Original research

DIGITAL HEALTH

DIGITAL HEALTH
Volume 9: 1–11

DOI: 10.1177/20552076231171239
journals.sagepub.com/home/dhj
SAGE

A cross-sectional study on reliability and quality of YouTube® videos related to hallux valgus and evaluation of newly developed hallux valgus-specific survey tool

Suat Batar, Mehmet Salih Söylemez, Bahattin Kemah and Serdar Kamil Cepni

Abstract

Objective: Previous studies have reported low quality and reliability on YouTube videos about various medical issues including videos related to hallux valgus (HV) treatment. Therefore, we aimed to evaluate the reliability and quality of YouTube videos on HV and develop a new HV-specific survey tool that physicians, surgeons, and the medical industry can use to create high-quality videos.

Methods: Videos viewed over 10,000 times were included in the study. We used the Journal of the American Medical Association (JAMA) benchmark criteria, global quality score (GQS), DISCERN tool, and new HV-specific survey criteria (HVSSC) developed by us to evaluate the quality, educational utility and reliability of the videos, the popularity of which was assessed using the Video Power Index (VPI) and view ratio (VR).

Results: Fifty-two videos were included in this study. Fifteen videos (28.8%) were posted by medical companies producing surgical implants and orthopedic products, 20 (38.5%) by nonsurgical physicians, and 16 (30.8%) by surgeons. The HVSSC indicated that the quality, educational value, and reliability of only 5 (9.6%) videos were adequate. Videos posted by physicians and surgeons tended to be more popular ($p = 0.047$ and 0.043). Although no correlation was detected among the DISCERN, JAMA, and GQS scores, or between the VR and VPI, we found correlations of the HVSSC score with the number of views and the VR ($p = 0.374$ and $p = 0.006$, respectively). A good correlation was detected among the DISCERN, GQS, and HVSSC classifications (rho = 0.770, 0.853, and 0.831, respectively, $p = 0.001$).

Conclusions: The reliability of HV-related videos on YouTube is low for professionals and patients. The HVSSC can be used to evaluate the quality, educational value, and reliability of videos.

Keywords

Hallux valgus, internet, patient education, quality, surgery, YouTube

Submission date: 3 November 2022; Acceptance date: 5 April 2023

Introduction

Hallux valgus (HV) is a common foot problem with a wide variety of complex and poorly understood causes. Occupation, trauma, shoe wear, genetic predisposition, pes planus and first ray hypermobility have been reported to cause HV. Biz et al. have showed the first ray hypermobility, and HV to be related to anatomical-hereditary factors

Department of Orthopaedics and Traumatology, Umraniye Training and Research Hospital, Istanbul, Turkey

Corresponding author:
Mehmet Salih Söylemez, Department of Orthopaedics and Traumatology, Umraniye Training and Research Hospital, Yaprak Street, Acıbadem District, No. 32, D: 12, 34660 Uskudar, Istanbul, Turkey.
Email: slhsylmz@gmail.com

and to incorrect technical execution in ballet practice.[1,2] HV is more common in women than men and may exhibit a progressive course; there is currently no treatment to slow or stop progression.[3] Many surgical and conservative treatment methods have been proposed; as there are many extrinsic and intrinsic etiologies, treatment must be individualized which can pose a challenge even to trained physicians.[4] Conservative treatment involves patient education, shoe modifications, toe pads, positioning devices, and activity modifications. The position of the first toe can be altered in a lasting effect only before closure of physis. After skeletal maturity, the position of the great toe cannot be altered and conservative treatments aim to alleviate the symptoms.[5] Surgical treatment is indicated for patients that conservative options have failed. Patients having pain that is not alleviated by conservative means and having limitations in daily activities can be treated by surgical procedures. The ideal age for surgery after skeletal maturity is debatable. However, it has been shown that patients operated for HV after 60 years of age have a higher risk for recurrence.[6] Over 120 surgical techniques have been described.[7] Although the best option remains unclear, surgery should be proportional to the extent of deformity. Isolated, open, or minimally invasive distal osteotomy is preferred for mild deformities, distal osteotomy with additional soft tissue interventions for moderate deformities, and multiple osteotomies with further soft tissue interventions and arthrodesis for advanced deformities.[8–10]

HV also causes esthetic issues for many patients.[11] Such patients can be demanding and it may not be easy to meet their expectations.[12] YouTube is the most popular video-hosting website, and is commonly searched for information on health issues. However, YouTube's main source of income is advertising and these advertisements are delivered to the public through original content producers channels who produce and share videos on their own YouTube channels. The revenue from advertisements is distributed as financial support in line with (the main parameters publicly are not yet known) the number of subscriptions to the channel. Additionally, the videos are not peer-reviewed and can be posted by physicians, patients, manufacturers and individuals aiming to increase the number of views and subscriptions rather than providing reliable and qualified information on any issue. Video reliability and quality are very important; use of an inappropriate treatment modality may worsen the condition or waste the patient's time. Very few studies have concluded that videos posted by professional physicians on particular aspects may be reliable. Morra et al. have found the YouTube videos to be reliable on bladder pain syndrome.[13] However, most of the studies on the reliability and quality of YouTube videos on various medical issues found that they were unreliable and failed to educate and inform patients appropriately.[14–17] Thus, physicians should use evaluation tools to guide the creation of videos on health-related topics.[18]

HV is a very common health/cosmetic problem;[11,12] but only a few studies have evaluated the reliability and quality of YouTube videos on this condition.[19–21] We hypothesized that YouTube originated videos related to HV treatment may not be sufficient to guide patients as reported previously. Therefore, we evaluated the reliability and quality of the HV-related YouTube videos to assess the reliability of our newly developed HV-specific survey tool (HVSSC).

Methods

Search strategy

We performed a cross-sectional study searching YouTube using the terms "Hallux valgus surgery" and "Hallux valgus treatment" and ordered the videos from the most to the least watched. Only videos viewed over 10,000 times were included assuming that videos watched more than 10,000 will appear on the first pages and are more visible by watchers.[18,22] Only videos with English audio and sub-titles were included. Videos posted by physicians, patients, and manufacturers; and commercials were included. However, videos dealing with foot conditions other than HV were excluded (Figure 1: Flow chart).

Study screening, data extraction, and evaluation

Two reviewers (both foot and ankle surgeon) independently screened all article titles, and then met to further evaluate the titles and exclude duplicates. Selected videos were viewed in full. A standardized form was used to extract data for assessment of quality and reliability. We recorded the Uniform Resource Locator (URL), title, time since uploading, run time, numbers of views, likes, dislikes,

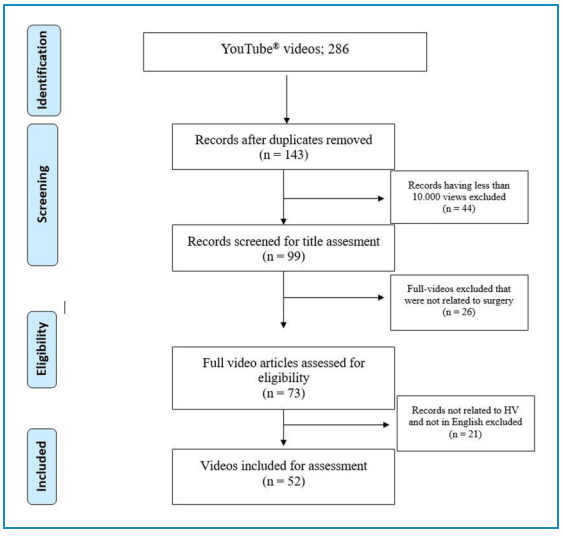

Figure 1. Flow chart diagram showing video selection process (HV: hallux valgus).

and comments, additional sources, content (surgical technique/information on HV/surgical animation/patient experience/advertisement), and type of surgery. Data on the numbers of views and likes were extracted same day by two authors between 22:00 and 24:00 to obtain an exact cross-sectional data as these numbers can change very rapidly. All videos were evaluated in terms of reliability and quality. The Video Power Index (VPI), view ratio (VR) and like ratio (LR) were used to evaluate popularity. These three tools, which use the time since uploading, number of views, likes and dislikes in the calculation, formulate the cross-sectional data obtained and aims to reduce the effect of elapsed time on the reliability of the data, even if the number of views and other parameters change. The VPI was employed to measure views and likes. The VPI was calculated as follows: likes × views/100. The like ratio (LR) was calculated as 100 × likes/[likes + dislikes].[23] The VR was calculated as the number of views/time since uploading (days).[14]

The Journal of the American Medical Association (*JAMA*) benchmark criteria, used to evaluate the accuracy, utility, and reliability of each video, includes four criteria. Satisfied and unsatisfied criteria receive 1 and 0 points, respectively. The scores were used to classify the videos as containing insufficient data (score of 1), partially sufficient data (score of 2–3) or completely sufficient data (score of 4).[24] The global quality score (GQS) was used to evaluate educational value. The GQS is a five-point Likert scale previously employed to evaluate the flow and ease of use of video websites. The GQS classifies videos as very poor, poor, suboptimal, good, or excellent, in terms of patient utility.[25] The DISCERN tool is widely used to evaluate video reliability and information quality.[26] DISCERN includes 16 questions; the first 8 evaluate reliability; questions 9 to 15 explore information quality; and the final question evaluates overall quality. Videos are classified as very poor (16–26 points), poor (27–38), fair (39–50), good (51–62), or excellent (≥63 points).[27] As there was no specific survey tool to assess the quality, educational value, or reliability of YouTube videos on HV, we developed HV-specific survey criteria (HVSSC) adopted from previous orthopedic studies evaluating YouTube videos on lower-extremity conditions (Table 1).[22,28–31] The previous literature including DISCERN and GQS have been focused on information on common patient presentations and symptoms, anatomy, diagnosis, evaluation, treatment options, and post-treatment and/or postoperative course and expectations. From these valid tools we extracted the most important points and created HVSSC that concerned only HV. Our tool has 20 questions that include the most important aspects of HV that a patient and/or a physician must learn about this pathology. The HVSSC evaluates information on (1) common patient presentations and symptoms, (2) the anatomy of the first toe, (3) HV diagnosis and evaluation, (4) treatment options, and (5) the postoperative and/or post-treatment course and expectations. One point is assigned to each item; the maximum possible score is 20. The HVSSC categorizes videos as good (16–20), fair (11–15), poor (6–10), or very poor (0–5)[12] (Table 1).

Data were independently abstracted and organized into spreadsheets using Excel 2016 (Microsoft Corp., Redmond, WA, USA) by the two reviewers. After evaluation of data, a meeting was held by both observers. For debatable categorical data, the videos were watched again and a consensus was established and deemed as valid. For quantitative data, the mean value of both observations was deemed as valid.

Statistical analyses

Data were analyzed using SPSS software (ver. 22.0; IBM Corp., Armonk, NY, USA). Interobserver reliability for qualitative data was determined using the kappa coefficient (κ). A κ value of 0.81–1.00 indicates almost perfect agreement, while 0.61–0.80 reflects substantial agreement, 0.21–0.60 moderate agreement, and ≤0.20 slight agreement. The intraclass correlation coefficient (ICC) was calculated to determine the reliability of the quantitative data evaluations by the observers. An ICC < 0.50 indicates poor agreement, while a value of 0.50–0.74 reflects fair agreement, 0.75–0.90 good agreement, and 0.91–1 excellent agreement. The DISCERN, JAMA, GQS, and HVSSC scores were analyzed separately. The normality of the data distribution was evaluated by the Shapiro–Wilk test. The Kruskal–Wallis *H*-test combined with Monte Carlo simulations was used for comparison of nonparametric variables. Dunn's test was used for post hoc analysis. The Kruskal–Wallis test was employed to compare quantitative data between independent groups. Categorical variables were compared using the Pearson Chi-squared test and Monte Carlo simulations with Fisher's exact test. Spearman's correlation coefficient (p) was used to evaluate the associations between quantitative variables, and the Pearson correlation coefficient (rho) was calculated as a measure of the associations between categorical variables (1 = perfect positive correlation and -1 = perfect negative correlation). Quantitative variables are expressed as mean ± standard deviation or median with interquartile range (IQR). Qualitative variables are expressed as frequencies or ratios. p-Values < 0.05 were considered to indicate statistical significance.

Results

Fifty-two videos were included in the final analysis and 91 were excluded (Figure 1). The mean duration was 7.21 ± 9.084 minutes and the mean number of views was 101,191.53 ± 154,519.038. The details of the extracted data are shown in Tables 2 and 3. Of the 52 included

Table 1. Hallux valgus-specific survey criteria.

	Score (20)
Presentation of the condition	
Describes symptoms	1
Describes patient population	1
Information about condition	
Describes anatomy and/or function of the first toe	1
Differentiates bunion hypertrophy, bursitis and hallux valgus	1
Explains healing potential of bursitis and hallux valgus	1
Mentions progression to arthritis or other deformities like ovverriding second toe	1
Diagnosis and evaluation	
Mentions physical examination and findings	1
Discusses inability for routine non-weight-bearing radiographs to evaluate	1
Discusses use of weight-bearing X-rays as gold-standard diagnostic imaging modality	1
Discusses types of deformity (congenital HV, HV interfalangeus, TM joint instability)	1
Describes surgical candidates (middle aged adults or patients with symptoms impacting function or quality of life)	1
Describes nonsurgical candidates (young adults with few symptoms)	1
Treatment	
Mentions conservative treatment	1
Mentions bunionectomy and other pathologies that may be addressed concomitantly	1
Describes single osteotomy	1
Describes multilevel osteotomies and soft tissue balance surgeries	1
Postoperative and/or post-treatment course	
Describes succes rate, complications and outcomes	1
Mentions weight-bearing restrictions	1
Mentions physical therapy	1
Outlines return-to-function timeline	1

TM, tarso-metatarsal; HV, hallux valgus.
Classification; Sum of points indicates that the content is good for 16–20 points, fair for 11–15 points, poor for 6–10 points and very poor for 0–5 points.

videos, 15 were posted by medical companies involved in the production of surgical implants and orthopedic products, while 20 were posted by non-surgeon physicians, 16 by surgeons presenting their techniques, and 1 by a patient who recounted his personal experience during HV surgery (Table 3).

Table 2. Descriptives of the data and statistical analysis of data among different video sources.

	Mean	Std. deviation	Minimum	Maximum	Median	IQR	*P*
DISCERN score	32.71	±16.725	16	78	25.5	18	**0.02**
JAMA score	1.85	±0.826	1	4	2	1	0.309
GQS score	2.02	±1.180	1	5	2	2	**0.018**
HVSSC score	4.83	±4.536	0	20	3	4	**0.002**
Time since upload (days)	2012.79	1071.216	195	4300	2057	1948	0.092
Runtime (minutes)	7.21	±9084	1	48	4	7	**0.033**
View (*n*)	101,191.53	±154,519.038	10,760	798,843	30459.5	119058	0.321
Like (*n*)	398.21	±861.102	0	5529	106	339	**0.043**
Dislike (*n*)	35.46	±56.855	0	249	8	45	0.177
Like ratio	83.9652	±22.83751	0.00	98.33	91.14	10.8	0.61
View ratio	56.6011	±82.08954	0.04	373.64	24.06	41.68	0.205
VPI	42.9140	±64.82490	0.00	311.91	18.69	39.23	**0.047**

P: Kruskal–Wallis Test evaluating significance of the data between different video sources; VPI: Video Power Index; IQR: interquartile range; JAMA: Journal of the American Medical Association benchmark criteria; GQS: global quality score, DISCERN tool score; HVSSC: hallux valgus-specific survey criteria.
Bold indicates statistically significant values.

For the DISCERN tool, JAMA benchmark criteria, GQS, and HVSSC, good-to-excellent agreement was observed between the two observers (ICC = 0.87, 0.81, 0.89, and 0.95, respectively). Almost perfect agreement was observed for video source, content, and surgery type ($\kappa = 0.98$, 0.94, and 0.97, respectively). Of the 52 videos, 28 were concerned with particular surgical techniques including distal osteotomies, soft tissue procedures, proximal osteotomies, the chevron and Lapidus techniques, minimally invasive distal osteotomies, and bunionectomy. The remaining 24 offered general information on HV treatment including non-specific surgical options, advertisements, personal experiences, and postoperative rehabilitation (Table 3). Most videos posted by surgeons and physicians dealt with surgical techniques and general treatment options, with the goal of educating patients. However, the commercial videos were mainly implant and orthosis advertisements ($p = 0.0001$). Animated surgical videos and videos describing regular surgical techniques were similarly popular (Table 3).

The VPI scores and numbers of likes were significantly higher for videos posted by physicians and surgeons ($p = 0.047$ and 0.043, respectively). Runtime was significantly longer for videos posted by surgeons ($p = 0.043$) (Table 2).

DISCERN: Only 10 (19.2%) of the 52 videos were classified as good or excellent in terms of reliability and quality; 6 of these were posted by physicians and 4 by surgeons. The remaining videos were unreliable and of poor quality.

JAMA: Only two (3.8%) videos were completely accurate, useful, and reliable. One video was posted by a physician and the other by a surgeon. Twenty (38.5%) videos were unacceptable and 30 (57.6%) somewhat acceptable.

GQS: The educational value of only six (11.6%) videos was good or excellent; five of these were posted by physicians and one by a surgeon.

HVSSC: The quality, educational value, and reliability of only five (9.6%) videos were fair or good; four of these were posted by physicians and one by a surgeon. The remaining 47 (90.4%) videos were very poor or poor.

The DISCERN, GQS, and HVSSC scores were significantly higher for videos posted by surgeons and physicians ($p = 0.02$, 0.018, and 0.02, respectively). However, the JAMA scores were similar for all videos (Table 2). The DISCERN, JAMA, and HVSSC classifications did not differ by video source (Table 3). Although the Chi-squared test indicated that significantly more good and excellent videos (as rated by the GQS) were posted by physicians ($p = 0.045$) (Table 3), the Pearson correlation test revealed no correlation between video source and the classification of any survey tool (rho = 0.02, $p = 0.889$). No significant correlation was detected among video source, video content, DISCERN, JAMA, or HVSSC

Table 3. Descriptives and analysis of survey tools after classification of the videos among different video sources.

Evaluation tool		Video source					
		Physician ($n=20$)	Commercial ($n=15$)	Patient ($n=1$)	Allied surgeon ($n=16$)	Total ($n=52$)	*P*
DISCERN class	Very poor (n/%)	9 (45%)	12 (80%)	1 (100%)	4 (25%)	26 (52)	0.213
	Poor (n/%)	4 (20%)	3 (20%)	0 (0%)	5 (31.25%)	12 (23.1%)	
	Fair (n/%)	1 (5%)	0 (0%)	0 (0%)	3 (18.75%)	4 (7.7%)	
	Good (n/%)	3 (15%)	0 (0%)	0 (0%)	3 (18.75%)	6 (11.5%)	
	Excellent (n/%)	3 (15%)	0 (0%)	0 (0%)	1 (6.25%)	4 (7.7%)	
JAMA class	Insufficient (n/%)	9 (45%)	8 (53.3%)	0 (0%)	3 (18.75%)	20 (38.5%)	0.461
	Partially sufficient (n/%)	10 (50%)	7 (46.7%)	1 (100%)	12 (75%)	30 (57.6%)	
	Completely sufficient (n/%)	1 (5%)	0 (0%)	0 (0%)	1 (6.25%)	2 (3.8%)	
GQS class	Very poor (n/%)	9 (45%)	10 (66.7%)	1 (100%)	2 (12.5%)	22 (42.3%)	0.045
	Poor (n/%)	2 (10%)	4 (26.7%)	0 (0%)	9 (56.25%)	15 (28.8%)	
	Suboptimal (n/%)	4 (20%)	1 (6.7%)	0 (0%)	4 (25%)	9 (17.3%)	
	Good (n/%)	3 (15%)	0 (0%)	0 (0%)	0 (0%)	3 (5.8%)	
	Excellent (n/%)	2 (10%)	0 (0%)	0 (0%)	1 (6.25%)	3 (5.8%)	
HVSS class	Very poor (n/%)	10 (50%)	15 (100%)	1 (100%)	11 (68.75%)	37 (71.2%)	0.189
	Poor (n/%)	6 (30%)	0 (0%)	0 (0%)	4 (25%)	10 (19,2%)	
	Fair (n/%)	2 (10%)	0 (0%)	0 (0%)	0 (0%)	2 (3.8%)	
	Good (n/%)	2 (10%)	0 (0%)	0 (0%)	1 (6.25%)	3 (5.8%)	
Video content	Surgical techniques (n/%)	5 (25%)	1 (6.7%)	0 (0%)	15 (93.75%)	21 (40.4%)	0.0001
	General treatment options including surgery (n/%)	10 (50%)	1 (6.7%)	0 (0%)	0 (0%)	11 (21.2%)	
	Animated surgical techniqes (n/%)	2 (10%)	3 (20%)	0 (0%)	1 (6.25%)	6 (11.5%)	
	Advertisement of orthosis (n/%)	0 (0%)	7 (46.7%)	0 (0%)	0 (0%)	7 (13.5%)	
	Advertisement of implant (n/%)	0 (0%)	3 (20%)	0 (0%)	0 (0%)	3 (5.8%)	
	Postop rehabilitation (n/%)	3 (15%)	0 (0%)	0 (0%)	0 (0%)	3 (5.8%)	
	Patient experience (n/%)	0 (0%)	0 (0%)	1 (100%)	0 (0%)	1 (1.9%)	

P: Chi-squared Test evaluating the significance of classes of survey tools and video content for different video sources; JAMA: Journal of the American Medical Association benchmark criteria; GQS: global quality score, DISCERN tool score; HVSSC: Hallux valgus-specific survey criteria.

Table 4. Correlations between HVSSC, video source, video content and other survey tools.

	Video source	Video content	DISCERN class	JAMA class	HVSSC class	GQS class
Video source (rho;P)	**1**	**−0.354;** 0.010	0.087; 0.537	0.237; 0.091	−0.145; 0.304	0.020; 0.889
Video content (rho;P)	**−0.354;** 0.010	**1**	−0.203; 0.150	−0.197; 0.162	−0.213; 0.129	−0.200; 0.156
DISCERN class (rho;P)	0.087; 0.537	−0.203; 0.150	**1**	**0.576; 0.0001**	**0.770; 0.0001**	**0.831; 0.0001**
JAMA class (rho;P)	0.237; 0.091	−0.197; 0.162	**0.576; 0.0001**	**1**	**0.425; 0.002**	**0.593; 0.000**
HVSSC class (rho;P)	−0.145; 0.304	−0.213; 0.129	**0.770; 0.0001**	**0.425; 0.002**	**1**	**0.853; 0.0001**
GQS class (rho;P)	0.020; 0.889	-0.200; 0.156	**0.831; 0.0001**	**0.593; 0.0001**	**0.853; 0.0001**	**1**

rho;P: Pearson Correlation Coefficient; Significance level (*P* value); JAMA: Journal of the American Medical Association benchmark criteria; GQS: global quality score, DISCERN tool score; HVSSC: hallux valgus-specific survey criteria.
Bold indicates statistically significant values.

classification, number of views, number of dislike or the like ratio (Table 4). Also, we found no correlation among time since uploading, number of views, or the DISCERN, JAMA, GQS, or HVSSC scores (Table 5). We found a positive correlation between DISCERN, HVSSC and GQS scores and video durations ($p < 0.05$, rho = 0.439, 0.299, and 0.289, respectively). No relationship between JAMA scores and video durations was found. Also, there was no statistically significant relationship between duration of video and number of views, likes, VR, LR, and VPI.

Weak correlations of video source with the VR and VPI were found (Table 5). Videos posted by physicians and surgeons tended to be more popular. Although no correlations of the DISCERN, JAMA, or GQS scores with the VR or VPI were detected, the HVSSC score was correlated with both the number of views and the VR. We found no correlations of popularity indicators with the DISCERN or JAMA classifications**.** However, the HVSSC and GQS classifications were correlated with both the VR and VPI (Table 5). We found moderate correlations of the JAMA classifications and with those of DISCERN, GQS, and HVSSC. The correlations among the DISCERN, GQS, and HVSSC classifications were relatively strong (rho = 0.770, 0.853, 0.831, $p = 0.001$) (Table 4).

Discussion

The study evaluated the reliability and quality of YouTube videos on HV to assess the efficiency and reliability of the new HVSSC tool for determining whether the videos were reliable and of high quality. Videos posted by physicians and surgeons seem to be more popular than those posted by medical companies. However, even though videos posted by physicians and surgeons tend to be of better quality, they may still be unreliable and misleading. The HVSSC alone could be used to evaluate the quality, educational value, and reliability of videos on HV. Moreover, when producing such videos, referring to the HVSSC could promote quality and thus more views.

Most HV patients expect that treatment will relieve pain in the great toe and allow conventional shoes to be worn.[32,33] Esthetic appearance is the second most important factor.[34,35] Physicians treating HV should assess all factors that might affect the results and manage patient expectations via preoperative consultations. Thorough assessment is mandatory if a patient requests corrective surgery.[33] However, many patients do not consult physicians and thus lack adequate information; often, they search the Internet (particularly YouTube) for possible treatments and outcomes. Many studies have documented misinformation on the Internet, characterized by poor quality and reliability. Our PubMed search yielded 349 studies, from all branches of the health sciences, published over the past decade evaluating the reliability of YouTube videos. The general consensus was that online videos provided poor information on surgical interventions. However, we detect two studies that have concluded that videos posted by professional physicians on particular aspects may be reliable. Morra et al. have found the YouTube videos to be reliable on bladder pain syndrome.[13] Gerundo et al. have evaluated the quality of YouTube videos on personal protective equipment in COVID-19. Authors have found the posted

Table 5. Correlations between scores and classes of survey tools and popularity of the videos.

	View	Like	Dislike	Like ratio	View ratio	VPI
Video source (p);P	−0.184;	**−0.367;**	−0.204;	−0.132;	**−0.279;**	**−0.371;**
	0.191	**0.009**	0.147	0.35	**0.045**	**0.007**
DISCERN score (p);P	0.211;	−0.004;	0.001;	0.097;	0.148;	0.156;
	0.133	0.979	0.996	0.493	0.295	0.269
JAMA score (p);P	0.128;	−0.142;	−0.025;	−0.183;	0.096;	0.048;
	0.365	0.316	0.858	0.194	0.498	0.737
HVSSC score (p);P	**0.374;**	0.193;	0.258;	−0.014;	**0.290;**	0.229;
	0.006	0.171	0.065	0.923	**0.037**	0.102
GQS score (p);P	0.212;	0.089;	0.133;	0.113;	0.215;	0.238;
	0.131	0.533	0.346	0.425	0.125	0.09
DISCERN class (rho);P	0.125; 0.379	0.009; 0.950	0.082; 0.564	0.164; 0.247	0.132; 0.349	0.183; 0.194
JAMA class (rho);P	−0.001; 0.996	−0.214; 0.128	−0.097; 0.493	−0.134; 0.342	0.066; 0.640	0.006; 0.968
HVSSC class (rho);P	**0.373; 0.007**	0.166 0.239	**0.377; 0.006**	0.040; 0.777	**0.379; 0.006**	**0.381; 0.005**
GQS class (rho);P	0.233; 0.096	0.040; 0.780	0.186; 0.186	0.128; 0.366	**0.282; 0.043**	**0.274; 0.049**

(p);P: Spearman Correlation Coefficient; significance level (*P* value). (rho);P: Pearson Correlation Coefficient; significance level (*P* value). VPI: Video Power Index; JAMA: Journal of the American Medical Association benchmark criteria; GQS: global quality score, DISCERN tool score; HVSSC: hallux valgus-specific survey criteria. Bold indicates statistically significant values.

videos to be reliable as a source of information despite the fact that some videos are inaccurate.[36] Of the studies, 13 focused on orthopedic conditions, assessing the reliability and educational value of videos on HV,[19–21] rotator cuff surgery,[14] anterior cruciate ligament reconstruction,[22,29] meniscus surgery,[28] knee arthroplasty,[30] lower-limb amputation,[31] physical shoulder examination,[37] femoroacetabular impingement,[38] and knee arthrocentesis.[39] All studies focused on orthopedic conditions were agreed that online videos provided poor information. Moreover, a study evaluating the quality of the videos on testicular cancer have shown that the quality of the videos doesn't increase by the time despite the fact that the number of studies showing misinformation on YouTube videos is increasing.[16]

All three studies published between 2021 and 2023 had evaluated only the reliability and quality of HV videos on YouTube rather than presenting a better tool for creating accurate videos.[19–21] Those three studies had used similar methods with our study to evaluate the reliability and quality of videos and, in line with our study, concluded that the quality and the reliability of HV-related videos are low. Although authors had developed various specific tools for HV videos in all three studies, no detailed comments were made about their reliability and whether they were correlated with JAMA, DISCERN, GQS and other tools in order to produce better videos. In our study, it was revealed that the developed HVSSC was correlated with DISCERN, GQS and other parameters and could be effective in producing a reliable information source.

Given the rampant misinformation on video platforms, physicians must provide good information to patients. Surgeons, medical students, and residents lacking experience of specific medical situations also access YouTube.[40] However, it must be emphasized that most videos are unhelpful.[31] Few high-quality and thorough educational videos on HV are available. It is important to create peer-reviewed educational and surgical videos. There are a

few sites sharing high-quality videos. However, only professionals have access to these high-quality videos using member-only sites like VuMedi.[40] Only high-quality, relevant, reviewed, and approved videos are posted on these sites to avoid misinformation. Sites, similar to scientific journals, can be set up to inform ordinary people, where videos are submitted, reviewed, and presented if found appropriate. During the COVID-19 pandemic, most medical associations, universities, and training hospitals "went online," creating a large amount of information. After a review process carried out by relevant health associations, this source of information could be placed on high-quality websites accessible by patients. We believe that creating these kinds of sites supervised by medical associations could prevent unnecessary outpatient visits, inappropriate industry advertisements, unnecessary competition among physicians and surgeons, and ethical violations.

The use of pathology-specific tools to evaluate online information is important, and was a novel aspect of this study. The DISCERN tool and JAMA criteria may be inappropriate for assessment of specific pathologies, as they were developed to evaluate text.[22] DISCERN, the GQS, and the HVSSC evaluate videos in a similar manner, while the JAMA criteria assess video accessibility. As such, strong correlations were evident among the DISCERN, GQS, and HVSSC scores, while correlations with the JAMA scores were weak. In addition, there are different tools that are used to evaluate the quality of posted videos. Of these, patient education materials assessment tool (PEMAT), and misinformation scale are reported to be valid for evaluation of quality of video contents.[16,41]

Video source was weakly correlated with both the VR and VPI (Table 5). Videos posted by physicians and surgeons tended to be more popular. Although videos can be posted by industry for commercial uses, videos of specific aspects of some procedures are generally posted by professional physicians.[42] The relationships between the popularity indices and survey scores were evaluated before and after classification; we aimed to identify the best index for future users. Some studies evaluated tools in terms of their classifications, while others focused on their scores.[29–31,37–39] We found that neither the scores nor classifications of the DISCERN tool and JAMA criteria correlated with the popularity indices, similar to the GQS scores. However, the GQS classifications correlated with both the VR and VPI. The HVSSC scores correlated with both the number of views and VR. The HVSSC classifications correlated with the number of views, numbers of dislikes, VR, and VPI; the HVSSC performed better than the other tools. During the conversion of scores to classifications, cases with wide lower and upper boundaries undermined possible correlations between the survey data and popularity indicators. The HVSSC employs a 20-point scale, which may explain why it performed best. This study is the first of its kind; future studies on videos pertaining to HV should also evaluate the accuracy of survey tools in terms of assessing video quality, educational value, and reliability.

While creating a qualified video, one of the factors that should be considered apart from the content is the video length. As expected, a long video can deliver quality content as needed.[14] However, it can be expected that the number of views in a video that is kept too long will be less than a short one.[43] Sarı and Umur have reported a positive correlation between DISCERN scores and duration of videos related to HV.[21] We also found a positive correlation between DISCERN, HVSSC and GQS scores as expected. But there was no statistically significant relationship between duration of video and number of views, likes, VR, LR, and VPI. This can be attributed to people being more selective about their health issues and paying more attention to the content rather than the video duration. However, the absence of a significant statistical relationship between the JAMA and DISCERN scores and the number of view, dislike, like, VR, LR, and VPI contradicts this judgment. Nevertheless, the fact that there is a positive correlation between HVSSC, GQS and number of views, dislike, VR, and VPI (Table 5) suggests that this theory may be true particularly for people searching videos related to HV. As a result, there is no definite information about how long the ideal duration of an informative video on medical subjects should be, and this issue should be examined in future studies.

Our work had several limitations. First, the YouTube search algorithm is unknown; the order in which results appear may be affected by the IP address, regional codes, personal search history, and commercials. Although our inclusion only of videos with over 10,000 views may have reduced the risk of bias, YouTube might preferentially show channels with larger numbers of subscribers' videos for commercial reasons. This affects the number of views that alters the VPI and VR values as those values are calculated using number of views. Second, as this is the first study using the novel HV pathology-specific HVSSC tool, further studies are needed to validate our results.

Conclusion

The reliability of HV-related videos on YouTube is low for professionals and patients. Neither the scores nor classifications of the DISCERN tool and JAMA criteria correlated with the popularity indices. However, the GQS classifications correlated with both the VR and VPI. The HVSSC classifications correlated with the number of views, numbers of dislikes, VR, VPI and performed better than the other tools, The HVSSC tool can be used to evaluate the quality, educational value, and reliability of videos on HV. Videos created based on the HVSSC should be more reliable and of higher quality.

Acknowledgements: All authors had made appropriate contributions to the conception and design of the study. The English in this document has been checked by at least two professional editors, both native speakers of English. For a certificate, please see: http://www.textcheck.com/certificate/kV4vxT.

Contributorship: MSS: design of the study, statistical analysis, development of checklists, interpretation of data, writing of the article, drafting and revision of the article. SB: acquisition of data, co-writing of the article. BK: contributed to the acquisition and analysis of the data. SKÇ: contributed to the interpretation of the data and design of the research. All authors critically revised the manuscript, agree to be fully accountable for ensuring the integrity and accuracy of the work, and read and approved the final manuscript.

Declaration of conflicting interests: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethics approval: This article does not contain any studies with human participants performed by any of the authors.

Funding: The author(s) received no financial support for the research, authorship, and/or publication of this article.

Guarantor: MSS

Informed consent: Not applicable. This research does not involve any human participant or animal subject.

ORCID iDs: Mehmet Salih Söylemez https://orcid.org/0000-0002-0828-0145

References

1. Biz C, Favero L, Stecco C, et al. Hypermobility of the first ray in ballet dancer. *Muscles Ligaments Tendons J* 2013; 2: 282–288.
2. Biz C, Maso G, Malgarini E, et al. Hypermobility of the first ray: the Cinderella of the measurements conventionally assessed for correction of hallux valgus. *Acta Biomed* 2020; 91: 47–59.
3. Hecht PJ and Lin TJ. Hallux valgus. *Med Clin North Am* 2014; 98: 227–232.
4. Smyth NA and Aiyer AA. Introduction: why are there so many different surgeries for hallux valgus? *Foot Ankle Clin* 2018; 23: 171–182.
5. Wülker N and Mittag F. The treatment of hallux valgus. *Dtsch Arztebl Int* 2012; 109: 857–867.
6. Goh GS, Tay AYW, Thever Y, et al. Effect of age on clinical and radiological outcomes of hallux Valgus surgery. *Foot Ankle Int* 2021; 42: 798–804.
7. Curtin M, Murphy E, Bryan C, et al. Scarf osteotomy without internal fixation for correction of hallux valgus: a clinical and radiographic review of 148 cases. *Foot Ankle Surg* 2018; 24: 252–258.
8. Biz C, Crimì A, Fantoni I, et al. Functional and radiographic outcomes of minimally invasive intramedullary nail device (MIIND) for moderate to severe hallux valgus. *Foot Ankle Int* 2021; 42: 409–424.
9. Biz C, Fosser M, Dalmau-Pastor M, et al. Functional and radiographic outcomes of hallux valgus correction by mini-invasive surgery with reverdin-isham and akin percutaneous osteotomies: a longitudinal prospective study with a 48-month follow-up. *J Orthop Surg Res* 2016; 11: 57.
10. Uygur E, Özkan NK, Akan K, et al. A comparison of chevron and lindgren-turan osteotomy techniques in hallux valgus surgery: a prospective randomized controlled study. *Acta Orthop Traumatol Turc* 2016; 50: 255–261.
11. Makhdom AM, Sinno H, Aldebeyan S, et al. Bilateral hallux valgus: a utility outcome score assessment. *J Foot Ankle Surg* 2016; 55: 944–947.
12. Radl R, Leithner A, Zacherl M, et al. The influence of personality traits on the subjective outcome of operative hallux valgus correction. *Int Orthop* 2004; 28: 303–306.
13. Morra S, Collà Ruvolo C, Napolitano L, et al. YouTube™ as a source of information on bladder pain syndrome: a contemporary analysis. *Neurourol Urodyn* 2022; 41: 237–245.
14. Celik H, Polat O, Ozcan C, et al. Assessment of the quality and reliability of the information on rotator cuff repair on YouTube. *Orthop Traumatol Surg Res* 2020; 106: 31–34.
15. Cilio S, Collà Ruvolo C, Turco C, et al. Analysis of quality information provided by "Dr. YouTubeTM" on phimosis. *Int J Impot Res* 2022; 24: 1–6.
16. Di Bello F, Collà Ruvolo C, Cilio S, et al. Testicular cancer and YouTube: what do you expect from a social media platform? *Int J Urol* 2022; 29: 685–691.
17. Collà Ruvolo C, Califano G, Tuccillo A, et al. YouTube™ as a source of information on placenta accreta: a quality analysis. *Eur J Obstet Gynecol Reprod Biol* 2022; 272: 82–87.
18. Eksi Ozsoy H. Evaluation of YouTube videos about smile design using the discern tool and Journal of the American Medical Association benchmarks. *J Prosthet Dent* 2020; 20: 22–26.
19. Tekin SB and Bozgeyik B. Quality and content analysis of hallux valgus videos on YouTube®. *J Foot Ankle Surg* 2023; 62: 85–90.
20. Uzun M, Cingoz T, Duran ME, et al. The videos on YouTube® related to hallux valgus surgery have insufficient information. *Foot Ankle Surg* 2022; 28: 414–417.
21. Sari E and Umur LF. Quality analysis of hallux valgus videos on YouTube. *J Am Podiatr Med Assoc* 2021; 111(5). doi: 10.7547/20-191
22. Cassidy JT, Fitzgerald E, Cassidy ES, et al. YouTube provides poor information regarding anterior cruciate ligament injury and reconstruction. *Knee Surg Sports Traumatol Arthrosc* 2018; 26: 840–845.
23. Erdem H and Sisik A. The reliability of bariatric surgery videos in YouTube platform. *Obes Surg* 2018; 28: 712–716.
24. Silberg WM, Lundberg GD and 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: 1244–5.

25. Bernard A, Langille M, Hughes S, et al. A systematic review of patient inflammatory bowel disease information resources on the World Wide Web. *Am J Gastroenterol* 2007; 102: 2070–2077.
26. Charnock D, Shepperd S, Needham G, et al. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. *J Epidemiol Community Health* 1999; 53: 105–11.
27. Olkun HK and Demirkaya AA. Evaluation of internet information about lingual orthodontics using DISCERN and JAMA tools. *Turk J Orthod* 2018; 31: 50–54. doi:10.5152/TurkJOrthod.2018.17042
28. Kunze KN, Krivicich LM, Verma NN, et al. Quality of online video resources concerning patient education for the meniscus: a YouTube-based quality-control study. *Arthroscopy* 2020; 36: 233–238.
29. Springer B, Bechler U, Koller U, et al. Online videos provide poor information quality, reliability, and accuracy regarding rehabilitation and return to sport after anterior cruciate ligament reconstruction. *Arthroscopy* 2020; 36(12): 3037–3047.
30. Wong M, Desai B, Bautista M, et al. YouTube is a poor source of patient information for knee arthroplasty and knee osteoarthritis. *Arthroplast Today* 2018; 5: 78–82.
31. Yammine K and Assi C. Educational assessment of the major lower limb amputations videos on YouTube. *Vascular* 2020; 28: 536–541.
32. Harb Z, Kokkinakis M, Ismail H, et al. Adolescent hallux valgus: a systematic review of outcomes following surgery. *J Child Orthop* 2015; 9: 105–112.
33. Schneider W and Knahr K. Surgery for hallux valgus. The expectations of patients and surgeons. *Int Orthop* 2001; 25: 382–385.
34. Saro C, Johnson DN, Martinez De Aragón J, et al. Reliability of radiological and cosmetic measurements in hallux valgus. *Acta Radiol* 2005; 46: 843–851.
35. Robinson C, Bhosale A and Pillai A. Footwear modification following hallux valgus surgery: the all-or-none phenomenon. *World J Methodol* 2016; 6: 171–180.
36. Gerundo G, Collà Ruvolo C, Puzone B, et al. Evidence-based quality and analysis of YouTube videos after one year of pandemic. *Am J Infect Control* 2022; 50: 300–305.
37. Urch E, Taylor SA, Cody E, et al. The quality of open-access video-based orthopaedic instructional content for the shoulder physical exam is inconsistent. *HSS J* 2016; 12: 209–215.
38. MacLeod MG, Hoppe DJ, Simunovic N, et al. YouTube as an information source for femoroacetabular impingement: a systematic review of video content. *Arthroscopy* 2015; 31: 136–142.
39. Fischer J, Geurts J, Valderrabano V, et al. Educational quality of YouTube videos on knee arthrocentesis. *J Clin Rheumatol* 2013; 19: 373–376.
40. Rogers MJ, Zeidan M, Flinders ZS, et al. Educational resource utilization by current orthopaedic surgical residents: a nationwide survey. *J Am Acad Orthop Surg Glob Res Rev* 2019; 3: 41.
41. Capece M, Di Giovanni A, Cirigliano L, et al. YouTube as a source of information on penile prosthesis. *Andrologia* 2022; 54: e14246.
42. Turco C, Collà Ruvolo C, Cilio S, et al. Are videos a reliable information tool for internet users? *Arch Ital Urol Androl* 2022; 94: 57–61.
43. Biggs TC, Bird JH, Harries PG, et al. YouTube as a source of information on rhinosinusitis: the good, the bad and the ugly. *J Laryngol Otol* 2013; 127: 749–754.