*Original Research Article*

DIGITAL HEALTH

DIGITAL HEALTH
Volume 11: 1–14

DOI: 10.1177/20552076251317578
journals.sagepub.com/home/dhj

Sage

# Health information analysis of cryptorchidism-related short videos: Analyzing quality and reliability

**Yuqiang Sun†, Xingjian Liu†, Xintao Zhang, Qiongqian Xu and Aiwu Li**

## Abstract

**Objective:** Cryptorchidism is a common congenital anomaly in children. Short video content has increased awareness and urged timely intervention, but there is a lack of understanding about the quality and reliability of these videos. This survey assesses the quality and credibility of cryptorchidism-related videos on Chinese short video platforms to ensure accurate information for caregivers and healthcare professionals.

**Methods:** We analyzed 228 videos from TikTok, Bilibili, and Kwai on 16 May 2024. Using the Journal of the American Medical Association (JAMA) criteria, Global Quality Score (GQS), and modified DISCERN score, we evaluated the videos based on six factors: definition, risk factors, symptoms, tests, treatment, and outcome.

**Results:** Video quality varied significantly across platforms ($P = .03$). Medical professionals' videos had higher JAMA scores than those by nonmedical professionals ($P < .001$), but no significant differences were found in GQS ($P = .633$) or modified DISCERN scores ($P = .853$). Overall, video quality was suboptimal, focusing mainly on treatment, symptoms, and prognosis, while neglecting tests and risk factors. Longer videos correlated with higher GQS ($r = .34$, $P < .001$) and modified DISCERN scores ($r = .39$, $P < .001$), while newer videos were of higher quality (GQS: $r = -.19$, $P < .005$; modified DISCERN: $r = -.25$, $P < .001$). Higher engagement (comments: $r = -.16$, $P = .01$; shares: $r = -.18$, $P = .04$) was associated with lower quality.

**Conclusion:** Overall, the videos' quality and reliability were moderate, with professional content providing more accurate insights. This highlights the need for improved regulation of medical information on short video platforms.

## Keywords

Digital health, eHealth, health informatics, social media, pediatrics, public health

Submission date: 13 September 2024; Acceptance date: 16 January 2025

## Introduction

Cryptorchidism, often also referred to as undescended testis, is one of the most common disorders in newborns, and there are two types of cryptorchidism: congenital cryptorchidism and acquired cryptorchidism.[1] The prevalence of cryptorchidism in the neonatal population is 90–100 per 10,000.[2] Notably, cryptorchidism is an important risk factor for infertility and testicular germ cell tumors.[3] Early detection of cryptorchidism is critical for the efficacy of surgical treatment and postoperative outcomes.[4] However, public awareness of cryptorchidism remains limited, often leading to delays in surgical treatment.[5] Therefore, there is an urgent need for effective public health education approaches to increase public awareness of cryptorchidism.[6]

The increasing popularity of the Internet among the population has made it an important channel for patient

Department of Pediatric Surgery, Qilu Hospital of Shandong University, Jinan, China
†The first two authors contributed equally to this work and should be considered co-first authors.

**Corresponding author:**
Aiwu Li, Department of Pediatric Surgery, Qilu Hospital of Shandong University, Wenhua West Road 107#, Jinan 250012, China.
Email: liaiwu@qiluhospital.com

education.[7] Short videos in particular have become an important medium for disseminating information to residents.[8] This includes health-related content, and short videos play a crucial role in disseminating healthcare information to diverse communities.[9] However, a lack of regulation of the sources of video content can lead to a proliferation of unreliable information that has the potential to distort public perceptions.[10] Furthermore, the dynamic nature of health information requires video platforms to continuously improve quality and reliability.[11] Outdated or incorrect health information can adversely affect public health management.[12,13]

Upon retrieval and discovery, although the quality of English health science popularization short videos about cryptorchidism on YouTube has already been evaluated,[14] the quality and reliability of Chinese short videos related to cryptorchidism on TikTok, Bilibili, and Kwai have not yet been assessed. Hence, we selected the Chinese short video platforms for our research.[15] TikTok, Bilibili, and Kwai stand as China's most frequented short-video platforms, boasting the highest active user counts.[16] These platforms grant users easy access to a plethora of health-related short videos, including those about cryptorchidism. To date, the quality of cryptorchidism-related content on these short video platforms has not been thoroughly evaluated. This study aims to appraise the content, quality, and reliability of cryptorchidism-related videos on TikTok, Bilibili, and Kwai.

## Method

### Search strategy and data collection

In this cross-sectional analysis, the Chinese term "隐睾" (cryptorchidism) was employed as a search keyword by YQ S on 16 May 2024, across TikTok, Bilibili, and Kwai. The top 100 short videos (duration of more than 10 s to less than 10 min) retrieved for each keyword from each platform were compiled. Exclusion criteria encompassed duplicate content, irrelevance, or non-Chinese language videos (as depicted in Figure 1). XJ L identified short videos that could be included in the study according to the exclusion criteria, as illustrated in Figure 1.

The collated data for each video encompassed the title, uploader's identity, duration, counts of likes, comments, saves, shares, and days since the video was uploaded. The uploaders were divided into two categories, medical professionals and nonmedical professionals, where medical professionals include professional individuals (professional doctors, nonprofessional doctors and other healthcare professionals) and professional institutions, whereas nonmedical professionals include non-professionals (patients and Science bloggers) and nonprofessional institutions (news agencies and nonprofit organizations). Detailed definitions of these uploader classifications are delineated in Table S1 in the supplementary online material.

### Evaluation methodology

The assessment of video integrity, reliability, and quality utilized an established scoring system that was identical to the methodology used in a previous video assessment study.[17] The system categorized video content into six main categories: definitions, tests, risk factors, outcomes, signs/symptoms, and treatment/management. The specific delineation of these main content categories is shown in Table S2 in the online supplementary material. The quality of the content was assessed by two researchers who rated the designated content type of each video on a five-point Likert scale[18] on the following scale: 0 (no content), 0.5 (little content), 1 (some content), 1.5 (most content) and 2 (extensive content).

The Journal of the American Medical Association (JAMA) criteria were utilized to determine the videos' reliability. Each video could accrue up to four points, with one point allocated per criterion met; a higher tally signified enhanced reliability.[19] The Global Quality Score (GQS) employs a five-tiered scale to evaluate the overall information quality and its applicability for patients, with superior levels indicating elevated quality, coherence, inclusivity, and patient utility.[20] The modified DISCERN instrument, a standard for appraising consumer health information quality, has been extensively applied to video content.[21] It concentrates on the content's accessibility, impartiality, and source traceability. The scoring involves five queries, with affirmative responses scored as 1 and negatives as 0, culminating in a range from 0 to 5, where higher scores denote superior reliability and quality. Comprehensive details are available in Tables S3–S5 in the online supplementary material.

### Evaluation process

Evaluation of the video was performed by two senior physicians, XJ L and YQ S, from the Department of Pediatric Surgery at one of the largest tertiary care teaching hospitals in China. Their assessments were grounded in established guidelines pertinent to cryptorchidism. Discrepancies between the researchers were reconciled through consultations with AW L, a seasoned pediatric surgeon with over three decades of expertise. To quantify the consistency of the evaluations.

### Statistical analysis

Data analysis was conducted using SPSS software (version 25.0, IBM, Armonk, New York, USA). Given the nonparametric distribution of the data, the median and interquartile range (IQR) served as the descriptive statistics.

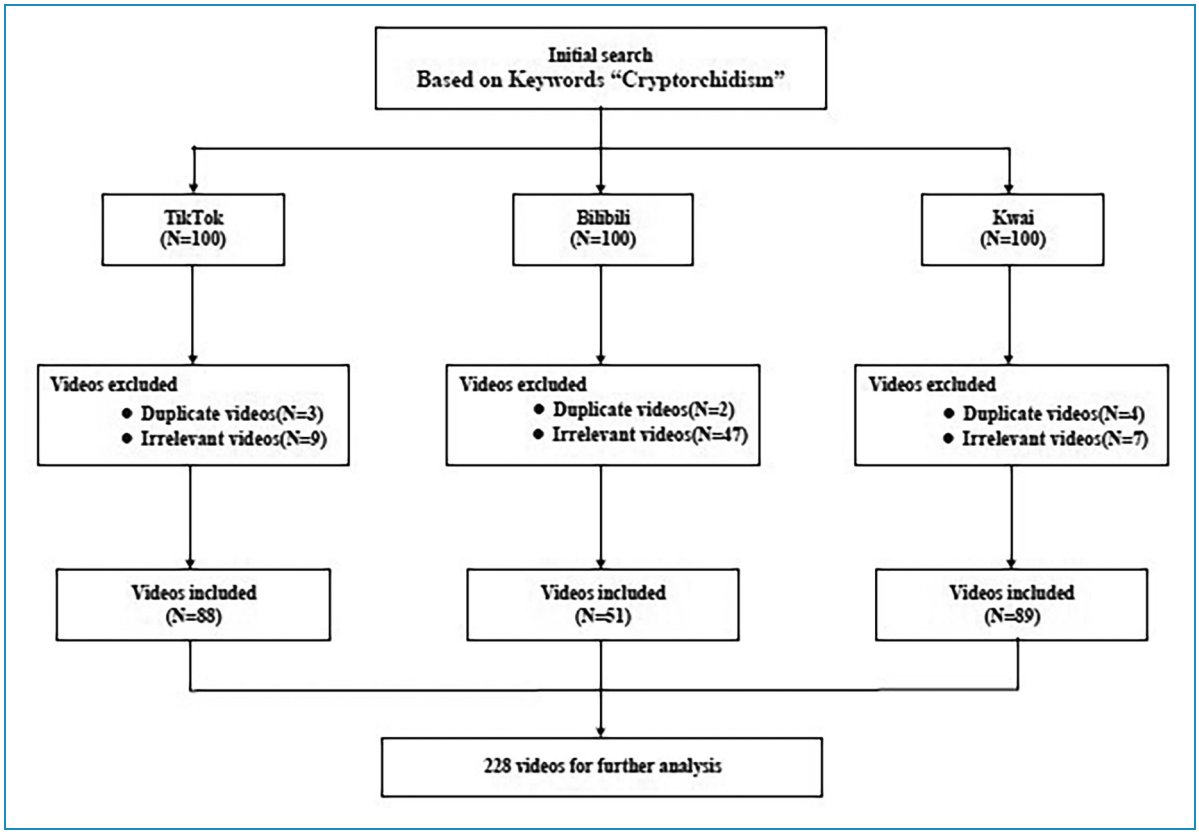


**Figure 1.** The selection process of short videos.

**Table 1.** Detailed characteristics of videos and comparison across platforms.

| Variable | TikTok ($n = 88$), median (IQR) | Bilibili ($n = 51$), median (IQR) | Kwai ($n = 89$), median (IQR) | *P* value |
|---|---|---|---|---|
| Likes, median (IQR) | 123.5 (55, 275.25) | 12 (4, 32) | 57 (27, 186.5) | <.001 |
| Comments, median (IQR) | 18 (6, 54.25) | 1 (0, 4) | 9 (2, 32.5) | <.001 |
| Shares, median (IQR) | 53 (14.25, 166.75) | 2 (0, 5) | NA | <.001 |
| Favorites, median (IQR) | 22 (10, 105) | 3 (1, 8) | 11 (6, 32) | <.001 |
| Days since published, median (IQR) | 267 (141.25, 474) | 587 (161, 1120) | 497 (296, 926.5) | <.001 |
| Duration (s), median (IQR) | 57 (32, 95) | 82 (47, 163) | 44 (31.5, 62) | <.001 |
| GQS score, median (IQR) | 3 (3, 4) | 3 (2, 4) | 3 (2, 3) | 0.03 |
| Modified DISCERN score, median (IQR) | 2 (2, 2) | 2 (2, 3) | 2 (2, 3) | <.001 |
| JAMA score, median (IQR) | 3 (3, 3) | 3 (2, 3) | 3 (3, 3) | 0.03 |

The Kolmogorov–Smirnov test verified the normalcy of the quantitative data's distribution. For normally distributed data, the independent samples *t*-test compared quantitative variables between two groups, while the one-way analysis of variance (ANOVA) compared variables across multiple groups. The Mann–Whitney *U* test was employed for two

**Table 2.** Characteristics of the videos across sources and content in TikTok.

| Variable | Likes | Comments | Shares | Favorites | Days Since Published | Duration |
|---|---|---|---|---|---|---|
| **Video sources ($n = 88$), median (IQR)** | | | | | | |
| Professional individuals ($n = 72$) | 111.5 (54, 252) | 16 (5, 46.5) | 45.5 (13, 150.5) | 21 (10, 93) | 249 (129.5, 430) | 49 (34, 79) |
| Nonprofessional individuals ($n = 9$) | 149 (94, 267) | 52 (24, 359) | 65 (15, 100) | 14 (6, 59) | 267 (182, 482) | 101 (22, 120) |
| Professional institutions ($n = 5$) | 150 (111, 215) | 15 (6, 20) | 96 (38, 160) | 38 (36, 97) | 392 (193, 544) | 89 (81, 95) |
| Nonprofessional institutions ($n = 2$) | 41500 (13000, 70000) | 10645 (1289, 20000) | 32909.5 (2819, 63000) | 1434 (363, 2505) | 702 (433, 971) | 5 (4, 6) |
| **Video content ($n = 88$), median (IQR)** | | | | | | |
| Risk factors ($n = 5$) | 38 (72, 126) | 6 (7, 19) | 21.5 (42, 92) | 11 (12, 19.5) | 189 (267, 596.5) | 43 (77, 303) |
| Definition ($n = 53$) | 51.5 (118, 327.5) | 4 (14, 50) | 14.5 (59, 211) | 10 (27, 120) | 148 (268, 489.5) | 44 (66.5, 106.5) |
| Signs/Symptoms ($n = 66$) | 55.75 (149, 434.25) | 5 (19, 78.5) | 15.75 (74.5, 346.25) | 12.75 (28.5, 152.5) | 151.75 (276, 497) | 43.5 (66, 108.5) |
| Test ($n = 36$) | 77.25 (162, 518) | 6 (21, 82.75) | 20 (94.5, 456.5) | 15.25 (58.5, 162.25) | 157.75 (269, 571) | 44 (74, 120) |
| Treatment/management ($n = 77$) | 52.5 (112, 260.5) | 5.5 (18, 68) | 14.5 (56, 156.5) | 9.5 (21, 81) | 142.5 (268, 466) | 43 (62, 100.5) |
| Outcome ($n = 64$) | 56.75 (111.5, 229.25) | 5.25 (17.5, 60.25) | 15 (45.5, 146.75) | 10.25 (19, 52) | 154.5 (304, 497) | 43 (62, 101) |

**Table 3.** Characteristics of the videos across sources and content in Bilibili.

| Variable | Likes | Comments | Shares | Favorites | Days Since Published | Duration |
|---|---|---|---|---|---|---|
| **Video sources ($n = 51$), median (IQR)** | | | | | | |
| Professional individuals ($n = 28$) | 16.5 (4.5, 35) | 1 (0, 3) | 1.5 (0, 6.5) | 3 (1, 6) | 534.5 (176, 924) | 49.5 (36.5, 105) |
| Nonprofessional individuals ($n = 10$) | 8.5 (6, 83) | 2.5 (0, 25) | 4 (1, 70) | 9 (1, 75) | 648 (209, 1089) | 228.5 (111, 467) |
| Professional institutions ($n = 2$) | 3 (1, 5) | 0 (0, 0) | 1 (0, 2) | 1 (1, 1) | 917.5 (44, 1791) | 120 (60, 180) |
| Nonprofessional institutions ($n = 11$) | 12 (2, 30) | 0 (0, 5) | 2 (0, 4) | 3 (1, 7) | 1212 (44, 1736) | 127 (50, 240) |
| **Video content ($n = 51$), median (IQR)** | | | | | | |
| Risk factors ($n = 8$) | 2.5 (8, 89.5) | 0 (0.5, 9.25) | 0.25 (1, 45.75) | 1 (4.5, 16.75) | 140.75 (323, 986.75) | 112.25 (164, 427.5) |
| Definition ($n = 33$) | 4 (7, 24) | 0 (0, 4) | 0 (2, 4.5) | 1 (2, 8.5) | 176 (587, 1179.5) | 39 (91, 224.5) |
| Signs/Symptoms ($n = 40$) | 5 (12.5, 31.75) | 0 (0.5, 4.75) | 0 (2, 7.25) | 1 (3, 8.75) | 143 (648, 1112.25) | 49 (105, 201.75) |
| Test ($n = 22$) | 6.25 (18.5, 36) | 0 (0.5, 4.75) | 0 (2.5, 8) | 1.5 (6, 9.75) | 126.5 (266, 719.5) | 68.75 (123.5, 285.5) |
| Treatment/ management (n = 33) | 5 (18, 44) | 0 (2, 12) | 0.5 (4, 17) | 1 (5, 15) | 143.5 (491, 939) | 63 (127, 241) |
| Outcome ($n = 36$) | 4.25 (17, 36) | 0 (1, 8) | 0 (3, 13.25) | 1 (4, 9.75) | 145.25 (532.5, 949.5) | 50 (117, 232.25) |

groups of data not adhering to a normal distribution. The Kruskal–Wallis test analyzed variance among multiple groups with skewed distribution data. The Pearson chi-square test evaluated differences across multiple groups. Bonferroni correction-adjusted *P*-values were utilized for post hoc binary comparisons in multi-group assessments. Spearman correlation analysis examined the relationships between quantitative variables. A *P*-value of less than .05 was considered statistically significant. Kappa statistics were computed, with values exceeding 0.8 indicating commendable interrater reliability for the assessment criteria.

## Results

### *General information and characteristics of the videos*

The general characteristics and quality reliability scores of the reviewed videos and summary of the analysis of video characteristics and quality reliability across platforms are listed in Table 1. Exclusions included 12 videos from TikTok (three duplicates, nine irrelevant), 49 videos from Bilibili (two duplicates, 47 irrelevant), and 11 videos from Kwai (four duplicates, seven irrelevant), with a final selection of 72 videos. Thus, a total of 228 videos were assigned for in-depth analysis, 88 from TikTok, 89 from Kwai, and 51 from Bilibili. Among the entire selection, TikTok videos secured the highest median figures for likes (123.5, interquartile range (IQR): 55–275.25, ($P < .001$)), comments (18, IQR: 6–54.25 ($P < .001$)), and favorites (22, IQR: 10–105 ($P < .001$)), while Bilibili videos boasted the lengthiest median duration (82, IQR: 47–163 ($P < .001$)). The general characteristics of the videos among the different video uploaders on the three platforms, as well as the general characteristics of the videos that have the main video evaluation content are detailed in

**Table 4.** Characteristics of the videos across sources and content in Kwai.

| Variable | Likes | Comments | Shares | Favorites | Days Since Published | Duration |
|---|---|---|---|---|---|---|
| **Video sources ($n = 89$), median (IQR)** | | | | | | |
| Professional individuals ($n = 79$) | 64 (30, 215) | 10 (2, 41) | NA | 11 (7, 34) | 497 (303, 918) | 43 (32, 63) |
| Nonprofessional individuals ($n = 4$) | 22 (15.5, 59.5) | 10 (5, 64.5) | NA | 10 (2, 18) | 629 (199, 1135.5) | 45 (26, 241.5) |
| Professional institutions ($n = 5$) | 45 (35, 52) | 3 (0, 9) | NA | 7 (6, 13) | 610 (216, 1303) | 50 (47, 56) |
| Nonprofessional institutions ($n = 1$) | 495 (495, 495) | 13 (13, 13) | NA | 33 (33, 33) | 275 (275, 275) | 23 (23, 23) |
| **Video content ($n = 89$), median (IQR)** | | | | | | |
| Risk factors ($n = 7$) | 19 (27, 1085) | 2 (6, 100) | NA | 6 (9, 108) | 389 (436, 1222) | 40 (49, 216) |
| Definition ($n = 48$) | 26.25 (50.5, 465.75) | 2 (8.5, 58.25) | NA | 6 (9.5, 33.25) | 330.25 (569, 1066.75) | 32.25 (49, 73) |
| Signs/Symptoms ($n = 62$) | 27 (53, 312) | 2 (10, 42.25) | NA | 7 (12, 37.25) | 296 (515, 903.75) | 31.75 (44, 64) |
| Test ($n = 18$) | 31.5 (48.5, 122.5) | 1.75 (6, 18.5) | NA | 9 (17, 37.25) | 198 (407.5, 630.25) | 33.5 (46, 92.75) |
| Treatment/management ($n = 71$) | 26 (62, 169) | 3 (9, 25) | NA | 6 (13, 31) | 297 (533, 997) | 35 (48, 71) |
| Outcome ($n = 69$) | 26.5 (62, 173.5) | 2 (10, 32.5) | NA | 6 (11, 29.5) | 298 (497, 1049) | 35 (48, 62) |

Tables 2–4, respectively. Sources from TikTok ($n = 72$; 81.8%), Bilibili ($n = 28$; 54.9%), and Kwai ($n = 79$; 88.8%) were predominantly professional individuals. Among them, the videos uploaded by professional individuals on Bilibili have the highest likes with median (IQR) at 16.5 (4.5–35). However, Bilibili's contributions included a notable share from nonprofessional institutions ($n = 11$; 21.6%) and individuals ($n = 10$; 19.6%), surpassing the other platforms in this regard. Videos uploaded by nonprofessional individuals have the highest comments, shares and favorites, at 2.5 (0, 25), 4 (1, 70) and 9 (1, 75), respectively. Content-wise, TikTok ($n = 77$; 88%) and Kwai ($n = 77$; 80%) predominantly highlighted cryptorchidism treatment, whereas Bilibili placed a greater emphasis on the symptoms of cryptorchidism ($n = 40$; 78%). The days since the publication of treatment-related videos in TikTok is shorter with median (IQR) at 142.5 (268, 466), while in those in Kwai has the most comments with median (IQR) at 3 (9, 25).

## Assessment of interrater reliability

For the six content-related items, the kappa coefficients varied from .871 to .903 ($P < .001$) for the GQS, modified DISCERN, and JAMA scores, respectively. The range was .790 to .912 ($P < .001$), .860 to .918 ($P < .001$), and .847 to .961 ($P < .001$), indicating a high level of interrater agreement for each assessment item.

## Assessment of video content

The content scores for the six assessment components are summarized in Table 5. Figure 2(a) illustrates the scores for video content between platforms. Figure 2(b) to (f) demonstrate the scores of video uploaders on content of videos across different classifications. On average, the mean content score for each section was less than 1.5, indicating a moderate level of content coverage. Most of the videos described cryptorchidism symptoms and treatments, but the level of comprehensiveness of the content varied. Specifically, 60.2% and 75.0% of TikTok videos contained content on the definition and symptoms of cryptorchidism, respectively, while 64.7% and 78.4% of Bilibili's content dealt with the definition and symptoms of cryptorchidism, respectively. About 53.9% and 69.7% of Kwai's cases described the definition and symptoms of cryptorchidism, respectively. In terms of treatment and outcomes, 87.5% and 72.7% of TikTok videos featured these aspects, while

**Table 5.** Completeness of video content from TikTok, Kwai and Bilibili.

| Video Content | Risk Factors (*n* (%)) | Definition (*n* (%)) | Symptoms (*n* (%)) | Test (*n* (%)) | Treatment (*n* (%)) | Outcomes (*n* (%)) |
|---|---|---|---|---|---|---|
| **TikTok** | | | | | | |
| No content (0 points) | 83 (94.3%) | 35 (39.8%) | 22 (25%) | 52 (59.1%) | 11 (12.5%) | 24 (27.3%) |
| Little content (0.5 points) | 1 (1.1%) | 6 (6.8%) | 10 (11.4%) | 3 (3.4%) | 6 (6.8%) | 6 (6.8%) |
| Some content (1 point) | 1 (1.1%) | 24 (27.3%) | 33 (37.5%) | 21 (23.9%) | 27 (30.7%) | 20 (22.7%) |
| Most content (1.5 points) | 1 (1.1%) | 13 (14.8%) | 9 (10.2%) | 4 (4.5%) | 13 (14.8%) | 10 (11.4%) |
| Extensive content (2 points) | 2 (2.3%) | 10 (11.4%) | 14 (15.9%) | 8 (9.1%) | 31 (35.2%) | 28 (31.8%) |
| **Bilibili** | | | | | | |
| No content (0 points) | 43 (84.3%) | 18 (35.3%) | 11 (21.6%) | 29 (56.9%) | 18 (35.3%) | 15 (29.4%) |
| Little content (0.5 points) | 0 (0%) | 3 (5.9%) | 4 (7.8%) | 1 (2%) | 2 (3.9%) | 6 (11.8%) |
| Some content (1 point) | 2 (3.9%) | 12 (23.5%) | 12 (23.5%) | 14 (27.5%) | 12 (23.5%) | 7 (13.7%) |
| Most content (1.5 points) | 0 (0%) | 9 (17.6%) | 8 (15.7%) | 3 (5.9%) | 9 (17.6%) | 5 (9.8%) |
| Extensive content (2 points) | 6 (11.8%) | 9 (17.6%) | 16 (31.4%) | 4 (7.8%) | 10 (19.6%) | 18 (35.3%) |
| **Kwai** | | | | | | |
| No content (0 points) | 82 (92.1%) | 41 (46.1%) | 27 (30.3%) | 71 (79.8%) | 18 (20.2%) | 20 (22.5%) |
| Little content (0.5 points) | 1 (1.1%) | 4 (4.5%) | 1 (1.1%) | 2 (2.2%) | 12 (13.5%) | 17 (19.1%) |
| Some content (1 point) | 4 (4.5%) | 38 (42.7%) | 28 (31.5%) | 11 (12.4%) | 34 (38.2%) | 31 (34.8%) |
| Most content (1.5 points) | 0 (0%) | 4 (4.5%) | 17 (19.1%) | 2 (2.2%) | 14 (15.7%) | 9 (10.1%) |
| Extensive content (2 points) | 2 (2.2%) | 2 (2.2%) | 16 (18%) | 3 (3.4%) | 11 (12.4%) | 12 (13.5%) |

Bilibili and Kwai featured 64.7% and 70.6%, and 79.8% and 77.5%, respectively. Notably, content on risk factors and screening for cryptorchidism was largely absent across the three platforms. In TikTok videos, risk factors and screening content were missing in 94.3% and 59.1% of cases, respectively, compared to 84.3% and 56.9% in Bilibili and 92.1% and 79.8% in Kwai.

## Analysis of video quality and reliability

*Comparison of platforms.* As indicated in Table 1 and Figure 3(a) to (c), the evaluation of video reliability revealed that TikTok videos outperformed Bilibili videos in terms of modified DISCERN scores ($P = .003$), surpassed Kwai in GQS ($P = .002$), and exceeded Bilibili in JAMA scores ($P < .001$). Additionally, Kwai attained higher JAMA scores compared to those from Bilibili ($P < .001$).

*Comparison of uploaders.* As delineated in Table 6 and Figure 3(d) to (l), among the professional individuals, specialized physicians accounted for the highest percentage ($n = 131$, 57.5%), and among the nonprofessional individuals, the percentage of science bloggers ($n = 12$, 5.3%) was higher than the percentage of patients ($n = 11$, 4.8%). Also, science bloggers had the highest median GQS score of 4 (3, 5) and the highest median modified DISCERN score of 4 (3, 4).

The JAMA scores for videos uploaded by medical professionals were notably higher than those by nonmedical professionals ($P < .001$). No significant disparities were observed in the GQS and modified DISCERN scores when comparing videos from medical and nonmedical professionals (($P = .63$) and ($P = .85$), respectively. Within the healthcare professional category, videos from specialized physicians received higher GQS scores compared to those

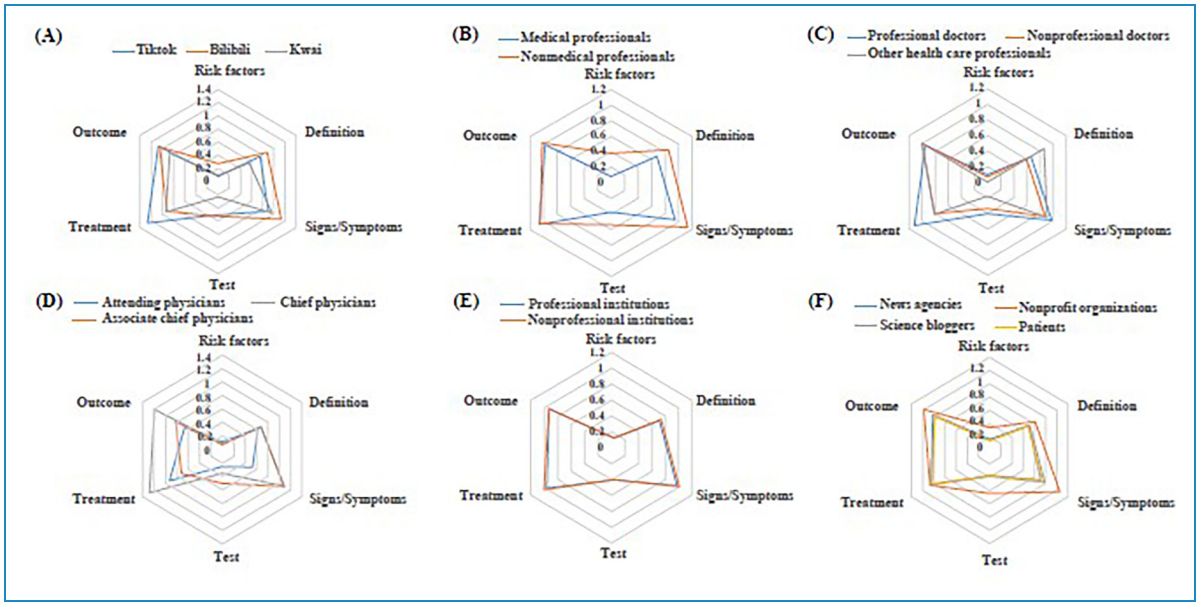


**Figure 2.** Radar chart of health information for short videos.

from non-specialized physicians ($P < .001$) and other healthcare professionals ($P < .001$). Moreover, professional doctors' videos had elevated modified DISCERN scores relative to nonprofessional doctors ($P = .04$). The JAMA scores did not significantly differ among the three subtypes of healthcare professionals ($P = .22$). Regarding the nonmedical professional subtypes, videos from patients had lower GQS scores than those from science bloggers ($P < .001$) and nonprofit organizations ($P = .04$). Science bloggers' videos boasted higher modified DISCERN scores compared to those from patients ($P < .001$) and news agencies ($P = .04$), and videos from nonprofit organizations had more favorable modified DISCERN scores than those from patients ($P = .03$). Additionally, patient-uploaded videos had inferior JAMA scores in contrast to those from news agencies ($P = .02$), nonprofit organizations ($P = .049$), and science bloggers ($P = .004$).

### *Correlation analysis and Poisson regression analysis*

In the Pearson correlation test (Table 7), it was found that there was a significant negative correlation between the GQS and the number of comments ($r = -.16$, $P = .01$), shares ($r = -.18$, $P = .04$) and the days since published ($r = -.19$, $P = .005$), a statistically significant positive correlation ($r = .34$, $P < .001$) with the video duration and no statistically significant correlation with the number of favorites ($P = .70$) and likes ($P = .18$). The modified DISCERN had a significant positive correlation with video duration ($r = .39$, $P < .001$), and no statistically significant correlation with the number of likes ($P = .86$), comments ($P = .54$), shares ($P = .82$), favorites ($P = .20$), and days since published ($P = .12$). JAMA scores had a significant negative correlation with days since published ($r = -.25$, $P < .001$), and no statistically significant correlation with the number of likes ($P = .93$), comments ($P = .52$), shares ($P = .97$), favorites ($P = .36$), and days since published ($P = .55$).

The basic information of the videos was non-normally distributed, so we used Spearman correlation analysis (Table 8) to investigate the relationship between them. We found a positive correlation between the following characters: likes-comments ($r = .84$, $P < .001$), likes-shares ($r = .89$, $P < .001$), likes-favorites ($r = .87$, $P < .001$), comments-shares ($r = .85$, $P < .001$), comments-favorites ($r = .75$, $P < .001$), shares-favorites ($r = .92$, $P < .001$).

## Discussion

This comprehensive cross-sectional study meticulously evaluated the content of 228 short videos from three widely used short-video sharing platforms in China (TikTok, Bilibili, and Kwai) to scrutinize the health-related messages and ascertain the reliability of short videos concerning cryptorchidism. The analysis, depicted in a radar chart, focused on six core contents within the videos, revealing a predominant emphasis on symptoms and treatments of cryptorchidism, while discussions on risk factors and test were notably scarce. The GQS, modified DISCERN, and JAMA scores served as metrics to assess the videos' completeness and reliability. Collectively, the health information presented was deemed moderately reliable.

The study found that videos from Bilibili were less reliable and complete, potentially due to a higher proportion of uploads by nonmedical professionals. In contrast, TikTok videos demonstrated better reliability and completeness. The uploaders' background significantly influenced the

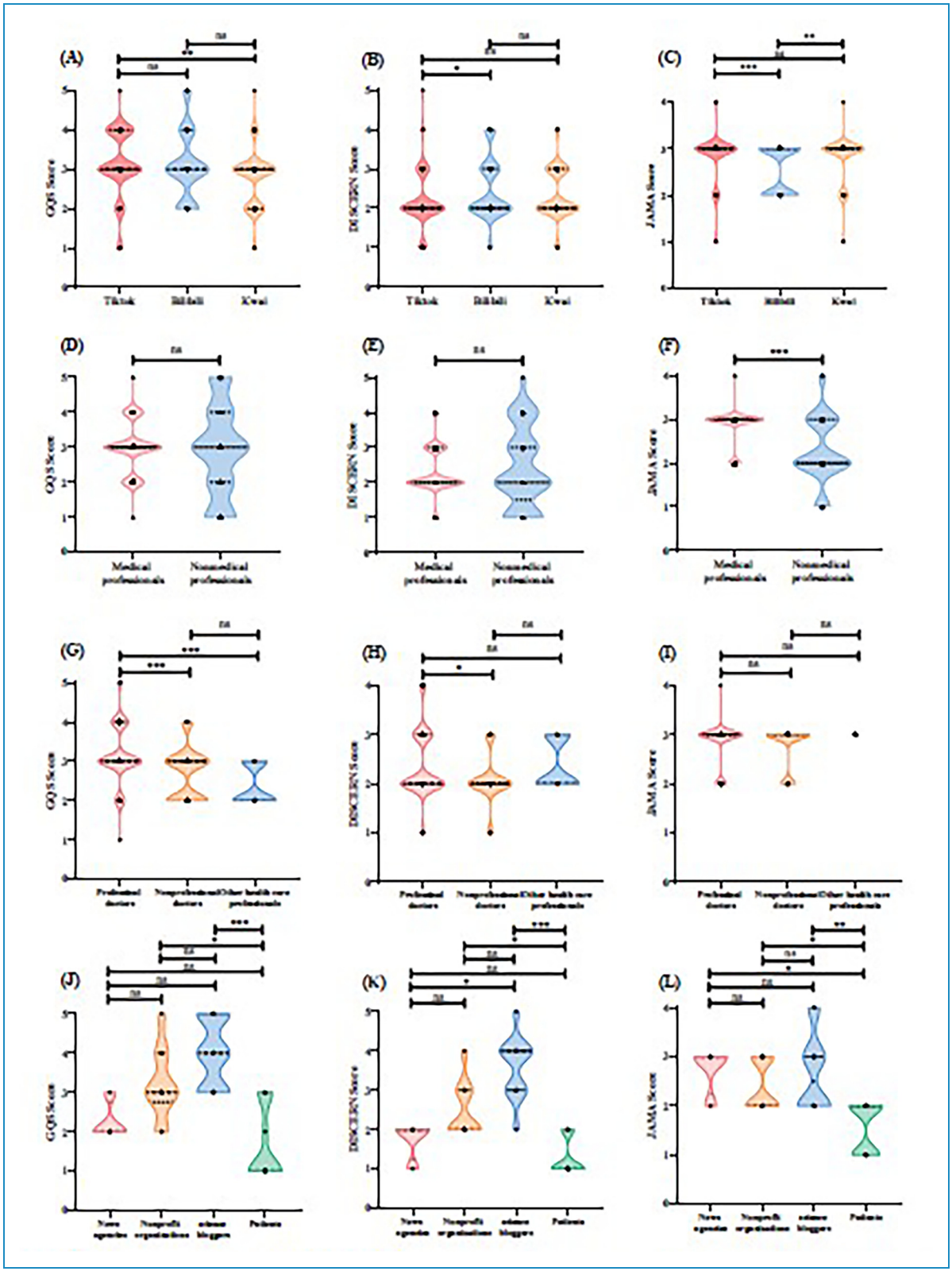

**Figure 3.** Violin diagram of health information on different short video platforms.

**Table 6.** Analysis of videos from diverse types of uploaders

| Subtypes (*n*) (%) | GQS Score (median (IQR)) | Modified DISCERN Score (median (IQR)) | JAMA Score (median (IQR)) |
|---|---|---|---|
| Professional individuals (*n* = 179) (78.5%) | 3 (3, 3) | 2 (2, 3) | 3 (3, 3) |
| Professional doctors (*n* = 131) (57.5%) | 3 (3, 3) | 2 (2, 3) | 3 (3, 3) |
| Nonprofessional doctors (*n* = 37) (16.2%) | 3 (2, 3) | 2 (2, 2) | 3 (2.5, 3) |
| Other health care professionals (*n* = 11) (4.8%) | 2 (2, 3) | 2 (2, 3) | 3 (3, 3) |
| Professional institutions (*n* = 12) (5.3%) | 3 (3, 4) | 2 (2, 3) | 3 (3, 3) |
| Nonprofessional individuals (*n* = 23) (10.1%) | 3 (1, 4) | 2 (1, 4) | 2 (2, 3) |
| News agencies (*n* = 4) (1.8%) | 2 (2, 2.75) | 2 (1.25, 2) | 3 (2.25, 3) |
| Nonprofit organizations (*n* = 10) (4.4%) | 3 (2.75, 4) | 2 (2, 3) | 2 (2, 3) |
| Science bloggers (*n* = 12) (5.3%) | 4 (3, 5) | 4 (3, 4) | 2.5 (2, 3) |
| Patients (*n* = 11) (4.8%) | 1 (1, 3) | 1 (1, 2) | 2 (1, 2) |
| Attending physicians (*n* = 41) | 3 (2, 3) | 2 (2, 2) | 3 (2, 3) |
| Associate chief physician (*n* = 52) | 3 (2, 3) | 2 (2, 3) | 3 (3, 3) |
| Chief physician (*n* = 86) | 3 (3, 4) | 2 (2, 3) | 3 (3, 3) |

videos' reliability; professional individuals, especially professional doctors, provided more reliable and comprehensive health information on cryptorchidism. Conversely, videos uploaded by patients scored lower on GQS, modified DISCERN, and JAMA, indicating a need for stricter regulation of health information dissemination by patients.

TikTok has a wider user group and a variety of content styles; Bilibili has young people as its main user group and its content is more inclined to entertainment; and Kwai focuses more on life situations and original content. This leads to differences in the content and style of health science videos on different platforms, thus affecting the reliability of the information.

The auditing rules for medical content on TikTok and Kwai are relatively strict, allowing only certified users to post medical science content, while noncertified users can only post health science content. In addition, there may be differences in the auditing methods and strength of different platforms, which may also affect the reliability of information in health science videos.[22]

Correlation analysis showed a positive correlation between video duration and higher GQS and modified DISCERN scores, suggesting that longer videos tend to be more reliable and complete. There was a negative correlation between the days since the video was uploaded and the quality score, implying that newer videos are more reliable in quality. Interestingly, less complete videos attracted more comments and shares, suggesting that the public tends to be exposed to potentially misleading health information. This emphasizes the importance of viewers critically evaluating health information in short videos.

The absence of rigorous review processes for video content has led to variability in the quality and reliability of short videos on platforms.[23] In contrast, longer videos tend to provide more comprehensive and valuable information,[24] and newer videos often reflect higher scientific quality.[25] These conclusions are also consistent with our findings. Many users of short-video platforms struggle to discern high-quality videos due to factors such as publicity and education level, resulting in popularity metrics such as likes, comments, shares, and favorites being poorly or negatively correlated with video quality and reliability.[23] This aligns with previous research[23,26] and underscores the need for public education on health and disease prevention, as misinformation can have detrimental social impacts.[27]

We have found that there is a positive correlation among the likes, comments, collections, and shares of short videos. There are statistical associations among the basic characteristics of the videos. This indicates that residents are willing to share and learn health knowledge from short videos related to cryptorchidism. This reflects that health short

**Table 7.** Pearson correlation analysis.

| Variable | Global Quality Score | Modified DISCERN Score | JAMA Score |
|---|---|---|---|
| Likes | | | |
| *r* | −0.09 | 0.01 | 0.01 |
| *P* value | .18 | .86 | .93 |
| Comments | | | |
| *r* | −0.16 | −0.04 | −0.04 |
| *P* value | .01[a] | .54 | .52 |
| Shares | | | |
| *r* | −0.18 | −0.02 | 0.003 |
| *P* value | .04[a] | .82 | .97 |
| Favorite | | | |
| *r* | 0.03 | 0.09 | 0.06 |
| *P* value | .70 | .20 | .36 |
| Days since published | | | |
| *r* | −0.19 | −0.10 | −0.25 |
| *P* value | .005[a] | .12 | <.001[a] |
| Duration | | | |
| *r* | 0.34 | 0.39 | −0.04 |
| *P* value | <.001[a] | <.001[a] | .55 |

aSignificant at $P < .05$.

videos related to cryptorchidism have good transmissibility in the residential community. Longer videos are capable of explaining health knowledge more comprehensively and in-depth.[24] They have more time to elucidate complex concepts, provide cases and empirical evidence, thereby helping the audience to better understand and master relevant content. In addition, some videos might require a relatively long time to display inspection method, operational steps or conduct interactive Q&A sessions to ensure that the audience can learn and apply the acquired knowledge correctly.

Professional doctors' videos showed higher quality and reliability, highlighting the importance of professional knowledge in healthcare videos. Platforms such as TikTok, Bilibili, and Kwai have certification processes for professional doctors. It ensures a more rigorous review of their content. Conversely, videos posted by patients often lack professionalism, focusing more on personal experiences and feelings, which diminishes their quality and reliability.

The general quality and reliability of short videos are not aimed at medical professionals, necessitating publishers to present information in an engaging manner, which may compromise professionalism.[28] This highlights the need for awareness about the credibility of short videos, and residents should be cautious when selecting information from these platforms.[29]

With the advancement of the Internet technology and increased health awareness among the population, Internet-based visual information dissemination is becoming more common. As a result, residents are changing from passive recipients to active knowledge seekers.[30] The rise of the multimedia technology and the widespread use of electronic devices such as smartphones have made visual social media an important source of patient information. Social media platforms facilitate access to healthcare information. However, the varying quality of videos poses many problems, and some videos have inaccurate content that can mislead viewers. An effective health promotion video should be scientifically accurate, reliable, and understandable. Therefore, it is crucial to assess the quality of the videos to enable viewers to get the message from these short videos more effectively. Such assessment also empowers professionals, raises public health awareness, and promotes the health care system. These findings highlight the necessity for more stringent regulation and quality control on such platforms. In the future, community health education should be further promoted so that residents have better access to comprehensive and accurate health information. Further research will be conducted to promote awareness of cryptorchidism and other diseases among the population and to build a social and public health system.

Our study analyzed three major short-video platforms in China, using the GQS, modified DISCERN score, and JAMA score to evaluate video quality, comprehensiveness, and reliability, and calculated Cohen's κ to ensure rater consistency. This study is the largest study of short videos to date and used several metrics to comprehensively assess the health messages and quality of short videos. However, there were several limitations in this study that merit attention. This study focuses on the top 3 popular Chinese short video platforms, which may not be representative of content in other languages. Despite using various quality indicators, controlling for confounders was challenging, and subjective differences in ratings between researchers could not be

**Table 8.** Spearman correlation analysis between the video variables.

| Variables | Likes | Comments | Shares | Favorites | Days Since Published | Duration |
|---|---|---|---|---|---|---|
| Likes | | | | | | |
| *r* | 1 | - | - | - | - | - |
| *P* value | - | - | - | - | - | - |
| Comments | | | | | | |
| *r* | 0.84 | 1 | - | - | - | - |
| *P* value | <.001[a] | - | - | - | - | - |
| Shares | | | | | | |
| *r* | 0.89 | 0.85 | 1 | - | - | - |
| *P* value | <.001[a] | <.001[a] | - | - | - | - |
| Favorites | | | | | | |
| *r* | 0.87 | 0.75 | 0.92 | 1 | - | - |
| *P* value | <.001[a] | <.001[a] | <.001[a] | - | - | - |
| Days since published | | | | | | |
| *r* | 0.08 | 0.12 | 0.14 | 0.02 | 1 | - |
| *P* value | .23 | .06 | .10 | .72 | - | - |
| Duration | | | | | | |
| *r* | −0.03 | 0.01 | 0.07 | 0.07 | −0.003 | 1 |
| *P* value | .64 | .90 | .39 | .32 | .96 | - |

[a]Significant at $P < .05$.

entirely eliminated. The rating scales were limited to textual content and did not account for images and audio. The small sample size for certain uploader subtypes (e.g., nonprofessional institutions) cannot thoroughly represent the generalizability across different demographics or regions, which adds to the study's uncertainty and potential errors.

## Conclusion

The study's examination of 228 videos related to cryptorchidism from three prominent short-form video-sharing platforms in China (TikTok, Bilibili, and Kwai) revealed that the overall quality and reliability of these videos were moderate. It was observed that content produced by medical professionals generally surpassed that of nonmedical professionals in terms of depth, quality, and reliability. Specifically, videos disseminating disease knowledge by specialized physicians were identified as being of superior quality. Consequently, individuals should approach health management information obtained from short-form video platforms with a degree of skepticism and prioritize verifying such information through credible sources.

**Contributorship:** Y.S. and X.L. contributed equally to this work. Conceptualization: A.L. Data curation: Y.S. and X.L. Formal analysis: Y.S. and X.L. Project administration: Y.S. and X.L. Supervision: A.L. Writing-original draft: Y.S. and X.L. Manuscript revision and language polishing: Y.S., X.L., Q.X., and X.Z. All authors read and approved the final manuscript.

**Data availability:** The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

**Declaration of conflicting interests:** The authors declared no potential conflicts of interests with respect to the research, authorship, and/or publication of this article.

**Ethical approval:** The study adhered to ethical guidelines by exclusively utilizing information from publicly available videos on TikTok, Bilibili, and Kwai, ensuring that no personal privacy was compromised. The absence of clinical data, human specimens, or laboratory animals further negates the need for ethical review. Moreover, the lack of direct user interaction within the study's framework reinforces its compliance with ethical standards. This approach aligns with the ethical conduct of research, emphasizing the importance of privacy and the noninvolvement of sensitive biological materials or direct human interaction.

**Funding:** The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by the Special Fund for Pediatric Development of Qilu Hospital, Shandong University (34641390220001).

**Informed consent:** No clinical data, human specimens, or laboratory animals were used in this study. All information was obtained from publicly released TikTok and BiliBili videos, and none of the data involved personal privacy concerns. In addition, this study did not involve any interaction with users; therefore, no ethics review was needed.

**ORCID iD:** Xingjian Liu https://orcid.org/0009-0008-9395-7034
Aiwu Li https://orcid.org/0000-0002-1069-2026

**Supplemental material:** Supplemental material for this article is available online.

## References

1. Koch T, Hansen AH, Priskorn L, et al. A history of cryptorchidism is associated with impaired testicular function in early adulthood: a cross-sectional study of 6376 men from the general population. *Hum Reprod* 2020; 35: 1765–1780.
2. Gurney JK, McGlynn KA, Stanley J, et al. Risk factors for cryptorchidism. *Nat Rev Urol* 2017; 14: 534–548.
3. Rodprasert W, Virtanen HE and Toppari J. Cryptorchidism and puberty. *Front Endocrinol (Lausanne)* 2024; 15: 1347435.
4. Virtanen HE and Toppari J. Cryptorchidism and fertility. *Endocrinol Metab Clin North Am* 2015; 44: 751–760.
5. Ciongradi CI, Sârbu I, Iliescu Halițchi CO, et al. Fertility of cryptorchid testis-an unsolved mistery. *Genes (Basel)* 2021; 12: 1894.
6. Zhao TX, Liu B, Wei YX, et al. Clinical and socioeconomic factors associated with delayed orchidopexy in cryptorchid boys in China: a retrospective study of 2423 cases. *Asian J Androl* 2019; 21: 304–308.
7. Dee EC, Muralidhar V, Butler SS, et al. General and health-related internet use among cancer survivors in the United States: a 2013–2018 cross-sectional analysis. *J Natl Compr Canc Netw* 2020; 18: 1468–1475.
8. Ferhatoglu MF, Kartal A, Ekici U, et al. Evaluation of the reliability, utility, and quality of the information in sleeve gastrectomy videos shared on open access video sharing platform YouTube. *Obes Surg* 2019; 29: 1477–1484.
9. Basch CH, Hillyer GC and Jaime C. COVID-19 on TikTok: harnessing an emerging social media platform to convey important public health messages. *Int J Adolesc Med Health* 2020; 34: 367–369.
10. van Uden-Kraan CF, Drossaert CH, Taal E, et al. Health-related internet use by patients with somatic diseases: frequency of use and characteristics of users. *Inform Health Soc Care* 2009; 34: 18–29.
11. Bailey MA, Coughlin PA, Sohrabi S, et al. Quality and readability of online patient information for abdominal aortic aneurysms. *J Vasc Surg* 2012; 56: 21–26.
12. Bai G, Fu K, Fu W, et al. Quality of internet videos related to pediatric urology in mainland China: a cross-sectional study. *Front Public Health* 2022; 10: 924748.
13. Li HO, Bailey A, Huynh D, et al. YouTube as a source of information on COVID-19: a pandemic of misinformation? *BMJ Glob Health* 2020; 5: e002604.
14. Keten T and Erkan A. An investigation of the reliability of YouTube videos on undescended testis. *J Pediatr Urol* 2022; 18: 515.e1–515.e6.
15. Eghtesadi M and Florea A. Facebook, Instagram, Reddit and TikTok: a proposal for health authorities to integrate popular social media platforms in contingency planning amid a global pandemic outbreak. *Can J Public Health* 2020; 111: 389–391.
16. Ding R, Kong Q, Sun L, et al. Health information in short videos about metabolic dysfunction-associated steatotic liver disease: analysing quality and reliability. *Liver Int* 2024; 44(6): 1373–1382. DOI:10.1111/liv.15871.
17. Goobie GC, Guler SA, Johannson KA, et al. YouTube videos as a source of misinformation on idiopathic pulmonary fibrosis. *Ann Am Thorac Soc* 2019; 16: 572–579.
18. Subramanian T, Araghi K, Amen TB, et al. Chat generative pretraining transformer answers patient-focused questions in cervical spine surgery. *Clin Spine Surg* 2024 Jul 1;37(6): E278-E281; 37(6): E278–E281. DOI: 10.1097/BSD.0000000000001600.
19. 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: 1244–1245.
20. Kyarunts M, Mansukhani MP, Loukianova LL, et al. Assessing the quality of publicly available videos on MDMA-assisted psychotherapy for PTSD. *Am J Addict* 2022; 31: 502–507.
21. 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: 105–111.

22. Liu Z, Chen Y, Lin Y, et al. YouTube/bilibili/TikTok videos as sources of medical information on laryngeal carcinoma: cross-sectional content analysis study. *BMC Public Health* 2024; 24: 1594.
23. Sun F, Zheng S and Wu J. Quality of information in gallstone disease videos on TikTok: cross-sectional study. *J Med Internet Res* 2023; 25: e39162.
24. Chen Z, Pan S and Zuo S. Tiktok and YouTube as sources of information on anal fissure: a comparative analysis. *Front Public Health* 2022; 10: 1000338.
25. Barlas T, Ecem Avci D, Cinici B, et al. The quality and reliability analysis of YouTube videos about insulin resistance. *Int J Med Inform* 2023; 170: 104960.
26. Mueller SM, Hongler VNS, Jungo P, et al. Fiction, falsehoods, and few facts: cross-sectional study on the content-related quality of atopic eczema-related videos on YouTube. *J Med Internet Res* 2020; 22: e15599.
27. Chan S, Parish M and Yellowlees P. Telepsychiatry today. *Curr Psychiatry Rep* 2015; 17: 89.
28. Song S, Xue X, Zhao YC, et al. Short-Video apps as a health information source for chronic obstructive pulmonary disease: information quality assessment of TikTok videos. *J Med Internet Res* 2021; 23: e28318.
29. Gijsen V, Maddux M, Lavertu A, et al. Science: the potential and the challenges of utilizing social media and other electronic communication platforms in health care. *Clin Transl Sci* 2020; 13: 26–30.
30. Kong W, Song S, Zhao YC, et al. Tiktok as a health information source: assessment of the quality of information in diabetes-related videos. *J Med Internet Res* 2021; 23: e30409.