



A Corpus Linguistic Analysis of the Audience's Views on Breast Cancer Campaigns on Social Media

Eman Adil Jaafar*

Rajaa Essa Whieb

University of Baghdad / College of Education for Women

Article Info

Date of Article

2024/4/04 Received: Received in 2024/4/25 revised form: 2024/5/06

Accepted: Available

2024/7/10 online:

Keywords:

corpus linguistics, health communication, opinion-mining, social media campaigns

Abstract

Media awareness campaigns increase public knowledge of breast cancer care effectively and highlight the existing and potential roles of the media in spreading health awareness. These campaigns have become vital for health communication since they effectively use the internet to spread information and provide intervention opportunities to modify behaviour. This study aims to study the audience views or opinions towards these campaigns that are published on YouTube, which is one of the popular social media platforms. A corpus linguistic methodology and sentiment analysis are applied to examine these interactive views. The data is collected from the audience's comments on five selected videos about breast cancer companies that spread health awareness. In order to empower and persuade individuals, groups, and communities to make healthier decisions, health communication techniques encompass both written and spoken forms of communication. To encourage constructive alterations in attitudes and actions, health communication frequently incorporates elements of several theories and models. Social media health campaigns aim at altering behaviour and saving lives.

© All rights reserved to Muthanna University 2024

1-Introduction

1.1 Health Communication: Definitions and Scope

Health communication is the science and the art of preventing disease, prolonging life, and promoting physical health and efficiency through organized community efforts for the sanitation of the environment (Berridge, 2016; Essa & Jaafar, 2024). It aims to communicate about the control of community infections, the education of the individual in principles of personal hygiene, the organization of medical

and nursing services for the early diagnosis and preventative treatment of disease, and the development of the social machinery that will ensure to every individual in the community a standard of living adequate for maintenance of health. Moreover, health communication is defined as "The art and technique of informing, influencing, and motivating individual, institutional, and public audiences about important health issues. Its scope includes disease prevention, health promotion, health care policy, and business, as well as enhancement of the quality of life and health of individuals within the community" (Ratzan et al., 1994, p. 361).

Health communication covers various communicative activities. It encompasses language used in numerous contexts, involving casual conversations about disease wellness. online assistance groups, advertisements for healthcare services. discussions with doctors and other healthcare providers, and health documents such as policy papers (Brookes & Collins, 2023, p. 3). Simple and straightforward language is used communicate health recommendations people to ensure global spread of information(Essa & Jaafar, 2024). Thus, this study investigates public views about such awareness with regard to health campaigns.

1.2 Research Questions

The study aims to answer the following questions:

- 1-How do interlocutors view the social media health campaigns?
- 2- What are the main themes in the audience interactions and the frequent language patterns about the awareness videos?
- 1.3 Objectives of the Study

The study aims at:

- 1-investigating the communicators' general views about breast cancer campaigns on the selected YouTube videos.
- 2- examining the main themes and recurring language patterns in the audience's interactions related to health campaigns.

Method and Data

The study applies a mixed-method methodology. The integration of both qualitative and quantitative approaches ensures

objective and reliable results. Corpus linguistics techniques make it possible to accurately determine the frequency and patterning of language across large corpora with the aid of technology (Brookes & Collins, 2024). These methods can reveal, for instance, how frequently specific phrases or word chains occur in a corpus and which words, when compared to another corpus, occur with a notably high frequency. Qualitative analysis is performed after gaining results from the quantitative analysis, and it helps interpret statistical results and provide objective linguistic justification(Jaafar, 2014). The method of the study has two phases:

2.1 The Sentiment Phase

The sentiment phase in this study is applied to study the audience's view through carrying out a detailed sentimental analysis or opinionmining process. Liu (2021, p.11) defines it as "Sentiment analysis or opinion mining is the computational study of people's opinions, sentiments, evaluations, attitudes, moods, and emotions. It is one of the most active research areas in natural language processing, data mining, information retrieval, and Web mining". Positive and negative views or opinions that are expressed are the primary focus of sentiment analysis. In this context, the problem is defined. Opinion is a broad concept that encompasses evaluation, appraisal, and attitude, as well as its sentiment. associated information, including the opinion objective and the opinion holder. Sentiment, on the other hand, refers exclusively to the positive or negative sentiment that is implicit in an opinion. Sentiments frequently are manifested in terms of polarity, characterized by two opposing extremes, such

satisfied/unsatisfied, nice/awful, happy/unhappy, positive/negative, and pro/con (Le Lei & Liu, 2021).

This approach helps understand what the audience of health campaigns thinks. The comments that have been written by the consumers of information enable the researchers to grasp their opinions and give an idea about the effectiveness of these health campaigns. Gaining opinions and views from consumers helps to improve how the quality of information in awareness videos so they can reach a wider audience (Pang & Lee, 2008; Lightart et al., 2021). Opinion mining is one technique for obtaining emotions sentiments from a text. This approach uses text analysis. biometrics, natural language processing, and text extraction to methodically find, extract, evaluate, and investigate affective states. Discovering the emotional undertone of a text body is the aim of opinion mining (Lopes & Fonseca, 2023, p. 1). Sentence emotion is identified by a corpus-based technique that uses syntactic and semantic This method patterns. starts predetermined collection of sentiment terms and their position in the sentence, then looks for similar or grammatical patterns in a large corpus to uncover sentiment tokens and their orientation. It assists in solving a matter of opinion terms having orientations that change depending on the context (Wankhade, 2022, p. 5745). Fundamental sentiment analysis is characterized as a bipolar or multiclass categorization task that determines the essential polarity of a text structure (such as a phrase, a sentence, or a token). To incorporate associated characteristics and facets, such as opinion topics, viewpoint holders, and subjects not explicitly covered in the text, a supporting line of study expands upon the essential sentiment classification. These methods typically use sequence classification models to decide if the current token relates to a particular sentiment-related aspect. Therefore, finding appropriate parts or aspects of the topic and matching sentiment is crucial to this type of supportive work (Li & Hovy, 2017, p.45).

2.2 The Corpus Phase

The corpus phase in this study aims to identify frequencies, keywords and concordancs to support the qualitative analysis. This requires building the corpus of the study and processing it in a computer tool. AntConc 4.2.4 (Anthony, 2023) was used to conduct the corpus study and to examine the keywords, frequencies, and concordances. This corpus tool is feasible to use and enables to gain objective statistics effectively and freely (Oakey, 2022; Jaafar, 2022, 2020; Jaafar, 2019).

The data was collected from five YouTube videos published online in different periods. Four videos were published in 2020, 2018, and one in 2011. The main selection criterion is based on the high number of views and comments that reached thousands and millions of viewers who posted their opinions and reflections on the content of the videos. The corpus consists of (28228) tokens. comments are exported through an online website known as YouTube Export Comments https://exportcomments.com/ .Ethical guidelines have been followed by concealing interlocutors' names and details, though these comments are written in a public mood. The five videos provide awareness tips and the best

protection practices that should be taken to deal with breast cancer. For example, one of the videos tackles Breast Self-Examination (It can save your life)(RafflesHospital, 2018). The second one deals with Breast cancer type and stage: what you need to know(Breast Cancer School for Patients, 2018). Similarly, information about one of the common tests of breast cancer Mammogram what you need to know (Whatley, 2011) is mentioned in the third video. The fourth video covers The 12 Breast Cancer Symptoms and Signs - What to look for on your Self-Breast Exam(Cancer Rehab PT, 2020) and presents tips for self-examination. The last video posted on Nucleus Medical Media (2020) that tackles medical treatment about the disease.

3. Analysis and Discussion

In light of the comments provided, the following linguistic patterns and themes are the notable recurring ones in the data under investigation. It is important to note that there are health concerns that the audience share. For example, concerns about breast health are expressed in numerous comments; examples include pain, swelling, nipple (lumps, alterations, discharge, hardness, and sensitivity). Additionally, certain medical conditions such as breast cancer and lymph node enlargement, are addressed. Through other interactions, communicators often seek reassurance or advice from others, especially if they are young and worried about breast abnormalities. The individuals inquire about the characteristics of symptoms, whether it is necessary to consult a physician, and what steps to undertake. Moreover, other themes appear in this study that show gratitude and support. Certain responses express gratitude

towards educational films. medical professionals, herbal remedies, and supportive communities for their ability to reduce anxiety and offer direction. Additional remarks convey appreciation for the assistance provided to However, many commenters communities. express fear, anxiety, and sadness about potential health issues and concerns about pain and discomfort, especially related to breast cancer. Some mention family histories of cancer, adding to their concerns. Examples include

- "I'm so worried about the pain"
- "This is hard, uncomfortable and painful for sure."
- "its so damn hurt.!"
- "I'm nervous as fuck."
- "My boobs are always very tender, so having them squashed in a brutal machine was very unpleasant."
- "This looks horrible"
- "Getting mine done on the 17th.. I'm nervous as fuck.."
- "I had my first one two days ago (I'm 64) and in my case it was extremely painful during and after (and still!)."
- "no sign of cancer thank god i get one every year no cancer or std in my vagina hiv aids test results was negative thank you lord"

Several comments share personal experiences with breast health. medical diagnoses, treatments, and outcomes. These stories often details about consultations include with doctors. diagnostic procedures like mammograms and biopsies, and the emotional impact of health scares. Cultural and personal views are also part of a few comments, including religious or spiritual references, such as prayers, blessings, and mentions of Jesus Christ, indicating the emotional and spiritual coping mechanisms people use during health crises. Apparently, some comments reflect cultural attitudes toward healthcare, family roles in health monitoring, and the importance of self-examination awareness for example:

• "Please check yourselves I did and noticed something even after 1mqmagram!,, felt bigger & hurt Now after going again,2nd mammogram,,, I'm setting up a SURGERY DATE."

This comment highlights the significance of maintaining a self-examination and being aware of changes that occur within one's body. In order to demonstrate the value of selfawareness in identifying potential health problems, the individual took the effort to monitor their own health and swiftly took action when spotting something abnormal. Communicators also appreciate educational content like videos and discussions that raise awareness about breast health. selfexamination techniques, and the importance of seeking medical advice. In contrast, some views show that people do not encourage tests for mammograms; they believe these tests are dangerous and harmful in terms of spreading cancer, for example:

- "Mammograms are very dangerous compresses the breast tissue too much and even if you do have some kind of breast cancer there it would only spread it"
- "You are at risk getting mammogram cause if you have a tumor the the machine squashes your breast and breaks the tumor and its cancer it can spread over your entire system and you may 2 or 3 months to live"
- "Never get a mammogram. You are an idiot".

In general, the recurrent linguistic patterns and themes underscore the collective sentiments, anxieties, and experiences pertaining to breast health, medical consultations, and individual trajectories amidst health obstacles such as breast cancer and other ailments.

3.1 Corpus Analysis of the Data

The Word List tool In AntConc (version 4.2.4) generates the words. Determining how frequently the most common keywords occur in the corpus is achieved by excluding function words such as (articles, pronouns ..etc). Lexical words can give an idea about the corpus under investigation. Table (1) presents twenty keywords arranged according to their frequencies. Since the study is sentimental by its methodology, the corpus study helps to locate essential keywords that the audience in expressing their feelings uses experiences; for example, the word 'feel' occurs frequently in the comments and this is logically associated with specific nouns and adjectives like (pain, painful, scared, worried, hard, hurt)

Table 1. Wordlist of the most frequent keywords

KCy W	RCy WOIUS					
NO.	Rank	Frequency	Words			
1.	7	527	Breast			
2.	12	370	have			
3.	15	287	cancer			
4.	28	167	pain			
5.	31	152	do			
6.	33	143	lump			
7.	37	131	left			
8.	39	127	feel			
9.	45	103	right			
10	47	99	scared			

11.	49	95	know
12.	52	84	doctor
13.	53	82	thank
14.	55	81	get
15.	59	79	nipple
16.	74	62	help
17.	74	62	mammogram
18.	83	58	worried
19.	86	55	stage
20.	101	46	Hard, hurt

Using KWIC (Keyword In Context), or concordance lines, display the search word in the center of the screen with a predetermined number of characters surrounding it rather than an index listing every word's location in a corpus. Concordance lines display a word in its surrounding grammatical and lexical context. The concordance tool is valuable for examining the relationships between different words and their grammatical structures and meanings.

KWIC in Figure (1), shows some context information before and after the word. The word "breast" is a concrete noun. The noun "breast" occurs with 527 hits in the corpus and obtains the highest frequency. In lines (20), (21), and (23), declarative sentences are used to emphasize the value of having a wide range of treatments for breast cancer, such as radiation, hormone therapy, and chemotherapy. For example, "I'm fine. I was diagnosed with a DCIS Stage 1 Breast cancer in June of 2016. I had chemotherapy and radiation." Another example is, "I don't know if that didn't make the chemo worse. I am so thankful that today breast cancer is not automatically a death sentence. Very interesting."

The keyword "lump" is a concrete noun and has 143 hits of the concordance lines. In these lines, "lump" is mainly followed by a prepositional phrase (in my left breast, in my breast). Here, People who communicate frequently ask for knowledge from medical professionals, as shown in Figure 2:

Here, People who communicate frequently ask for knowledge from medical professionals, as shown in Figure 2:

In these lines, "feel" achieves the frequency with 127 hits of the concordance lines. The keyword "feel" is a stative verb used to express emotion or condition. Health campaigns aim to help educate the local community about early breast cancer identification and mammograms. For example, "don't know where. I feel extremely fatigued and I feel like I'm dyin' never felt this bad before." Another example is "cancer n all peoples have lots of questions with u feel like to share about atypical cell n." form mammogram Additionally, "sometimes but sometimes there are no lumps, but sometimes I feel I'm examine but feeling so many lumps," as shown in Figure 3:

The keyword "help" is a dynamic verb and appears with 62 hits in the corpus. Imperative sentences are utilised. In lines (1), (6), and (15), "help" is a request for assistance. For instance," any kinda wound...I don't know what to do...plz help me out. I have pain in my both the nipples," And "I move it specifically near my right breast. Can anyone help me? I didn't have my period yet." Similarly, "from my childhood. Even I am scared of needle also. So please help me to reply me. I am 55 yrs. I just," as shown in Figure (4):

A Corpus Linguistic Analysis of the Audience's Views on Breast Cancer...

The keyword "stage" is a concrete noun with 55 hits of the concordance lines. In line (51)," What you're saying, it sounds like mine is the stage 2 that hasn't attacked any lymph nodes. They haven't." In line (53) "yes, same us my situation about your situation I was stage three had chemo surgery then radiation last was 15 treatments." Declarative sentences are used to

demonstrate that early breast cancer is far more treatable and has a greater probability of survival than later stages, which demonstrates that the cancer has spread to other parts of the body. The health is threatened from this stage, as shown in Figure 5:

Figure 1.

The Keyword in Context Tool of "Breast"

	File	Left Context	Hit	Right Context
13	Audience views	What about the breast cancer that doesnt have a lump?	Breast	Cancer School for Patients I am wearing my breast
14	Audience views	l might have it can a 14 year old get it		Cancer School for Patients4 what kinds of type brea
5	Audience views	who I am to you Thanks you lat me know		Cancer School for Patients How do you determine
6	Audience views	can i do? Is the invasive Brest cancer is harmful?		Cancer School for Patients NoI want to learn
17	Audience views	and having a burning sensation Can 11 year old get cancer?		Cancer School for Patients Thank you I have pain
18	Audience views	guys if you're under 20 :) Is it possible to have		cancer in both of the breast. What if someone
9	Audience views	and I'm so grateful???? Is it possible to have		cancer in both of the breast. What if someone
20	Audience views	I'm fine. I was diagnosed with a DCIS Stage 1		cancer in June of 2016. I had chemotherapy and rac
21	Audience views	checked out medically because it might be something serious since		cancer in men are rare but still risky ans
22	Audience views	good health and please be strong always ,we got this 283		cancer is hereditary btw don't worry if you
23	Audience views	make the chemo worse. I am so thankful that today		cancer is not automatically a death sentence. Very i
24	Audience views	it plz help I'm a very high risk for		cancer is there any other things I should do
25	Audience views	video for my nursing research presentation?? Guys, I know that		cancer is very rare in people under 20 but if
26	Audience views	at the tip of the nipple I think I have	breast	cancer but I'm scared to tell my mom
27	Audience views	might swell. I don't know if this applies to	breast	cancer but lymphoma (another type of cancer) can

A Corpus Linguistic Analysis of the Audience's Views on Breast Cancer...

Figure 2.

The Keyword in Context Tool of "Lump"

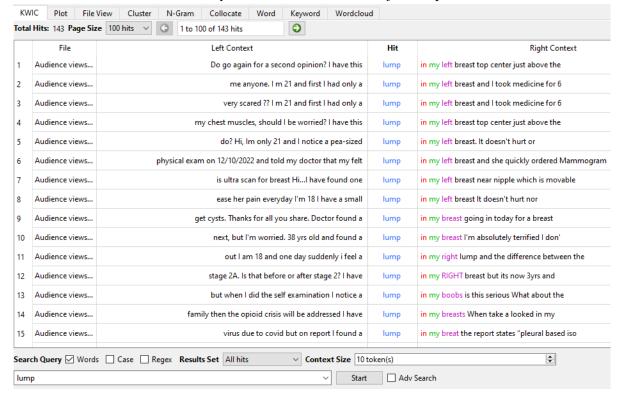


Figure 3.

The Keyword in Context Tool of "Feel"

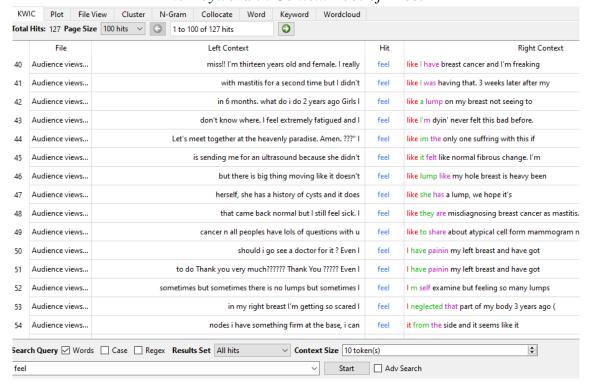
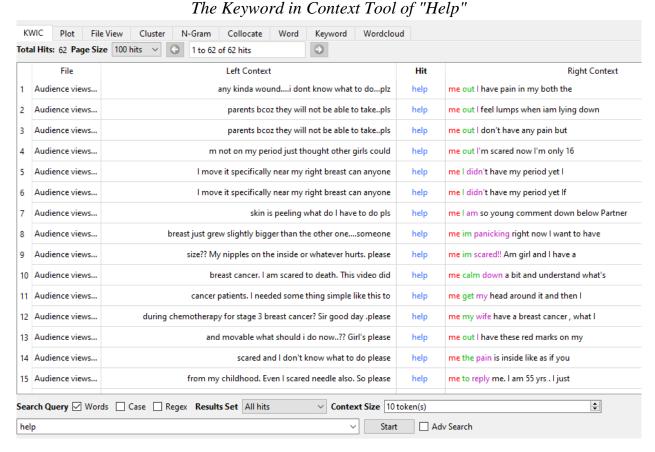


Figure 4.



4. Conclusion

The study aimed at revealing sentimental language patterns to know the views and the common themes of the audience on selected YouTube awareness videos on women's breast cancer. After building the corpus, it has been subjected to two phases of analysis: the sentimental and corpus phase. In general, the sentiment analysis uncovers a diverse variety of opinions and emotions that are exchanged in the comments, including thanks and gratitude to the poster of these videos. Moreover, optimism, support, fear, and anxiety are also expressed. This confirms the significance of incorporating education, awareness, compassion into dialogues concerning breast health and cancer consciousness. In addition, commentators communicate sentiments optimism, support, and hope. Promotion of determination, active pursuit of medical care, and mutual support among participants foster a constructive environment in the communication. To sum up, the study gained useful positive insights about the audience's views and should encourage awareness campaigns to provide more educational videos about breast cancer and general health care.

References

Anthony, L. (2023). AntConc (Version 4.2.4) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software Breast Cancer School for Patients. (2018, March 26). Breast cancer type and stage: what you need to know [Video]. YouTube. https://www.youtube.com/watch?v=ZWqfoBj2bsA

Brookes, G. and Collins, L. (2023) *Corpus Linguistics for Health Communication: A Guide for Research*. Routledge.

Brookes, G. and Collins, L. (2024) *Corpus Linguistics for Health Communication: A Guide for Research*. Routledge.

Brookes, G., Harvey, K. and Adolphs, S. (2023). Discourse and health(care). In M. Handford & J.P. Gee (Eds.), *The Routledge Handbook of Discourse Analysis* (second edition, pp. 553–567). Routledge. https://doi.org/10.4324/9781003035244-46

Berridge, V. (2016). 1. What is public health? In *Oxford University Press eBooks* (pp. 1–10). https://doi.org/10.1093/actrade/978019968846 3.003.0001

Cambria, E., Das, D., Bandyopadhyay, S., & Feraco, A. (Eds.). (2017). *A practical guide to sentiment analysis* (Vol. 5). Springer International Publishing.

Cancer Rehab PT. (2020, September 29). The 12 Breast Cancer Symptoms and Signs - What to look for on your Self-Breast Exam [Video]. YouTube.

https://www.youtube.com/watch?v=8T90GKk oqZ4

Demjén, Z., Atkins, S., & Semino, E. (2023). *Researching language and health: A Student Guide*. Routledge.

Download YouTube Comments. (n.d.). YouTube Comments Downloader. https://youtubecommentsdownloader.com/a

Essa, R. W. and Jaafar, E. A. (2024) Women's health on social media: A Corpus stylistic study of pink October campaigns. *Cogent Arts & Humanities*, 11 (1). 2341498 https://doi.org/10.1080/23311983.2024.234149

Whatley, G. (2011, October 26). *Mammogram what you need to know* [Video]. YouTube.

https://www.youtube.com/watch?v=175IiktUk 6w

Harvey, K. (2013). *Investigating adolescent health communication: A corpus linguistics approach*. Bloomsbury.

Jaafar, E. A.(2022). Book review of Ringrow and Pihlaja's contemporary media stylistics. *International Journal of Arabic-English Studies*, 22(2). 341-343. https://doi.org/10.33806/ijaes2000.22.2.19.

Jaafar, E. A. (2020). Schema theory and textworlds: A cognitive stylistic analysis of selected literary texts. *Journal of the College of Education for Women*, 31(2), 52–65. https://doi.org/10.36231/coedw/vol31no2.14

Jaafar, E. A. (2019). Examining Seamus Heaney"s Poem "A Herbal": *A Corpus stylistic approach. Alustath Journal for Human*

https://doi.org/10.36473/ujhss.v58i1.826

Jaafar, E. A. (2014). A stylistic analysis of Bisson"s bears discover fire. Journal of the College of Education for Women, 25(4), 1173-1182.

and Social Sciences, 58(1), 17–26.

https://doi.org/10.36231/coedw.v25i4.888 Jaafar, E. A. (2014). A stylistic analysis of two selected poems. *Journal of the College of Education for Women*, 25(1).

Le Lei, L., & Liu, D. (2021). *Conducting Sentiment Analysis*. Cambridge University Press.

Li, J., & Hovy, E. (2017). Reflections on sentiment/opinion analysis. In E., Cambria, D., Das, S., Bandyopadhyay, & A., Feraco (Eds.). *A practical guide to sentiment analysis* (pp. 41-59). Springer International Publishing Lighart, A., Catal, C., Tekinerdogan, B., 2021. Systematic reviews in sentiment analysis:

a tertiary study. *Artif. Intell. Rev.* 54 (7), 4997–5053. http://dx.doi.org/10.1007/s10462-021-09973-3

Liu, B. (2017). Many facets of sentiment analysis. In In E., Cambria, D., Das, S., Bandyopadhyay, & A., Feraco (Eds.). *A practical guide to sentiment analysis*, (pp.11-39). Springer International Publishing.

Lopes, A. G., & Fonseca, L. (2023). Tools and Techniques for Opinion Mining. In IS-EUD Workshops. https://ceur-ws.org/Vol-3408/short-s2-02.pdf

Nucleus Medical Media. (2020, October 9). Breast cancer treatment [Video]. YouTube. https://www.youtube.com/watch?v=zo0MNjG3jPY

Oakey, D. (2022). What can a corpus tell us about lexis? In O'Keeffe, A., & McCarthy, M. (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 185-203). Routledge.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135. https://doi.org/10.1561/1500000011

Raffles Hospital. (2018, November 16). *Breast Self-Examination (It can save your life)* [Video]. YouTube. https://www.youtube.com/watch?v=LrfE6JUw Ims

Ratzan, S. C., Stearns, N. S., Payne, J. G., Amato, P. P., Liebergott, J. W., & Madoff, M. A. (1994). Education for the health communication professional: A collaborative curricular partnership. *American Behavioral Scientist*, 38(2), 361-380

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, 55(7), 5731-5780.

https://link.springer.com/content/pdf/10.1007/s 10462-022-10144-1.pdf

تحليل لغوي مدون لآراء الجمهور حول حملات سرطان الثدي على وسائل التواصل الاجتماعي

ايمان عادل جعفر

رجاء عيسى وهيب

جامعة بغداد / كلية التربية للبنات

الخلاصة:

تزيد حملات التوعية الإعلامية الفيديوبة على منصة اليوتيوب من المعرفة العامة والتوعية للكشف المبكر عن سرطان الثدى بفعالية وتسليط الضوء على الأدوار الحالية لوسائل الإعلام في نشر الوعى الصحى لمحاربة هذا المرض. أصبحت حملات التوعية الإعلامية حيوبة للتواصل الصحى لأنها تستخدم الإنترنت بفعالية لنشر المعلومات وتوفير فرص التدخل لتعديل السلوك الفردي والجماعي. تهدف هذه الدراسة إلى التحري عن وجهات نظر الجمهور أو الآراء تجاه هذه الحملات التوعوبة التي يتم نشرها على منصات التواصل الاجتماعي. يتم تطبيق المنهجية اللغوية المدونة وتحليل الاراء لفحص هذه وجهات النظر التفاعلية.حيث يتم جمع البيانات من تعليقات الجمهور من خمسة مقاطع فيديوبة مختارة حول حملات التوعية من سرطان الثدى التي تنشر الوعى الصحى. من أجل تمكين وإقناع الأفراد والمجموعات لاتخاذ قرارات صحية ، تشمل تقنيات الاتصالات الصحية أشكال الاتصال المكتوبة والمنطوقة. لغرض تشجيع التعديلات البناءة في المواقف والإجراءات ، يتضمن التواصل الصحى عناصر من العديد من النظربات والنماذج. تهدف حملات صحة وسائل التواصل الاجتماعي إلى تغيير السلوك وإنقاذ الأرواح.

الكلمات المفتاحية: اللغويات والآراء , حملات التوعية الصحية . لغة وسائل التواصل الاجتماعي, علم اللغة المدون.