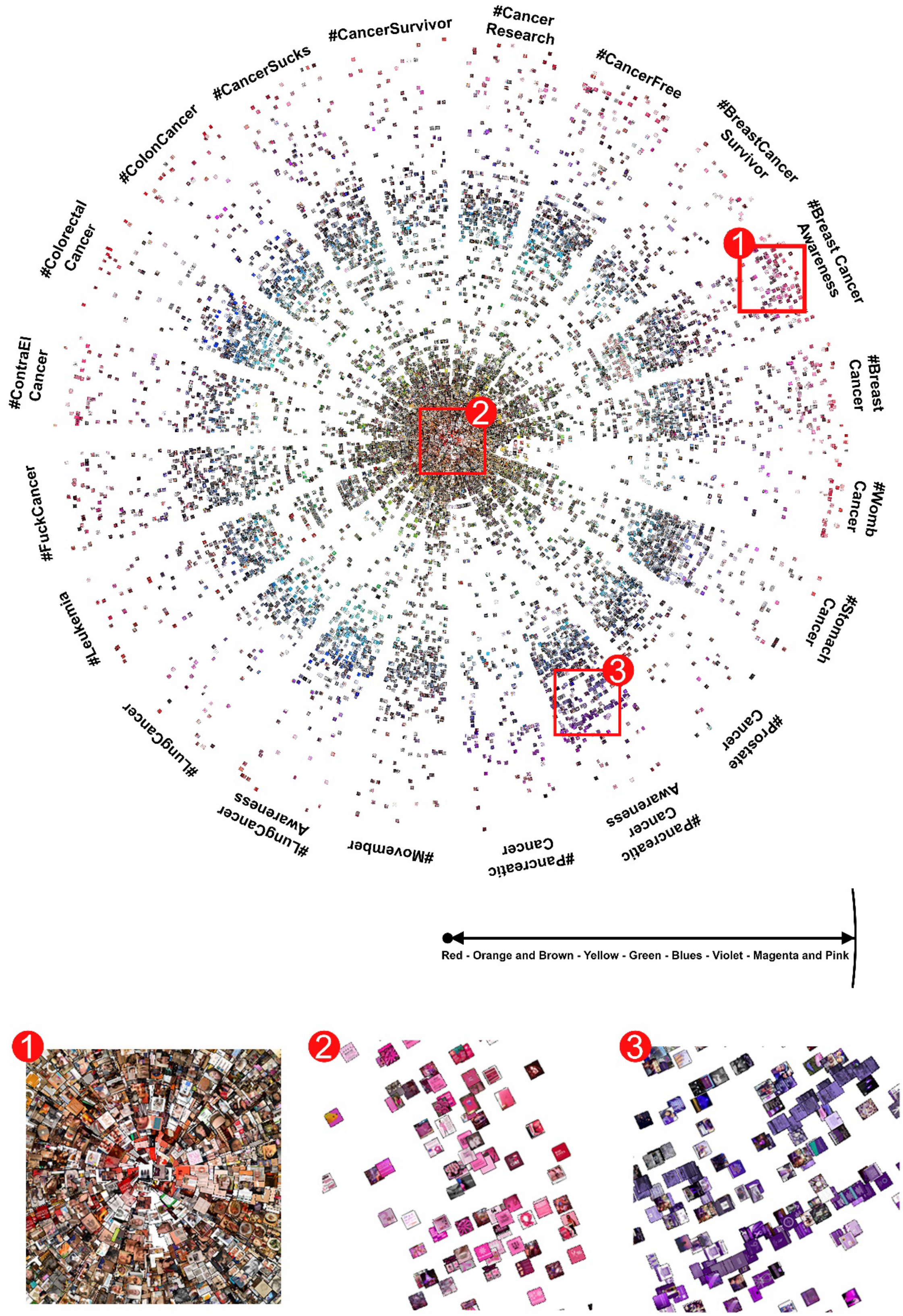
Visual traits

On their own, the images collected tell individual stories of recovery, hope and fear. Collectively, they reveal patterns about the representation of cancer on social media (Figure 1).

Most images in Figure 1 appear concentrated around the middle of the graph: these are portraits, group photos and anatomical images that contain hues closer to orange and brown emerging from the colors of the skins in the picture. Another cluster appears in the “blues” section (half-radius): these are often photographs taken outdoors, with the blue sky in frame. There is some concentration in the “greens”, too, from pictures taken outdoors (around trees and on grass) and uploaded by the AECC (whose corporate color is green).

Importantly, however, the graph reveals that posts associated with breast cancer make more use of highly saturated magentas and pinks, through pink ribbons and banners associated with breast cancer prevention and research campaigns. But it is not the only hue strongly associated with a given cancer site in the dataset: violet hues appear in photographs labelled as “pancreatic cancer”, a color that has been used to campaign for more research, while dark blue is linked to campaigns to raise awareness on lung cancer. Figure 1 shows that these two campaigns have a strong visual identity, like those related to breast cancer.

Figure 1. Visual distribution of images in the sample per cancer site and phrase



Distribution per hashtag

While Figure 1 can reveal patterns about color, it does not account for total hits, giving the impression that tags like “Pancreatic Cancer Awareness” have similar visibility to “Breast Cancer Awareness”. To compensate this fact, Figure 2 is adjusted to show the total number of images on the Explore page for each cancer site (Tables 3 and 4).

Table 3. Results per cancer site on Instagram (search term: “#[CancerSite]”)

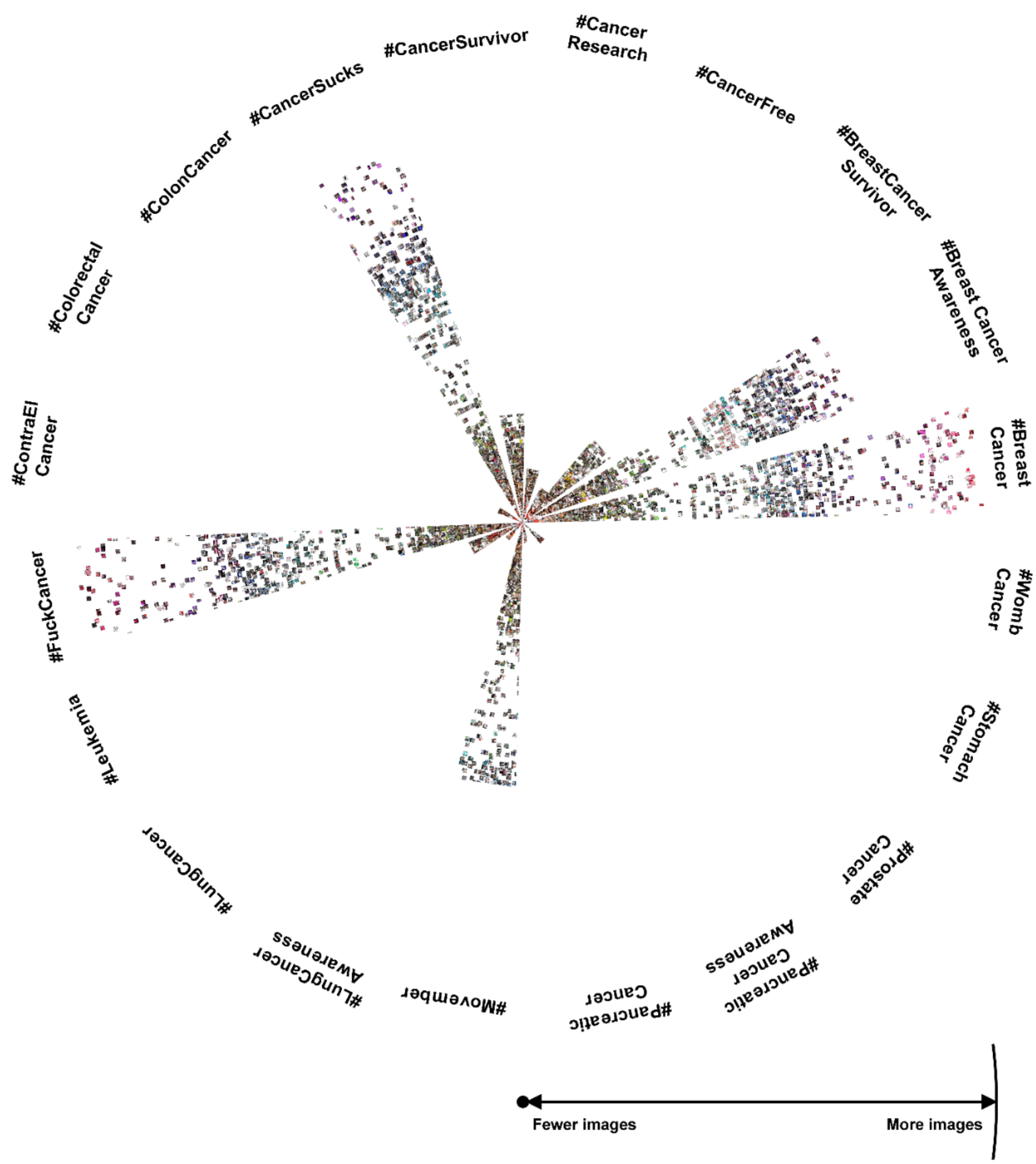
Site	Hits
Bladder	13,294
Brain	117,436
Breast	2,313,834
Uterus (Cervix or Corpus)	1,554
Colorectum	148,581
Esophagus	1,627
HodgkinsLymphoma	80,788
Hypopharynx	0
Kidney	19,562
Larynx	392
Leukemia	244,593
Lip-oral	20,028
Liver	22,112
Lung	112,823
Melanoma	176,836
Mesothelioma	12,664
Myeloma	18,378
Nasopharynx	7
Non-Hodgkins Lymphoma	24,016
Nonmelanoma	228
Oropharynx	22
Ovary	379
Pancreas	107,711
Penis	106
Prostate	146,097
Sarcoma	49,679
Stomach	12,361
Testis	49,558
Thyroid	108,169
Vagina	181
Vulva	1,053

304

Table 4. Results per phrase on Instagram (search term: "#[Phrase]")

Tag	Hits
BreastCancerAwareness	2,595,655
FuckCancer	3,405,325
CancerSucks	2,870,578
Movember	1,926,142
CancerSurvivor	796,272
CancerResearch	395,925
cancerfree	288,427
BreastCancerSurvivor	280,880
PancreaticCancerAwareness	35,257
ContraElCancer	26,837
LungCancerAwareness	26,385

Figure 2. Volume of images on Instagram per cancer site and phrase



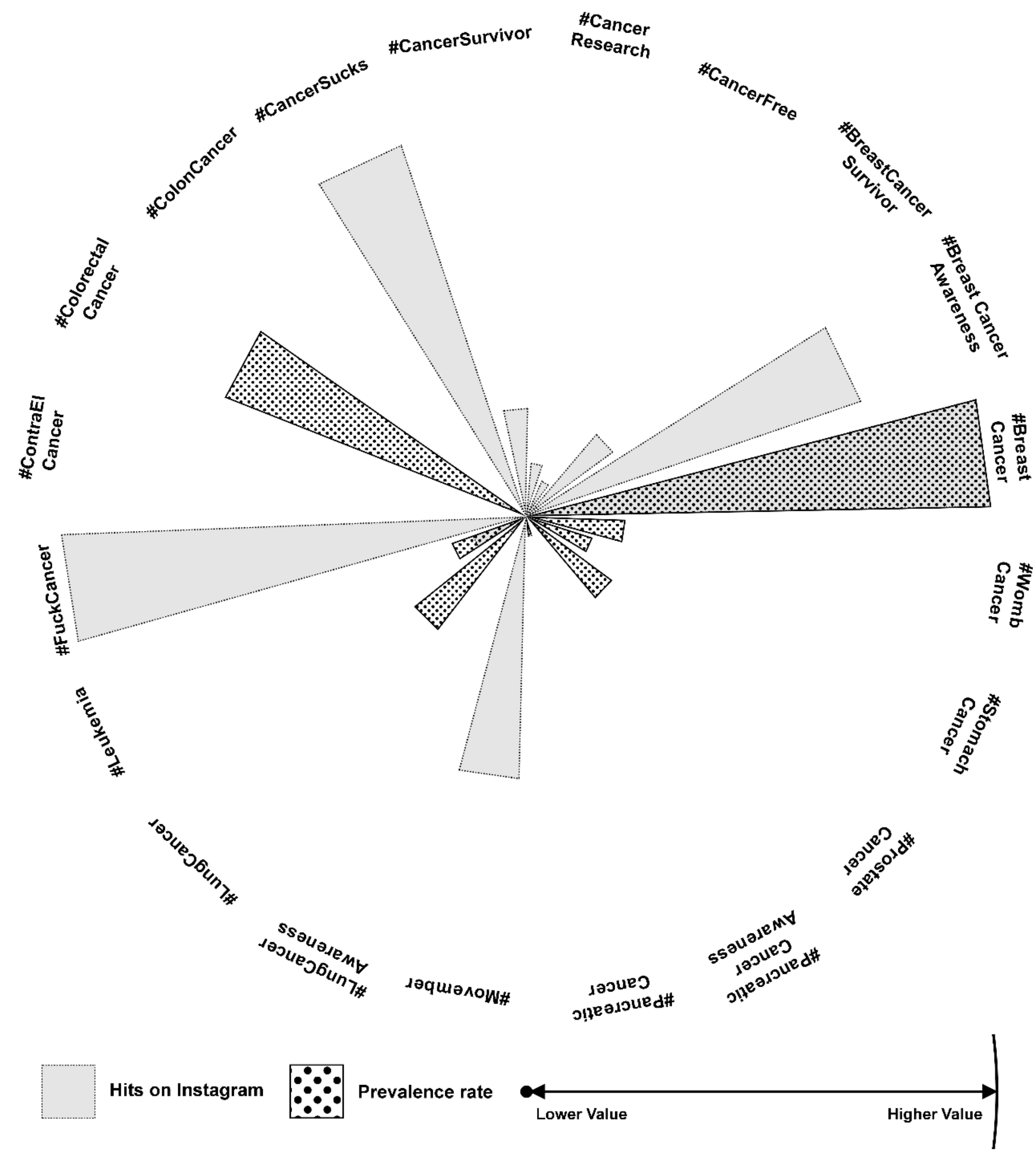
If Figure 2 were to encapsulate the visibility of cancer on Instagram, only one cancer site would be distinctly identifiable: breast cancer. On Instagram, there are almost ten times more posts with the hashtag “#breastcancer” than posts with the hashtag “#leukemia”, which follows as second most mentioned, and 20 times more than highly prevalent and mortal sites such as lung cancer.

Meanwhile, the only other visible tags would be general phrases such as “cancer sucks”, “fuck cancer”, “breast cancer awareness” and “movember”, with the latter being the only one that does not show results linked to breast cancer.

How is the number of images mentioning a cancer site related to its incidence, prevalence, and mortality rates?

Figure 3 overlays the prevalence of each cancer site (>5 years) in the sample with the number of images returned on Instagram, to test the hypothesis that this overrepresentation may be linked to the prevalence of breast cancer.

Figure 3. Volume of images on Instagram (Hits) and Prevalence rate (>5 years) for the cancer site represented



Indeed, with the highest prevalence rates of the cancer sites collected in the sample, breast cancer also shows the highest number of images of all cancer sites studies. Yet another cancer site with high prevalence rates, colorectal cancer, returned virtually no results.

A Pearson correlation test using all 31 cancer sites rejects the null hypothesis of a lack of correlation between Prevalence and number of images on Instagram (Hits), seemingly indicating that the higher presence of breast cancer online is related to its higher long-term prevalence rate:

$$R\ (29)\ =\ .66, p\ <\ .01$$

However, without accounting for breast cancer, the results of the test change dramatically, and the null hypothesis of a lack of correlation can no longer be rejected:

$$R\ (28)\ =\ .23, p\ <\ .221$$

Looking at both Figure 3 and the distribution plots below (Figures 4 and 5), it appears that breast cancer behaves as an outlier, with an abnormal number of hits returned. This behavior is relevant to understanding the imbalance in representation of cancer online.

Figure 4. Distribution plot of Hits vs Prevalence (>5 years, 2018) per cancer site

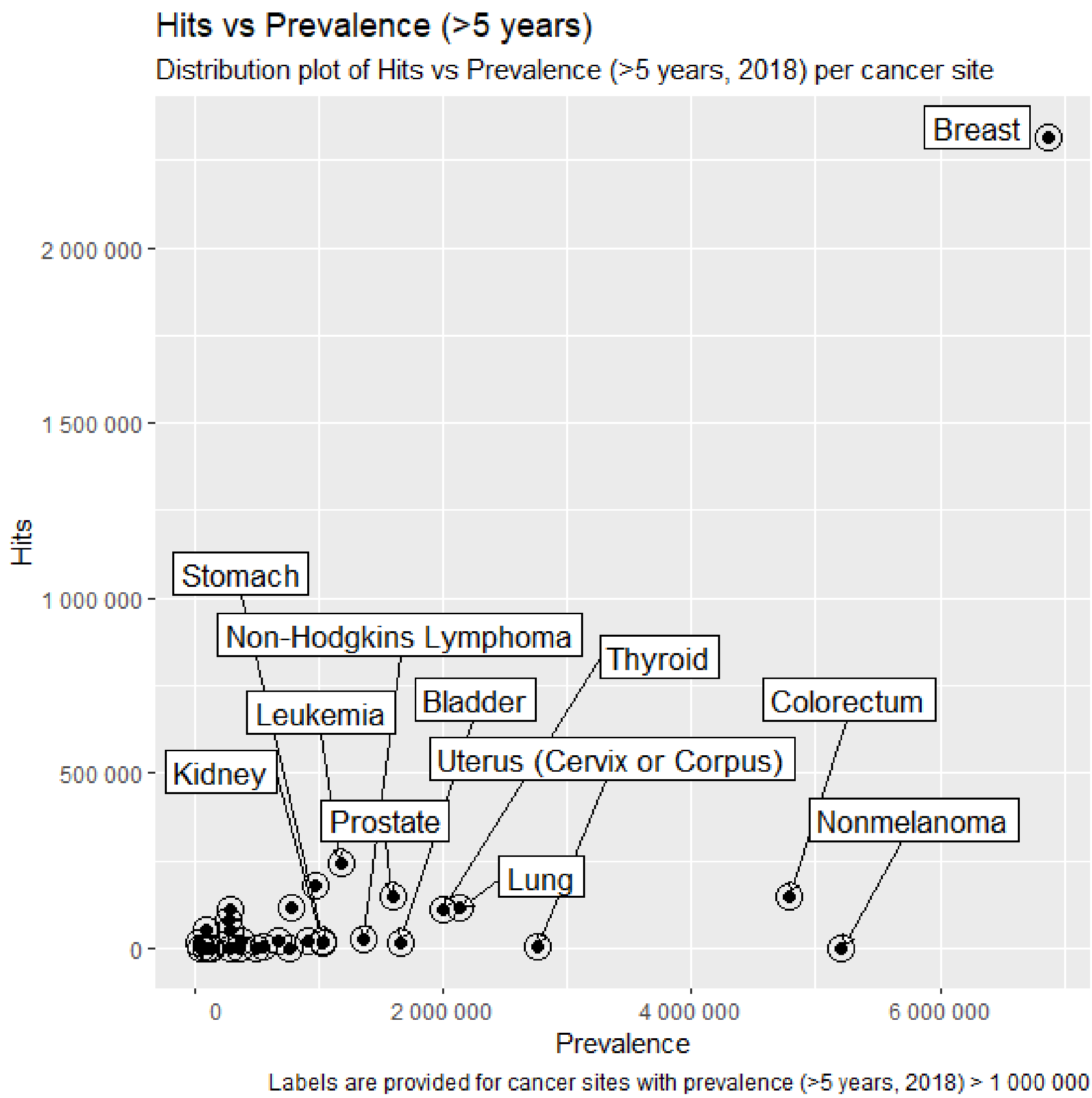
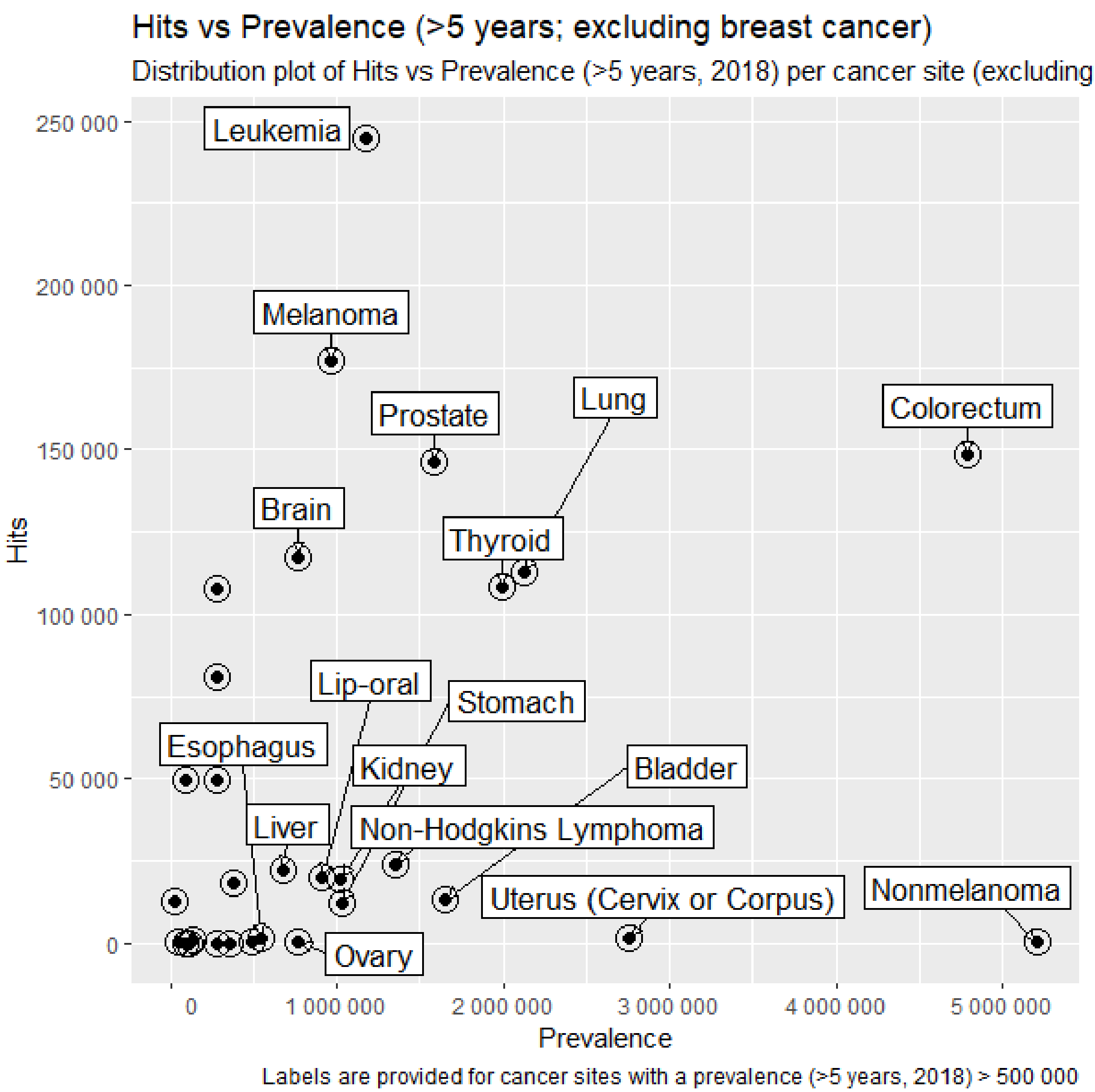
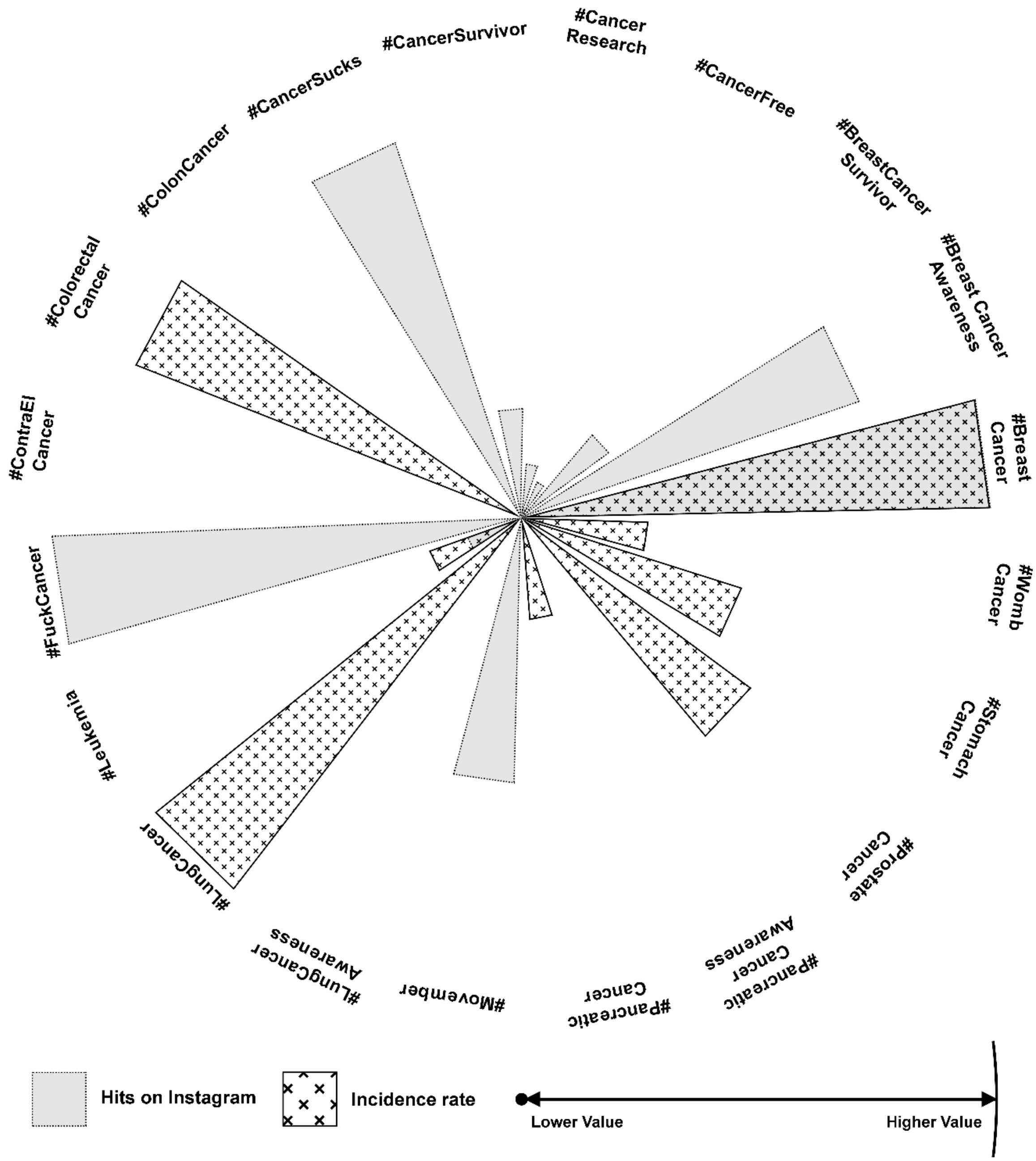


Figure 5. Distribution plot of Hits vs Prevalence (>5 years, 2018) per cancer site, excluding breast cancer



Similarly, the literature points to incidence as a possible reason for the higher visibility of breast cancer. Yet contrasting the number of results per site with the incidence of each cancer site does not appear to show a clear correlation (Figures 6, 7 and 8).

347 **Figure 6. Volume of images on Instagram (Hits) and Incidence rate (2018) for the cancer site**
348 **represented**



349

350 In 2018, lung cancer had the highest incidence rate worldwide, followed by breast
351 cancer and colorectal cancer. Once again, breast cancer dwarves all other cancer sites
352 in terms of images on Instagram. A Pearson correlation test for Hits on Instagram and
353 Incidence (2018) cannot reject the null hypothesis of a lack of correlation between the

two, indicating a possible correlation where higher incidence rates lead to more visibility online:

$$r\ (29)\ =\ .523, p\ <\ .01$$

As with prevalence, however, the significance of the test is lost (although by a small margin) when breast cancer is removed from the equation:

$$r\ (28) = .348, p\ = .06$$

Figure 7. Distribution plot of Hits vs Incidence rate per cancer site

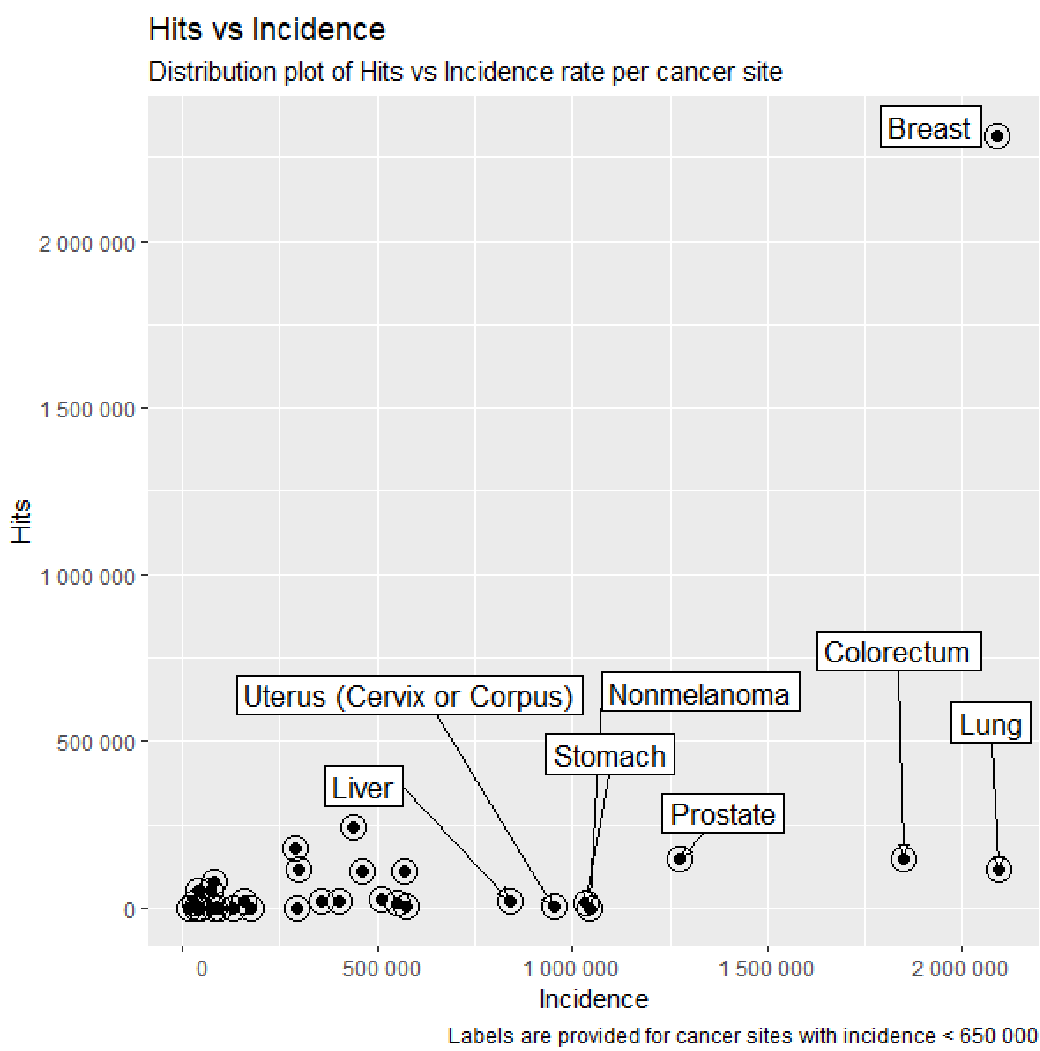
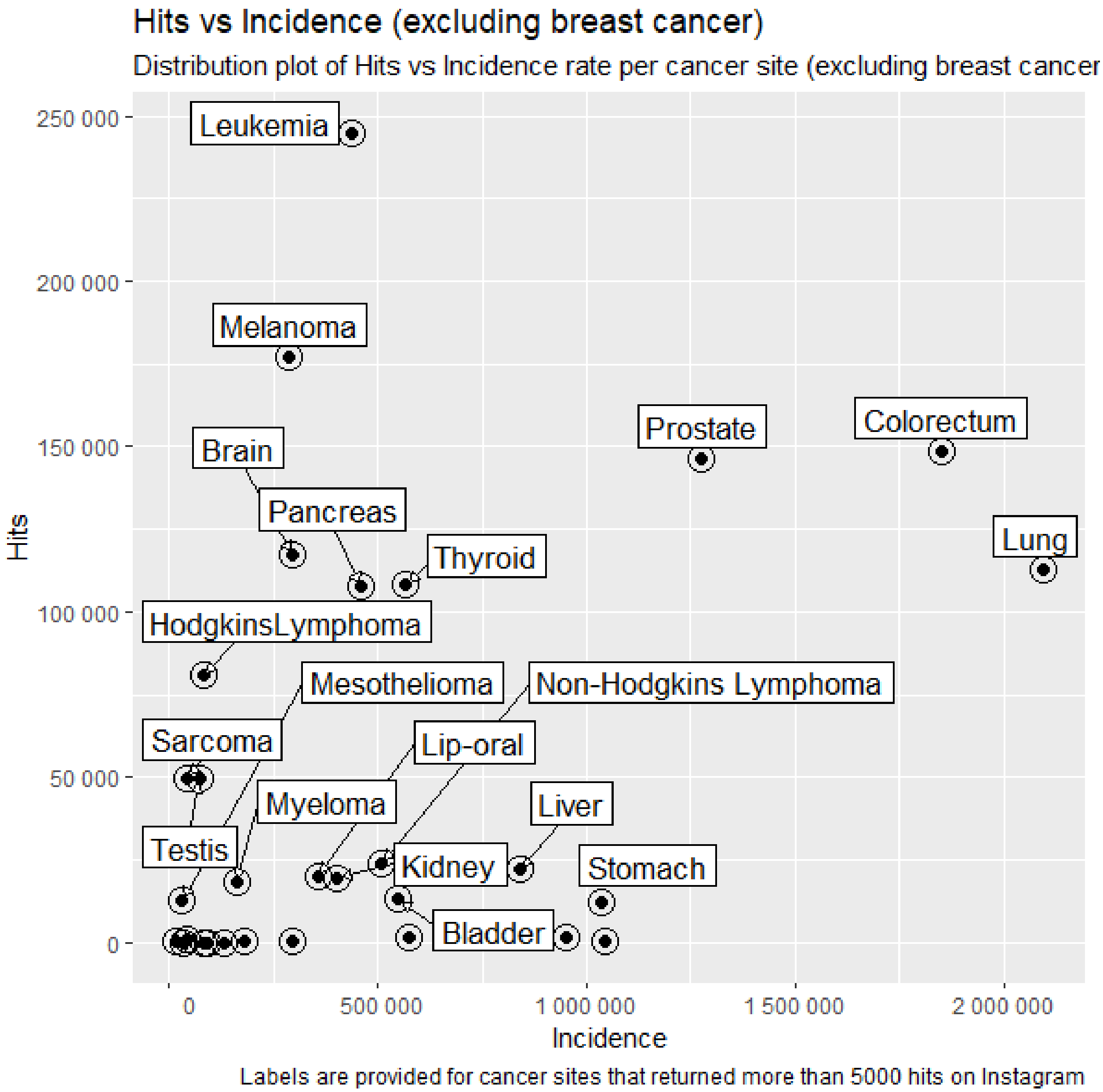
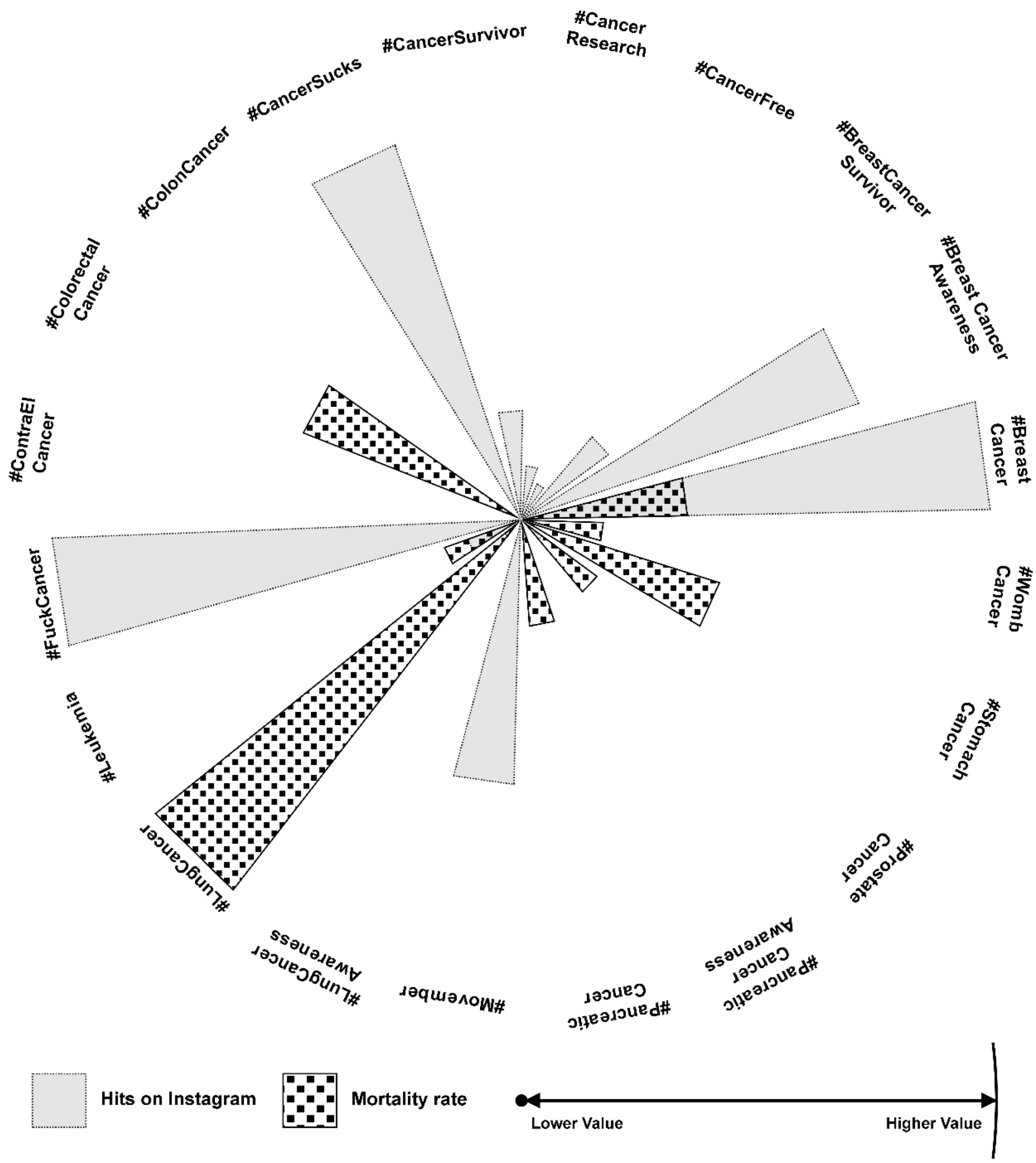


Figure 8. Distribution plot of Hits vs Incidence rate per cancer site, excluding breast cancer



Finally, a visualization contrasting Mortality with Hits presents the most contradictory image (Figure 9).

Figure 9. Volume of images on Instagram (Hits) and Mortality rates (2018) per cancer site



While lung cancer, stomach cancer and colorectal cancer have the highest mortality rates, they also have the lowest visibility online. This indicates that online visibility is not linked to the net human cost of the illness and may validate Sugawara's interpretation that online presence may be linked to survival rates. A Pearson correlation cannot reject the hypothesis of a lack of correlation between these two variables, be it maintaining breast cancer in the data:

$$r\,(29) = .212, p = .251$$

or removing it from the equation:

$$r\,(28) = .276, p = .140$$

Figure 10. Distribution plot of Hits vs Mortality rate per cancer site

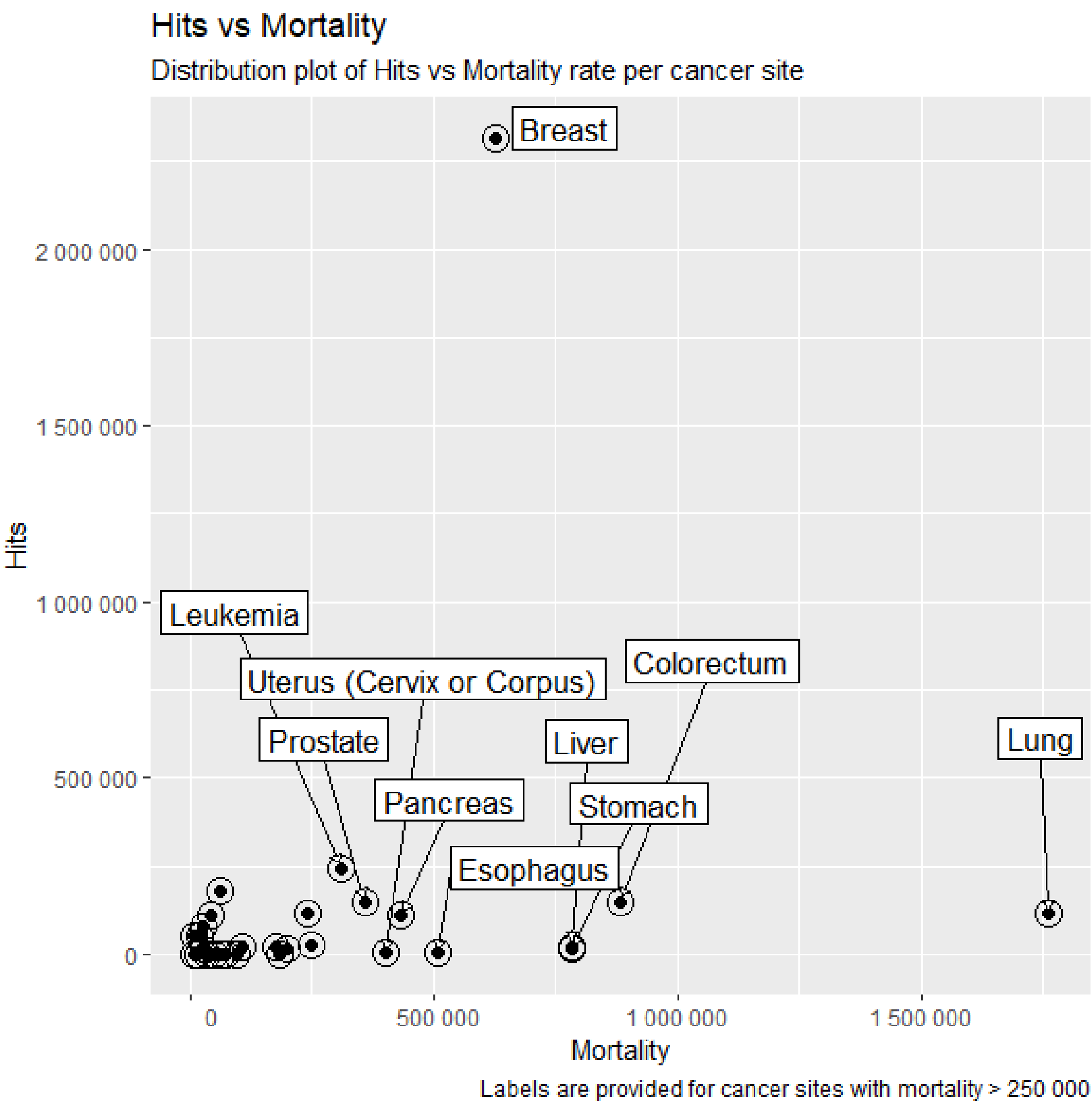
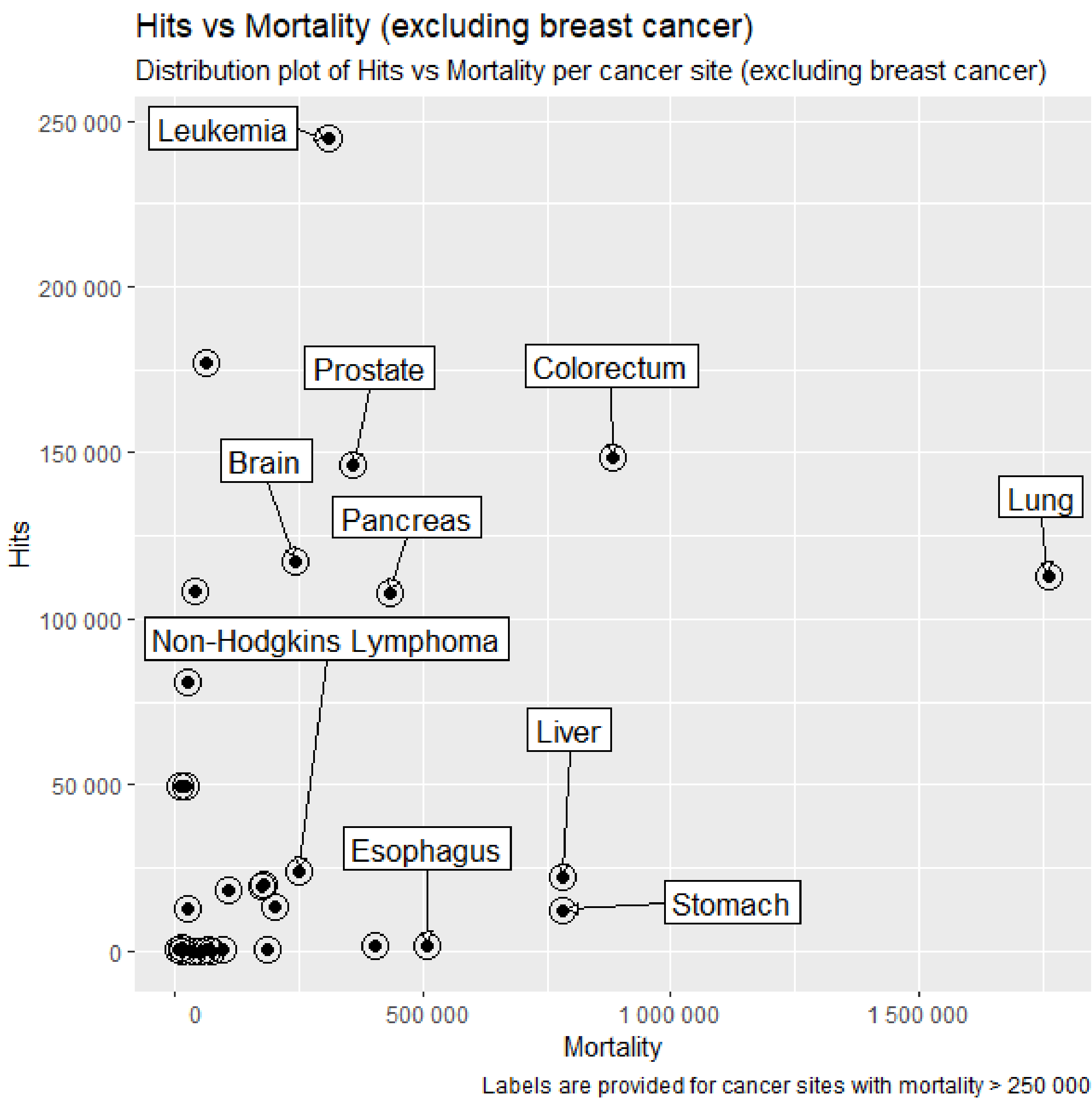


Figure 11. Distribution plot of Hits vs Mortality rate per cancer site, excluding breast cancer



Consequently, it cannot be concluded that online visibility of a given cancer site is related to its incidence, prevalence nor mortality rates.

Discussion

Previous research already indicated a dominance of breast cancer in online discussions—findings in this paper help visualize such imbalance and give it visual entity. That breast cancer appears to behave as an outlier in online representation is an important finding. It is the most mentioned type of cancer on Instagram, outweighing the next in line (leukemia) by a factor of almost ten. Meanwhile, sites with high incidence like lung or prostate cancer appear to have little visibility on Instagram.

The explanation for this imbalance appears to be social rather than epidemiological, supporting previous findings in the literature. Discussing Twitter posts, Crannell et al. find that "only 1% of all breast cancer tweets were patient tweets" (Crannell et al., 2016, p. 539). Their interpretation is that "nondiagnosed individuals are tweeting about the disease for fundraising purposes or sharing feelings about a loved one" (Crannell et al., 2016, p. 539), revealing the greater public awareness of breast cancer as compared to other sites. This suggests that content discussing cancer sites with less public awareness may be falling into an echo-chamber and not reaching the public discourse, which may have implications for research funding and access to information by patients.

While the literature points to social media and the internet as a key source of information, community and support, the lack of representation of some cancer sites may dissuade patients from sharing content. While breast cancer is prominent online, other patients face the task of building a community and challenging common perceptions of cancer that do not necessarily apply to their experience.

The implications are both individual (for patients) and social. Firstly, adding to the burden of blame that patients of lung cancer and other sites carry with them, the lack of support and visibility online for their illness can be demotivating and detrimental to their emotional wellbeing.[6] Secondly, the higher visibility of certain sites may impact research funding, as individuals and organizations turn their attention to certain sites and neglect others.[7]

---

6 For instance, Pertl, Quigley and Hevey (2014) reflect on this challenge when discussing cancer-related fatigue and how it contradicts social discourses related to survivorship.

7 Much work has been done on this issue with relation to breast cancer. See Sweeney & Killoran-McKibbin (2016) or King (King, 2008).

Movember, a case of dissociation between campaign and illness

While not the only campaign for awareness present in the sample, Movember is the only hashtag that compares in results to breast cancer-related tags. The Movember campaign, launched in the early 2000s, has been historically tied to raising awareness on prostate cancer, but has since moved to other issues related to the mental and physical health of males. Data from this study returns 105,214 results on Instagram carrying the hashtag “#ProstateCancer”, while “#Movember” soars to close to two million mentions. This result supports previous research that outlines the dissociation between the aesthetics of awareness-raising campaigns and their intended impact (Bravo & Hoffman-Goetz, 2016; Vraga et al., 2018), in a fashion not too dissimilar to the critique that Ehrenreich (2001) would make of breast cancer awareness campaigns, which she found to have moved from cancer awareness to a “cult of pink kitsch”.

Limitations and opportunities for further research

This research is limited by the volume of data available online. Its method relies on mostly manual techniques, which may have led to important keywords being left out. Importantly, the hashtag “cancer” could not be studied in isolation due to the vast number of results and the use that is made of it to refer to other issues (such as the horoscope sign). Similarly, cancer-related content that is not hashtaged requires a deeper dive into the data.

Through observation, it is possible to detect the most used keywords when referring to cancer by a small sample of users, but it is not possible to collect them all. Similarly, Instagram does not allow for advanced searches, which impedes the use of limited time ranges or location-specific searches. Future research should seek to automate this process, enabling cross-references between hashtags and allowing for more hashtags (including more common names for certain cancer sites, such as “throat cancer”).

The demographics of social media need also be considered, with a higher number of users being adolescents and young adults, while the incidence of certain cancer sites (breast cancer included) is above the 50 years of age.

Further, social media access has been restricted in the last two years, which impedes the use of automated data extraction, ordering of images, or the collection of "likes" per image. These are all valuable research variables that should be sought in future work, particularly to assess the impact of different images, hashtags, and cancer sites. Researchers may explore a collaboration between Facebook/Instagram, data, social and medical scientists. This may allow for further data, such as the average duration of treatment, survival rates, and qualitative data on cancer discourse online.

Conclusion

This research indicates that the volume of mentions to a given cancer site on Instagram does not correlate to its prevalence, incidence nor mortality rates, contradicting some of the hypotheses in the literature. Instead, it demonstrates that breast cancer behaves as an outlier in social media, with a large volume of posts that appear to be motivated by the success of awareness raising campaigns.

Meanwhile, other cancer sites with high prevalence, incidence or mortality are virtually absent from Instagram. Notable cases are those of lung, prostate, pancreas, and stomach, some of the sites with highest mortality rates and lowest survival rates, and also with some of the lowest visibility online. Patients of some of these types of cancer (mainly lung cancer) are often burdened with social blame and find little support in the public online space. While initiatives to provide informational and emotional support to such patients using social media show positive results, it is important to bear this imbalance in mind to understand the impact that social media may have on them.

While it is beyond the scope of this study, the high visibility of breast cancer online merits further study on its causes, the shaping of social discourse on illness, and the impact on patients' well-being, in a similar vein to existing work done on Facebook (Chou et al., 2020).

In an ocean of information like social media, volume matters. Regardless of their features, campaigns to raise awareness will struggle to make themselves visible as they row against millions of images and the attraction mechanism. When using social media to connect patients, researchers must ask “whose cancer?” and anticipate the effects that exposure to content related to other sites may have, and how that content shapes the public discourse of cancer.

Disclosure of interest

The authors report no conflict of interest

References

Attai, D. J., Cowher, M. S., Al-Hamadani, M., Schoger, J. M., Staley, A. C., & Landercasper, J. (2015). Twitter Social Media is an Effective Tool for Breast Cancer Patient Education and Support: Patient-Reported Outcomes by Survey. *Journal of Medical Internet Research*, *17*(7). https://doi.org/10.2196/jmir.4721

Bakhshi, S., Shamma, D., & Gilbert, E. (2014). Faces engage us: Photos with faces attract more likes and comments on Instagram. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/2556288.2557403

Banerjee, S. C., D’Agostino, T. A., Gordon, M. L., & Hay, J. L. (2018). “It’s Not JUST Skin Cancer”: Understanding Their Cancer Experience From Melanoma Survivor Narratives Shared Online. *Health Communication*, *33*(2), 188–201. https://doi.org/10.1080/10410236.2016.1250707

Blanch-Hartigan, D., & Viswanath, K. (2015). Socioeconomic and sociodemographic predictors of cancer-related information sources used by cancer survivors. *Journal of Health Communication*, *20*(2), 204–210. https://doi.org/10.1080/10810730.2014.921742

Bottorff, J. L., Struik, L. L., Bissell, L. J. L., Graham, R., Stevens, J., & Richardson, C. G. (2014). A social media approach to inform youth about breast cancer and smoking: An exploratory descriptive study. *Collegian*, *21*(2), 159–168. https://doi.org/10.1016/j.colegn.2014.04.002

Braun, L. A., Zomorodbakhsch, B., Keinki, C., & Huebner, J. (2019). Information needs, communication and usage of social media by cancer patients and their relatives. *Journal of Cancer Research and Clinical Oncology*, *145*(7), 1865–1875. https://doi.org/10.1007/s00432-019-02929-9

Bravo, C. A., & Hoffman-Goetz, L. (2016). Tweeting About Prostate and Testicular Cancers: Do Twitter Conversations and the 2013 Movember Canada Campaign Objectives Align? *Journal of Cancer Education*, *31*(2), 236–243. https://doi.org/10.1007/s13187-015-0796-1

Carneiro, B., & Dizon, D. S. (2019). Prostate Cancer Social Media: In YouTube We Trust? *European Urology*, *75*(4), 568–569. https://doi.org/10.1016/j.eururo.2019.01.004

Chou, W.-Y. S., Hunt, Y., Folkers, A., & Augustson, E. (2011). Cancer Survivorship in the Age of YouTube and Social Media: A Narrative Analysis. *Journal of Medical Internet Research*, *13*(1). https://doi.org/10.2196/jmir.1569

Chou, W.-Y. S., Trivedi, N., Peterson, E., Gaysynsky, A., Krakow, M., & Vraga, E. (2020). How do social media users process cancer prevention messages on Facebook? An eye-tracking study. *Patient Education and Counseling*, *103*(6), 1161–1167. https://doi.org/10.1016/j.pec.2020.01.013

Crannell, W. C., Clark, E., Jones, C., James, T. A., & Moore, J. (2016). A pattern-matched Twitter analysis of US cancer-patient sentiments. *Journal of Surgical Research*, *206*(2), 536–542. https://doi.org/10.1016/j.jss.2016.06.050

Döbrössy, B., Girasek, E., Susánszky, A., Koncz, Z., Győrffy, Z., & Bognár, V. K. (2020). “Clicks, likes, shares and comments” a systematic review of breast cancer screening discourse in social media. *PloS One*, *15*(4), e0231422. https://doi.org/10.1371/journal.pone.0231422

Firth, J., Torous, J., Stubbs, B., Firth, J. A., Steiner, G. Z., Smith, L., Alvarez-Jimenez, M., Gleeson, J., Vancampfort, D., Armitage, C. J., & Sarris, J. (2019). The “online brain”: How the Internet may be changing our cognition. *World Psychiatry*, *18*(2), 119–129. https://doi.org/10.1002/wps.20617

Fishman, J., Ten Have, T., & Casarett, D. (2010). Cancer and the Media: How Does the News Report on Treatment and Outcomes? *Archives of Internal Medicine*, *170*(6), 515–518. https://doi.org/10.1001/archinternmed.2010.11

Gibson, F., Hibbins, S., Grew, T., Morgan, S., Pearce, S., Stark, D., & Fern, L. A. (2016). How young people describe the impact of living with and beyond a cancer diagnosis: Feasibility of using social media as a research method. *Psycho-Oncology*, *25*(11), 1317–1323. https://doi.org/10.1002/pon.4061

Han, C. J., Lee, Y. J., & Demiris, G. (2018). Interventions Using Social Media for Cancer Prevention and Management: A Systematic Review. *Cancer Nursing*, *41*(6), E19–E31. https://doi.org/10.1097/NCC.0000000000000534

Hawkins, N. A., Pollack, L. A., Leadbetter, S., Steele, W. R., Carroll, J., Dolan, J. G., Ryan, E. P., Ryan, J. L., & Morrow, G. R. (2008). Informational needs of patients and perceived adequacy of information available before and after treatment of cancer. *Journal of Psychosocial Oncology*, *26*(2), 1–16. https://doi.org/10.1300/j077v26n02_01

International Agency for Research on Cancer. (2021). *Global Cancer Observatory: Cancer Today*. https://gco.iarc.fr/today/data-sources-methods#cancer-dictionnary

King, S. (2008). *Pink ribbons, Inc.: Breast cancer and the politics of philanthropy*. University of Minnesota Press. https://amzn.com/dp/B00IK7WS4G

Koskan, A., Klasko, L., Davis, S. N., Gwede, C. K., Wells, K. J., Kumar, A., Lopez, N., & Meade, C. D. (2014). Use and taxonomy of social media in cancer-related research: A systematic review. *American Journal of Public Health*, *104*(7), e20-37. https://doi.org/10.2105/AJPH.2014.301980

Lee-Won, R. J., Na, K., & Coduto, K. D. (2017). The effects of social media virality metrics, message framing, and perceived susceptibility on cancer screening intention: The mediating role of fear. *Telematics and Informatics*, *34*(8), 1387–1397. https://doi.org/10.1016/j.tele.2017.06.002

Manovich, L. (2011). What is visualisation? *Visual Studies*, *26*(1), 36–49. https://doi.org/10.1080/1472586X.2011.548488

Moorhead, S. A., Hazlett, D. E., Harrison, L., Carroll, J. K., Irwin, A., & Hoving, C. (2013). A New Dimension of Health Care: Systematic Review of the Uses, Benefits, and Limitations of Social Media for Health Communication. *Journal of Medical Internet Research*, *15*(4), e85. https://doi.org/10.2196/jmir.1933

Noar, S. M., Leas, E., Althouse, B. M., Dredze, M., Kelley, D., & Ayers, J. W. (2018). Can a selfie promote public engagement with skin cancer? *Preventive Medicine*, *111*, 280–283. https://doi.org/10.1016/j.ypmed.2017.10.038

Pardo, R. (2019). Fotografía y enfermedad: Iconografías en transformación. In *La imagen desvelada: Prácticas fotográficas en la enfermedad, la muerte y el duelo* (pp. 19–60).

Sans Soleil Ediciones. https://www.sanssoleil.es/tienda/la-imagen-desvelada-practicas-fotograficas-en-la-enfermedad-la-muerte-y-el-duelo-montse-morcate-y-rebeca-pardo-ed/

Pertl, M. M., Quigley, J., & Hevey, D. (2014). 'I'm not complaining because I'm alive': Barriers to the emergence of a discourse of cancer-related fatigue. *Psychology & Health*, *29*(2), 141–161. https://doi.org/10.1080/08870446.2013.839792

Pew Research Center. (2019, June 12). Demographics of Social Media Users and Adoption in the United States. *Pew Research Center: Internet, Science & Tech*. https://www.pewresearch.org/internet/fact-sheet/social-media/

Ridings, C. M., & Gefen, D. (2004). Virtual Community Attraction: Why People Hang Out Online. *Journal of Computer-Mediated Communication*, *10*(1), 00–00. https://doi.org/10.1111/j.1083-6101.2004.tb00229.x

Struck, J. P., Siegel, F., Kramer, M. W., Tsaur, I., Heidenreich, A., Haferkamp, A., Merseburger, A. S., Salem, J., & Borgmann, H. (2018). Substantial utilization of Facebook, Twitter, YouTube, and Instagram in the prostate cancer community. *World Journal of Urology*, *36*(8), 1241–1246. https://doi.org/10.1007/s00345-018-2254-2

Sugawara, Y., Narimatsu, H., Hozawa, A., Shao, L., Otani, K., & Fukao, A. (2012). Cancer patients on Twitter: A novel patient community on social media. *BMC Research Notes*, *5*, 699. https://doi.org/10.1186/1756-0500-5-699

Sweeney, E., & Killoran-McKibbin, S. (2016). Selling Pink: Feminizing the Non-Profit Industrial Complex from Ribbons to Lemonaid. *Women's Studies*, *45*(5), 457–474. https://doi.org/10.1080/00497878.2016.1186492

Taylor, J., & Pagliari, C. (2019). The social dynamics of lung cancer talk on Twitter, Facebook and Macmillan.org.uk. *Npj Digital Medicine*, *2*(1), 51. https://doi.org/10.1038/s41746-019-0124-y

Tifentale, A., & Manovich, L. (2018). Competitive Photography and the Presentation of the Self. In J. Eckel, J. Ruchatz, & S. Wirth (Eds.), *Exploring the Selfie: Historical, Theoretical, and Analytical Approaches to Digital Self-Photography* (pp. 167–187). Springer International Publishing. https://doi.org/10.1007/978-3-319-57949-8_8

Varela-Rodríguez, M., & Vicente-Mariño, M. (2021). Imágenes desgarradas: El uso de scrapers en investigación social en Instagram sobre cáncer. *Cuadernos.Info*, *49*, 72–97. https://doi.org/10.7764/cdi.49.27809

Varela-Rodríguez, M., & Vicente-Mariño, M. (2020). Automated image extraction from Instagram for social research: A technical and ethical exploration. *TEEM 2020 Online Conference: Technological Ecosystems for Enhancing Multiculturality, Salamanca 21-23 October 2020*, 5. https://doi.org/10.1145/3434780. 3436650

Vraga, E. K., Stefanidis, A., Lamprianidis, G., Croitoru, A., Crooks, A. T., Delamater, P. L., Pfoser, D., Radzikowski, J. R., & Jacobsen, K. H. (2018). Cancer and Social Media: A Comparison of Traffic about Breast Cancer, Prostate Cancer, and Other Reproductive Cancers on Twitter and Instagram. *Journal of Health Communication*, *23*(2), 181–189. https://doi.org/10.1080/10810730.2017.1421730

World Health Organisation. (2020). *Global Cancer Observatory: Cancer Today*. https://gco.iarc.fr/today/online-analysis-table?v=2018&mode=cancer&mode_population=continents&population=900&populations=900&key=asr&sex=0&cancer=39&type=0&statistic=5&prevalence=0&population_group=0&ages_group%5B%5D=0&ages_group%5B%5D=17&nb_items=5&group_cancer=1&include_nmsc=1&include_nmsc_other=1

Zade, H. A., Habibi, L., Arabtani, T. R., Sarani, E. M., & Farpour, H. R. (2017). Functions of Social Networks in a Community of Cancer Patients: The Case of Instagram. *International Journal of Networks and Communications*, *7*(4), 71–78.

Zaid, T., Burzawa, J., Basen-Engquist, K., Bodurka, D. C., Ramondetta, L. M., Brown, J., & Frumovitz, M. (2014). Use of social media to conduct a cross-sectional epidemiologic and quality of life survey of patients with neuroendocrine carcinoma of the cervix: A feasibility study. *Gynecologic Oncology*, *132*(1), 149–153. https://doi.org/10.1016/j.ygyno.2013.10.015