JOMDI

Journal of Medical and Dental Investigations

J Med Dent Invest 2025; 6: e250109
https://doi.org/10.5577/jomdi.e250109

Original Article Open Access

Analysis of YouTube videos on assisted reproductive techniques

Selim Demirtaş[1], **Mehmet Nuri Gördük**[2]

[1] Mardin Artuklu University, Faculty of Medicine, Department of Histology and Embryology, Mardin, Türkiye

[2] Mardin Artuklu University, Faculty of Medicine, Department of Public Health, Mardin, Türkiye

Received: 8 November 2025
Revised: 26 November 2025
Accepted: 5 December 2025
Published: 12 December 2025

Correspondence:

Dr. Selim DEMİRTAŞ
Mardin Artuklu University, Faculty of Medicine, Department of Histology and Embryology, Mardin, Türkiye.
E-mail: selimdemirtas@artuklu.edu.tr

How to cite this article:

Demirtaş S, Gördük MN. Analysis of YouTube videos on assisted reproductive techniques. J Med Dent Invest 2025;6:e250109. https://doi.org/10.5577/jomdi.e250109

Abstract

Aim: Assisted reproductive techniques (ART) have made significant progress in infertility treatment and have found a wide range of applications with various methods, especially in vitro fertilization. Today, digital platforms such as YouTube have become an important resource for individuals seeking information about ART. However, uncertainties about the reliability and quality of this content pose a potential risk for viewers. This study aims to assess the quality and reliability of ART-related videos on YouTube.

Methods: 76 videos were analyzed by applying elimination criteria from 350 videos uploaded between October 1, 2014, and October 1, 2024. These videos were obtained through the YouTube Data API v3 by searching seven different terms related to in vitro fertilization and infertility. In addition to descriptive statistics on video features, the Global Quality Scale (GQS), modified DISCERN, JAMA and UTvAC scales were used to evaluate the quality of the videos. Additionally, sentiment analysis was conducted based on the comments of the videos.

Results: It was determined that the majority of the analyzed videos were at a medium quality level. While the JAMA scores indicate that most of the videos are of low quality, the GQS and modified DISCERN scores suggest that most videos fall within the medium-quality range. According to UTvAC, most videos are just below the high-quality limit. The majority of videos were uploaded by doctor-independent influencers (32.9%), non-governmental organizations (25.0%), and news/media (21.1%) channels. The positive comment percentage was found to be 28%, indicating that the videos were generally poorly appreciated by viewers.

Conclusion: This study revealed that the information quality and reliability of ART videos on YouTube are generally moderate. The findings of the study show that there is a significant need for improvements in digital health communication and educational content production.

Keywords: Assisted reproductive techniques, in vitro fertilization, patient education, YouTube videos.

Introduction

Assisted reproductive techniques (ART) have made significant progress in the field of infertility treatment. After the first successful in vitro fertilization (IVF) procedure and the birth of the first "test tube baby" named Louise Brown in 1978, ART methods developed rapidly and found a wide range of applications (1, 2). Apart from IVF, techniques such as intrauterine insemination (IUI), intracytoplasmic sperm injection (ICSI), and genetic screening are also widely used in infertility treatment (3). Technological innovations in this field have made treatment processes safer, more effective, and more accessible. While ART applications make individuals' and couples' hopes of having children come true, increasing the level of knowledge and understanding of treatment processes has also gained importance.

Founded in 2005, YouTube has grown rapidly as the largest video-sharing platform on the Internet and has reached billions of users today (4). As of 2024, YouTube has approximately 2.7 billion active users worldwide. This means that approximately 52% of internet users worldwide access YouTube at least once every month, and 56.7% of social media users worldwide use YouTube (5). YouTube has become an active source of information not only for entertainment and social content, but also in important areas such as education and health (6). In particular, content in the field of health facilitates users' access to medical information (6, 7).

There is no exact percentage of YouTube users worldwide who watch health-related content. However, some studies and data show that health topics are followed by a significant audience on YouTube (8).

There are many videos about ART on YouTube, and we could not find any direct study on this subject in the literature. Existing studies have examined specific topics (9, 10). In this study, we aimed to evaluate how useful and reliable ART videos on YouTube for patients through four score analyses.

Materials and Methods

Data collection

We conducted a study of YouTube videos using the YouTube Data API v3 on October 1, 2024. The search was performed using the following keywords: 'IVF fertilization', 'infertility treatment', 'male infertility', 'female infertility', 'IVF baby', 'assisted reproductive technology', and 'IVF procedure.'

The YouTube API provided direct access to video metadata, including video ID, title, author, channel description, channel keywords, channel subscribers, publication date, duration, views, total comment count, and likes. The channels were categorized into six groups based on the authors of the videos.

Inclusion and Exclusion Criteria

1. **Video Selection:** For each keyword, the first 50 videos were selected, resulting in an initial pool of 350 videos. Duplicate videos were subsequently removed.
2. **Video Duration:** Videos shorter than 60 seconds were excluded to eliminate brief, non-informative clips.
3. **Recency:** Videos published more than 10 years before the search date were excluded to ensure the data reflect recent discussions and up-to-date medical information.
4. **Comment Analysis:** Videos with fewer than 10 comments were excluded, as the research focuses on sentiment analysis of comments.
5. **Language:** Only videos in English were included to ensure consistency and comprehensibility during the analysis.
6. **Final Dataset:** After applying all exclusion criteria, a total of 76 videos were included in the study (Fig. 1).

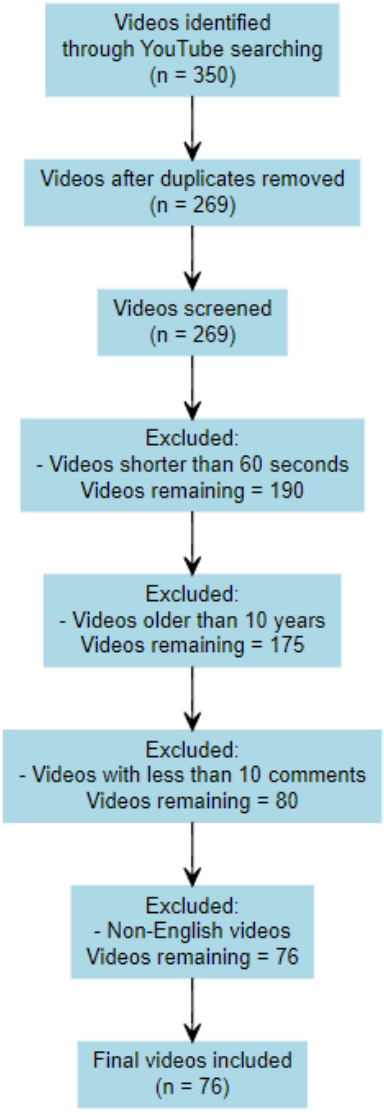

Figure 1. Video counts after inclusion and exclusion criteria.

Comment Retrieval, Preprocessing, and Sentiment Analysis

Using the YouTube Data API, we retrieved the top-level comments (i.e., main comments without replies) from each of the selected videos.

Comments were pre-processed to remove irrelevant elements, such as emojis, URLs, special characters, and non-alphanumeric symbols. Additionally, single-character comments (for example, "!", ".", etc.) were excluded from the analysis to enhance data relevance and clarity.

Sentiment analysis was performed on the cleaned comments to evaluate the tone of the viewer discussions. We used the Twitter-roberta-base-sentiment model, a transformer-based approach known for accurately capturing nuanced sentiments, including sarcasm and context-specific expressions. Comments were classified as positive, neutral, or negative, and the overall sentiment distribution and mean sentiment scores were calculated for each video.

Scales

Four different scales were used to evaluate the quality of the videos: Global Quality Score (GQS), modified DISCERN, Journal of American Medical Association (JAMA) benchmark criteria, and YouTube Video Assessment Criteria (UTvAC) (11-14). Each video was independently watched and assessed by the authors, and the average score of the authors was considered in the analysis.

A graphical illustration of the study is shown in Figure 2.

Statistical Analysis

Statistical analyses were performed using parametric and non-parametric tests, depending on the distribution characteristics of the data. The normality assumption of the variables was evaluated using the Shapiro-Wilk test, and since the normality assumption was not met, the Kruskal-Wallis test was applied for comparisons between groups. For variables with significant differences detected by the Kruskal-Wallis test, Dunn's post-hoc test with Bonferroni correction was applied to determine which group caused the difference between the groups. Central tendency measures of continuous variables are reported as mean ± SD and median (min-max). The relationships between video length and the number of views, total number of comments, number of positive comments, and scale scores were evaluated using Spearman's correlation analysis. The Kruskal-Wallis test was used to assess the differences in variables across the channel groups. The statistical significance level was set at $p<0.05$.

As this study utilized publicly available data, ethical approval was not required.

All data analyses were performed using the R programming language, along with their relevant libraries.

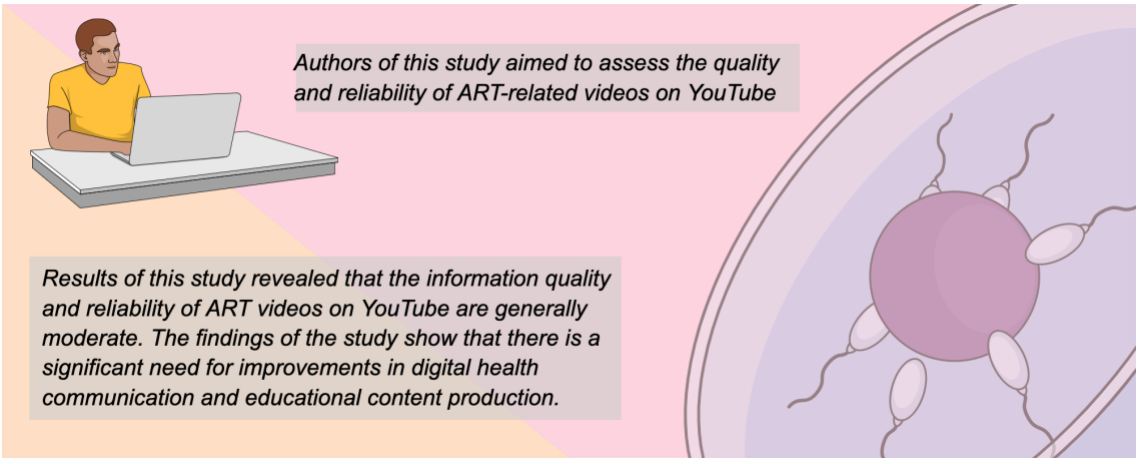

Figure 2. Graphical illustration of the study

Results

A word cloud was created to visualize the most frequently used keywords in the video channel descriptions. As shown in Figure 3, the most prominent keywords included IVF, medical, news, fertility, health, and infertility, followed by terms such as doctor, treatment, and biology.

In this study, 76 YouTube videos were analyzed, and the characteristics outlined in Table 1 were examined. After reviewing the videos, 31.6% were found to be direct advertisements or included advertisements alongside informational content.

Figure 3. Word cloud of keywords of the channels.

The distribution of video lengths exhibited a right-skewed pattern, with most videos concentrated in the 0-15 min range (Fig. 4).

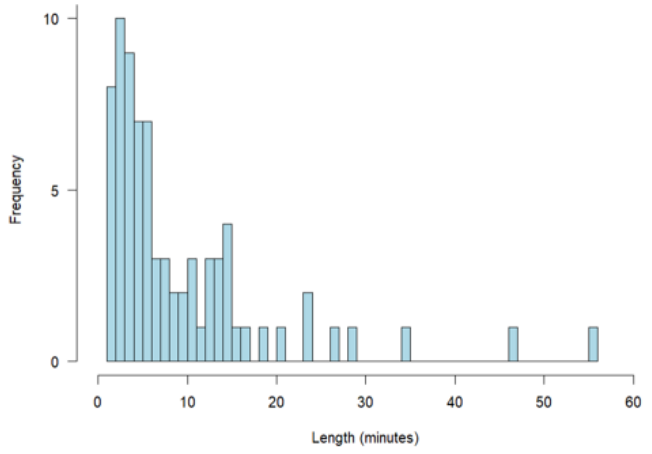

Figure 4. Distribution of the length of the videos.

The quality scores (GQS, modified DISCERN, JAMA, and UTvAC) indicated that most videos tended to fall within the moderate quality range across all scales. The GQS and modified DISCERN scores were distributed similarly, with most videos scoring between 3 and 4 points. However, the JAMA scores were notably low, with the majority of videos clustered at 2 points. This suggests that the videos generally meet the minimal quality standards according to the JAMA criteria. For the UTvAC, the scores were concentrated around 12 points, with fewer videos scoring at the lower or upper extremes. Given that the UTvAC considers 13 points and above as indicative of high-quality videos, this suggests that most videos fall just below the threshold for high quality, reflecting a moderate evaluation of video quality based on this scale (Fig. 4).

The channels were grouped according to their descriptions and keywords, and six different categories were created. Among these categories, unaffiliated doctors (grouped as “doctor independent influencers”) had the highest representation, followed by non-governmental organizations and news/media channels. In the analysis conducted according to channel groups, educational channels had the highest mean views, views per day, likes, and comments per video. However, the "doctor independent influencer" group received relatively more likes, although it received fewer views than the other groups.

The Kruskal-Wallis test was used to analyze the relationship between the variables in Table 2 and the channel groups, and only the comments were found to be statistically significant.

Approximately 9,000 main comments from the videos were subjected to sentiment analysis. According to the analysis, 28% of the comments were classified as positive comments. Comments with positive sentiments received higher scores, and it was understood that the comments grouped as positive by the model were highly positive according to the model's evaluation (Table 3).

The videos were categorized into four groups based on their length: 1-5 min, 5.01-10 min, 10.01-15 min, and greater than 15 min. For the Global Quality Score (GQS), no statistically significant difference was observed between the groups. However, for the JAMA, UTvAC, and modified DISCERN scales, the differences between the groups were statistically significant. For the UTvAC scale, videos in the 1-5 min group had statistically significantly lower scores compared to those in the 10-15 min group and the >15 min group. Similarly, for the JAMA scale, videos in the 1-5 min group had significantly lower scores than those in the 10-15 min group. For the modified DISCERN scale, videos in the 5-10 min group had statistically significantly lower scores than those in the >15 min group.

The correlation between video length (in seconds) and view counts was analyzed, resulting in a Spearman correlation coefficient of -0.0012. This value is extremely close to zero, indicating a negligible and non-meaningful relationship. The relationship between the video length groups and view counts was further examined using the Kruskal-Wallis test, which resulted in a p-value of 0.1171, indicating no statistically significant difference in view counts across the video length groups. Video length showed a weak to moderate positive association with both total comments (Spearman’s correlation coefficient: 0.2302) and positive comment percentages(positive comments/total comments) (Spearman’s correlation coefficient: 0.3089), although the relationships were not strong. This suggests that video length may slightly influence viewer engagement.

When comparing the video length groups with comments (p-value = 0.315) and sentiments, including both positive (p-value = 0.3791) and negative (p-value = 0.5729), using the Kruskal-Wallis test, none of the results were statistically significant.

When comparing positive comment percentages with scale scores, the correlations were found to be very weak(Spearman correlation coefficient: GQS: 0.0512, modified DISCERN: 0.0954, JAMA: 0.0117, UTvAC: 0.1319). Although the UTvAC scores showed a relatively higher correlation with positive sentiment percentages, the relationship remained weak overall.

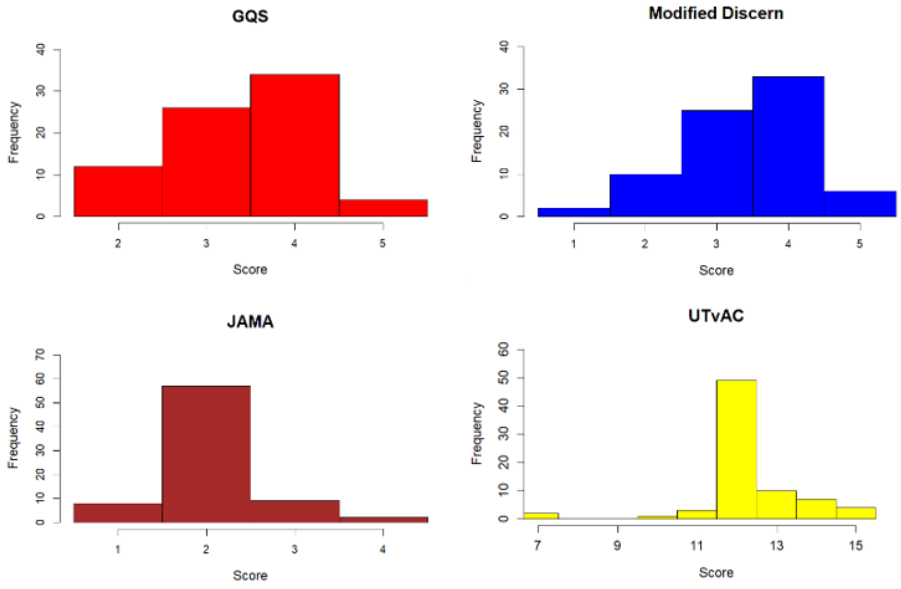

Figure 5. Distribution of evaluation scale score results.

Table 1. Descriptive statistics of video metrics and quality scores.

Variable	Mean	SD	Median	Min	Max
Subscribers	1636134	3557661	284000	934	20700000
Views	252454	582062	59392	2751	3155754
Likes	2389	6392	505	0	42319
Comments	228	762	49.5	10	6394
Days on Youtube	1535	887	1465	142	3455
Length(seconds)	554	594	324	63	3355
GQS	3.39	0.818	3.5	2	5
Modified DISCERN	3.41	0.912	4	1	5
JAMA	2.07	0.574	2	1	4
UTvAC	12.3	1.29	12	7	15

Table 2. Descriptive summary of variables by the channel group.

	Mean ± SD						
variable	Doctor independent influencer	Non-government organization	News/media	University/ hospital/ medical	Independent influencer	Education	p-value*
n (%)	25(32.9)	19(25.0)	16(21.1)	10(13.2)	4(5.3)	2(2.6)	
Likes	1644.5 ± 3668.9	1069.5 ± 2122.0	3113.4 ± 7880.9	1964.0 ± 2478.7	828.8 ± 1263.3	23682.0 ± 26356.7	0.054
Views	165967.2 ± 375118.4	267322.3 ± 663609.3	185843.6 ± 445910.0	304539.9 ± 507272.4	138205.2 ± 109789.7	1693231.5 ± 2068319.2	0.065
Views per day	141 ± 320	153 ± 296	211 ± 469	306 ± 747	83.2 ± 23.1	507 ± 574	0.266
Comments	99.6 ± 125.5	62.1 ± 76.1	564.8 ± 1580.8	230.0 ± 276.6	100.5 ± 142.6	959.0 ± 995.6	0.040
Positive Comments	18.8 ± 23.9	10.4 ± 13.6	53.1 ± 114.2	45.6 ± 69.6	30.0 ± 50.1	231.0 ± 164.0	0.844
GQS	3.5 ± 0.8	3.3 ± 0.7	3.5 ± 0.8	3.5 ± 1.0	2.2 ± 0.5	3.5 ± 0.7	0.151
Modified DISCERN	3.6 ± 0.9	3.2 ± 1.0	3.6 ± 0.7	3.4 ± 1.0	2.2 ± 0.5	3.5 ± 0.7	0.096
JAMA	2.2 ± 0.7	1.9 ± 0.3	2.1 ± 0.4	2.2 ± 0.6	1.5 ± 0.6	2.0 ± 0.0	0.120
UTvAC	12.3 ± 1.5	12.1 ± 0.5	12.2 ± 1.0	12.4 ± 2.1	12.0 ± 0.8	14.0 ± 1.4	0.090

Table 3. Distribution of the sentiment analysis results.

Type	n (%)	Sentiment mean (SD)
Positive	2556(28.0)	0.81(0.2)
Negative	3295(36.1)	0.70(0.2)
Neutral	3287(36.0)	0.68(0.1)

Discussion

This is the first study that analyses ART-related videos on YouTube in terms of different scales and the relationship between these scales and different characteristics of the videos.

It is an expected finding that videos are generally short. Video producers generally create short videos. It was observed that some of the videos analyzed were directly clinic or doctor advertisements or included advertisements in their content. It is not known whether this situation affects the algorithm and whether it brings videos containing advertisements to the forefront when people search for content.

In our study, the video quality was evaluated at a moderate level according to the UTvAC. The modified DISCERN scores ranged between 3 and 4 points, which can be considered moderate-good. The GQS scores were between 3 and 4, which can be considered as moderate-good useful for patients according to this scoring. Finally, the JAMA scores were mostly 2 points. This indicates that the videos were of a low quality.

When the data is examined, it is seen that more than 70% of them are individuals or organizations that can be the authority of this issue. Notably, most of the channel groups are doctor-independent influencers, and the least of them are in the education category. This may be related to the opportunities for content producers, the support they receive, or their professions. Despite the potential for educational impact, there is a shortage of content specifically categorized under the education category. This gap suggests an opportunity for more structured and authoritative educational content on health topics (8).

It is noteworthy that Education channels have the highest rates in terms of the number of likes, views and comments, although they are few in number among the channel groups. Here, the inclusion of students along with the patients researching the videos probably increases the number of views. There are studies showing that educational videos that are engaging and informative tend to receive more engagement (15, 16). The reason why the like/view rate of doctor independent influencers among the expert individuals and institutions is higher than the others may be the effect of doctors' own patients. It is noteworthy that there is no correlation between the channel groups and all scales used, and the results of the scales are similar across channels (17).

In this study, we performed a sentiment analysis on user comments, which has not been encountered in the literature. When comment analysis is performed, the fact that the rate of positive comments is low reveals that the viewers generally do not like the videos very much and indicate this in the comments.

Similar studies in the literature, in our study too, in general, short video length groups scored lower on the JAMA, modified DISCERN, and UTvAC (18, 19). This may be attributed to the insufficient information content of short videos. When the relationship between video lengths and video groups and the number of views was analyzed, the fact that there was no statistically significant result can be interpreted as the fact that video lengths do not affect the number of views of the video. When have been looked at the studies in the literature, there are studies showing that viewers do not prefer longer videos even if they are of higher quality (20, 21).

As a finding not found in other studies, when the percentage of positive comments was compared with the scores obtained from the scales, the very low correlation between them revealed that the positive comments of the viewers were not related to video quality. It was observed that the UTvAC was the most blind to the viewers' comments.

Limitations

This study has some limitations. First of all, the study was limited to videos published only in English on YouTube. This may lead us to ignore the possibility that videos in other languages may differ in terms of quality and content. In addition, the selected videos were found using only certain keywords; therefore, related videos uploaded with less common or alternative terms may not have been included.

The majority of the analyzed videos come from individual creators or small-scale organizations. The limited number of professional videos produced by large healthcare institutions or competent authorities may make it difficult to interpret the results in a broader perspective.

Future studies could provide a more comprehensive analysis with a wider range of languages, different platforms, and various assessment tools.

Conclusion

YouTube has emerged as a significant platform for health-related videos, providing a vast array of content ranging from professional medical advice to personal health experiences. This accessibility allows users to easily find information on various health topics, which can be both beneficial and challenging. YouTube can serve as a valuable resource for health education, offering visual and engaging content that can enhance understanding and awareness. However, the platform's open nature means that not all information is reliable or scientifically accurate, posing a risk of misinformation. Therefore, while YouTube can be a powerful tool for health communication, it is crucial for viewers to critically evaluate the credibility of the sources and content they consume.

In conclusion, when evaluating ART videos on YouTube, we can say that the video content is generally informative for patients. However, users should be careful when searching and should search for content from experts or healthcare organizations that refer to reliable sources for up-to-date information. It is important to focus on newer content because of the advances in ART technology.

Disclosures

Peer-review: Externally peer-reviewed.

Author Contributions: Concept – S.D.; Design – S.D.; Supervision – M.N.G.; Materials – S.D.; Data collection &/or processing – S.D.; Analysis and/or interpretation – M.N.G.; Literature search – S.D., M.N.G.; Writing – S.D.; Critical review – S.D., M.N.G.

Conflict of Interest: The authors declare no conflict of interest.

Funding: No support was received from any institution for this study.

References

1. Steptoe PC, Edwards RG. Birth after the reimplantation of a human embryo. Lancet. 1978;312(8085):366. https://doi.org/10.1016/S0140-6736(78)92490-1
2. Johnson MH. A short history of in vitro fertilization (IVF). Int J Dev Biol. 2019;63(3-4-5):83-92. https://doi.org/10.1387/ijdb.180364mj
3. Zegers-Hochschild F, Adamson GD, Dyer S, Racowsky C, de Mouzon J, Sokol R, et al. The international glossary on infertility and fertility care. Hum Reprod. 2017;32(9):1786-1801. https://doi.org/10.1093/humrep/dex234
4. Arthurs J, Drakopoulou S, Gandini A. Researching YouTube. Convergence. 2018;24(1):3-15. https://doi.org/10.1177/1354856517737222
5. Global Media Insight. YouTube Statistics 2024 (Demographics, Users by Country & More). Accessed December 30, 2024.
6. Osman W, Mohamed F, Elhassan M, Shoufan A. Is YouTube a reliable source of health-related information? A systematic review. BMC Med Educ. 2022;22(1):382. https://doi.org/10.1186/s12909-022-03446-z
7. Gabarron E, Fernandez-Luque L, Armayones M, Lau AYS. Identifying measures used for assessing quality of YouTube videos with patient health information: a review of current literature. J Med Internet Res. 2013;2(1):e6. https://doi.org/10.2196/ijmr.2465
8. Madathil KC, Rivera-Rodriguez AJ, Greenstein JS, Gramopadhye AK. Healthcare information on YouTube: a systematic review. Health Informatics J. 2014;21(3):173-194. https://doi.org/10.1177/1460458213512220
9. Kelly-Hedrick M, Grunberg PH, Brochu F, Zelkowitz P. "It's totally okay to be sad, but never lose hope": content analysis of infertility-related videos on YouTube in relation to viewer preferences. J Med Internet Res. 2018;20(5):e10199. https://doi.org/10.2196/10199
10. Ku S, Balasubramanian A, Yu J, Srivatsav A, Gondokusumo J, Tatem AJ, et al. A systematic evaluation of YouTube as an information source for male infertility. Int J Impot Res. 2021;33(6):611-615. https://doi.org/10.1038/s41443-020-0322-9
11. Azer SA, AlKhawajah NM, Alshamlan YA. Critical evaluation of YouTube videos on colostomy and ileostomy: can these videos be used as learning resources? Patient Educ Couns. 2022;105(2):383-389. https://doi.org/10.1016/j.pec.2021.05.023
12. Bernard A, Langille M, Hughes S, Rose C, Leddin D, van Zanten SV. A systematic review of patient inflammatory bowel disease information resources on the World Wide Web. Am J Gastroenterol. 2007;102(9):2070-2077. https://doi.org/10.1111/j.1572-0241.2007.01325.x
13. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health. 1999;53(2):105-111. https://doi.org/10.1136/jech.53.2.105
14. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the Internet: Caveant lector et viewor—let the reader and viewer beware. JAMA. 1997;277(15):1244-1245. https://doi.org/10.1001/jama.1997.03540390074039
15. Bello-Bravo J, Payumo J, Pittendrigh B. Measuring the impact and reach of informal educational videos on YouTube: the case of Scientific Animations Without Borders. Heliyon. 2021;7(12):e08508. https://doi.org/10.1016/j.heliyon.2021.e08508
16. Saurabh S, Gautam S. Modelling and statistical analysis of YouTube's educational videos: a channel owner's perspective. Comput Educ. 2019;128:145-158. https://doi.org/10.1016/j.compedu.2018.09.003
17. Horani K, Coskey A, Hagedorn JC. Evaluating the quality of ankle fracture education on short-form video platform YouTube Shorts. Foot Ankle Orthop. 2023;8(4): 24730 11423S00304. https://doi.org/10.1177/24730114 23S00304
18. Aydin MA, Akyol H. Quality of information available on YouTube videos pertaining to thyroid cancer. J Cancer Educ. 2020;35(3):599-605. https://doi.org/10.1007/s13187-019-01502-9
19. Duran MB, Kizilkan Y. Quality analysis of testicular cancer videos on YouTube. Andrologia. 2021;53(8):e14118. https://doi.org/10.1111/and.14118
20. Gokcen HB, Gumussuyu G. A quality analysis of disc herniation videos on YouTube. World Neurosurg. 2019;124:e799-e804. https://doi.org/10.1016/j.wneu.2019.01.146
21. Askin A, Sengul I, Tosun A. YouTube as a source of information for transcranial magnetic stimulation in stroke: a quality, reliability and accuracy analysis. J Stroke Cerebrovasc Dis. 2020;29(12):105309. https://doi.org/10.1016/j.jstrokecerebrovasdis.2020.105309