

The quality and reliability of Chinese short-video information on spinal robotic navigation surgery on TikTok and Bilibili: a cross-sectional content analysis

Qidi Geng

Air Force Medical University, Air Force Specialty Medical Center

Mingzhe Li

Air Force Medical University, Air Force Specialty Medical Center

Yaxuan Su

Air Force Medical University, Air Force Specialty Medical Center

Wei Guo

2003gw@163.com

Air Force Medical University, Air Force Specialty Medical Center

Research Article

Keywords: spinal navigation robot, short video, social media, health information quality, GQS, modified DISCERN

Posted Date: September 30th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10984080/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

The quality and reliability of Chinese short-video information on spinal robotic navigation surgery on TikTok and Bilibili: a cross-sectional content analysis

QidiGeng¹, MingzheLi¹, YaxuanSu¹, WeiGuo¹

1. Department of Orthopedics of Integrated Traditional Chinese and Western Medicine, Air Force Medical University, Air Force Specialty Medical Center; Beijing 100142

corresponding author: WeiGuo, E-mail: 2003gw@163.com

Abstract

Objective This cross-sectional content-analysis study aimed to systematically evaluate the quality and reliability of Chinese-language short-videos concerning spinal navigation robots on TikTok and Bilibili, and identify predictors of video information quality.

Methods Short-videos were retrieved from TikTok and Bilibili using predefined search terms on July 19, 2026. Videos were screened according to pre-established inclusion and exclusion criteria. Two blinded attending physicians scored eligible videos using the Global Quality Score (GQS) and modified DISCERN (mDISCERN) scale. Uploader type, content theme, video metadata and engagement metrics were coded. Inter-rater reliability was assessed by Cohen's κ . Non-parametric tests, Spearman's rank-order correlation and GEE-based robust Poisson regression were applied for statistical analysis.

Results A total of 52 eligible short-videos were finally enrolled, including 38 from TikTok and 14 from Bilibili. Inter-rater reliability was substantial for both GQS ($\kappa = 0.708$) and mDISCERN ($\kappa = 0.741$). Significant inter-platform differences existed in video duration, engagement indicators, uploader composition and content themes. Videos produced by clinicians achieved significantly higher GQS and mDISCERN scores. Benefit-risk analysis-oriented videos obtained the highest GQS, whereas pure technical-device-oriented videos showed poorer quality. Likes and comments were positively correlated with GQS, while only comment count was positively associated with mDISCERN. Multivariable GEE robust Poisson regression demonstrated that content type was an independent predictor for both quality scales.

Technology-equipment-focused content independently predicted lower GQS (adjusted IRR = 0.614, 95%CI 0.442-0.853, $P = 0.004$). Clinician-released videos were independently associated with higher mDISCERN scores compared with science-popularization bloggers (adjusted IRR = 1.713, 95%CI 1.014-2.892, $P = 0.044$). After adjustment, platform origin exerted no independent effect on quality outcomes.

Conclusions The overall quality and reliability of spinal-navigation-robot-related short-videos remain unsatisfactory on the two platforms. Uploader background and content theme are more critical determinants of information quality than platform source or user-engagement performance. High-interaction metrics cannot represent medically balanced and trustworthy health information. More clinician-participated

popular-science production, stricter creator-qualification audit and better exposure for balanced benefit-risk-analysis content are required to optimise online patient educational resources for robotic spinal surgery.

Keywords: spinal navigation robot; short video; social media; health information quality; GQS; modified DISCERN

1. Introduction

Traditional spinal surgery is limited by inaccurate screw placement, high-level radiation exposure, and elevated risks for complex scoliosis or revision operations^[1-2]. Spinal navigation robots provide three-dimensional reconstruction, precise screw insertion, reduced neurological injury risk, lower intraoperative radiation dosage, and shortened rehabilitation duration^[3], and such robotic systems have been increasingly adopted by hospitals across China with growing annual surgical volumes^[4].

A large proportion of patients actively search for preoperative information regarding robotic spinal surgery online, and short-video platforms have become their preferred information source^[5]. Spinal surgery carries high-risk characteristics^[6]; exaggerated narration, omitted contraindications, or concealed complications in popular-science videos may mislead patients' surgical decision-making, trigger distorted medical expectations, doctor-patient conflicts, and even severe intraoperative adverse events such as paralysis or screw malposition. TikTok and Bilibili have experienced explosive growth in user volume and health-popularization content^[7]. Short-video media deliver medical knowledge in an audiovisual, low-threshold manner that facilitates public understanding^[8]. Nevertheless, the two platforms differ substantially in ecological features: TikTok favours short-duration, algorithm-driven content prioritising traffic and emotional visual stimulation for younger audiences, resulting in fragmented popular-science outputs^[9]. By contrast, Bilibili features longer-form videos and a knowledge-oriented community, where audiences are receptive to in-depth, systematic medical explanations^[10]. However, short-video platforms suffer from inherent limitations including loose regulatory oversight and non-mandatory verification of creator qualifications. Recommendation algorithms tend to prioritise attention-grabbing content, which facilitates over-promotional and one-sided health communication^[11]. Multiple existing studies on chronic disease and joint-surgery-related short-video content have documented widespread information distortion across these platforms^[12].

Overseas studies mainly focus on YouTube and international-version TikTok, exploring short-video content for joint arthroplasty and chronic diseases, with only a small number addressing spinal-related social-media content^[13]. Few studies have performed dual-platform Chinese-language evaluations specifically for spinal navigation robots. Prior orthopaedic short-video investigations in China have predominantly evaluated knee-joint robotic content; no cross-sectional dual-platform analysis targeting spinal-navigation-robot-related short-form videos has been published. Given the high-risk and high-precision nature of spinal robotic

surgery, systematic quantitative assessments are lacking. Previous orthopaedic-video research failed to stratify surgical risk levels between spinal and joint procedures, thus ignoring potential clinical harms caused by misleading online popular-science materials^[14]. Evidence-based optimisation strategies for spinal-surgery-related social-media popular-science remain absent, limiting data-driven guidance for clinicians and platform regulators.

Against these research gaps, this cross-sectional content-analysis study systematically evaluated the quality of spinal-navigation-robot-related popular-science short-videos released on two major Chinese short-video platforms. Findings from this work aim to support standardised online health communication for spinal surgery and mitigate cognitive bias among patients.

2. Materials and Methods

2.1 Study design

This study adopted a cross-sectional content-analysis design. Chinese-language short-videos about spinal navigation robots from TikTok and Bilibili were enrolled. Video retrieval and sample collection were completed on July 19, 2026 to minimise confounding effects caused by platform-algorithm iteration and dynamic hot-content updates. Only publicly accessible metadata, visual-audio content and interaction metrics were extracted. No clinical medical records or personal-privacy-related data were involved; therefore, ethical-committee approval was waived.

2.2 Video retrieval, screening and sample collection

2.2.1 Platforms and search terms

Search platforms: domestic Chinese-language TikTok and Bilibili. Search keywords: spinal navigation robot OR minimally-invasive spinal robotic surgery. A brand-new browsing-history-free account was used to eliminate personalised-recommendation bias. Videos were retrieved under the comprehensive-ranking list weighted by completion rate, likes, comments and publication time.

2.2.2 Inclusion criteria

1. Videos presented in Chinese (with Chinese subtitles allowed);
2. Core themes covering spinal-navigation-robot devices, surgical procedures, indications, surgical benefits-risks, and relevant popular-science narratives;
3. Complete publicly available metadata including title, uploader information and interaction indicators (likes, comments, collections, shares);
4. Video duration ≥ 10 s with audible audio tracks.

2.2.3 Exclusion criteria

1. Pure foreign-language videos without Chinese subtitles;
2. Reproduced or second-edited videos lacking original content;
3. Pure commercial advertisements without objective popular-science components;
4. Content irrelevant to spinal navigation robots (general orthopaedics, joint robots, generic rehabilitation, etc.);
5. Defective videos with missing audio, black-screen frames or incomplete

evaluable information.

2.2.4 Sampling and screening workflow

The top 100 comprehensively-ranked videos were extracted separately from each platform. Each entry was manually screened against inclusion-exclusion criteria, and reasons for exclusion were documented for constructing the PRISMA-style flow diagram (Figure 1). Qualified videos from both platforms were aggregated as the final analytical sample, and all variables were imported into Excel for building a standardised database.

2.3 Standardised coding framework

All eligible videos were manually coded across three dimensions: basic video characteristics, uploader categories, and content themes. Uploader-type and content-theme classifications were consistent with the categories used in subsequent statistical analyses.

2.3.1 Basic video-characteristic variables

Platform source, video duration (seconds), days elapsed since publication, number of likes, comments, collections, and shares.

2.3.2 Four-category uploader classification

1. Clinician: Board-certified orthopaedic or spinal surgeons, clinical researchers specialising in spinal robotic surgery, and orthopaedic postgraduate trainees with clinical practice experience.

2. Institutional account: Official accounts of tertiary-hospital spinal surgery departments, orthopaedic research institutes, authorised orthopaedic medical associations, and official corporate accounts of spinal-navigation-robot manufacturers.

3. Commercial marketing account: Self-media agencies, rehabilitation clinics, medical-device sales organisations and other commercial-oriented accounts without independent clinical-physician authorship.

4. Science-popularization blogger: Non-medical individual content producers, including patients, popular-science enthusiasts and fitness-rehabilitation communicators without formal medical licensing.

2.3.3 Five-category content-theme classification

1. Technology&Equipment: Introduction and interpretation of robotic hardware, three-dimensional reconstruction, intraoperative registration, equipment principles and workflow of spinal navigation robots.

2. Intraoperative Surgical Footage: Intraoperative real-time surgical clips demonstrating robot-assisted pedicle screw placement, scoliosis correction and revision spinal surgery.

3. Indications&Contraindications: Eligible and ineligible clinical scenarios for robotic spinal surgery, comparisons between conventional open surgery and robot-assisted surgery, and key considerations for clinical decision-making.

4. Pros&Cons: Balanced discussion covering surgical advantages, limitations, intraoperative radiation exposure, economic cost, operative time, complications and surgeon learning-curve characteristics.

5. Promotional&Other: Hospital-or device-oriented promotional content with exaggerated descriptions for patient recruitment, as well as other videos that cannot be

assigned to the above four themes and contain limited evidence-based popular-science value.

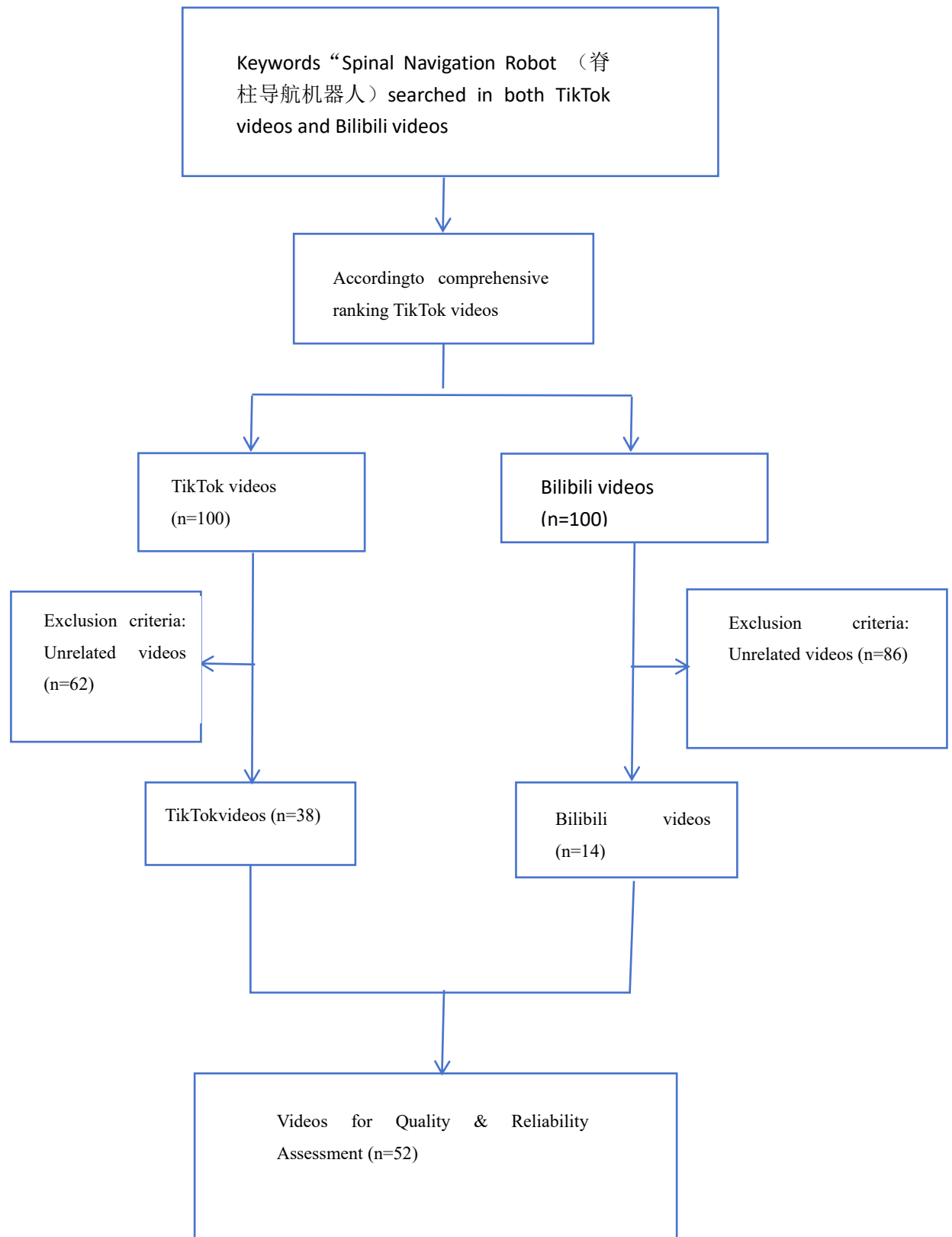


Fig. 1 Methodology for identifying short videos on Spinal Navigation Robot

2.4 Quality-and-reliability-assessment tools and scoring protocol

Two well-established standardised instruments widely adopted in orthopaedic-short-video research were applied:the Global Quality Score(GQS)^[15]and modified DISCERN(mDISCERN)^[16].

2.4.1 Description of assessment tools

1.GQS(Global Quality Score):scores ranging from 1 to 5(1 = very poor;2 = poor;3 = moderate;4 = good;5 = excellent).This scale evaluates overall popular-science quality regarding logical coherence,comprehensibility and practical value for lay-person audiences.

2.Modified DISCERN(mDISCERN):a short-video-adapted simplified scale with scores ranging from 1 to 5,assessing information reliability across five dimensions:source traceability,risk-disclosure completeness,presence of over-claiming,argument balance,and evidence-base sufficiency.

2.4.2 Standardised scoring procedure

1.Evaluators:two attending physicians with more than ten years of spinal-surgery clinical experience independently performed blinded assessments for all videos;platform information and traffic-related metrics were masked,and only video-audio content was reviewed.

2.Disagreement resolution:whenever total-score discrepancy ≥ 1 point existed for one video,a senior chief-physician reviewer conducted tertiary re-evaluation to achieve consensus final scores.

3.Inter-rater-reliability evaluation:Cohen's κ coefficient was interpreted according to Landis-Koch criteria: $\kappa > 0.80$ represented almost-perfect agreement;0.61–0.80 substantial agreement;0.41–0.60 moderate agreement; $\kappa \leq 0.40$ poor agreement^[17].

2.5 Statistical analysis

Non-normally-distributed continuous variables were summarised using median(interquartile range,IQR).Between-platform comparisons were performed with the Mann-Whitney U test.Group-level comparisons across uploader-type and content-theme subgroups were implemented via the Kruskal-Wallis H test followed by Dunn-Bonferroni post-hoc pairwise testing.Categorical-variable inter-group differences were assessed using Fisher's exact test.Spearman's rank-order correlation explored associations between engagement metrics,video duration,publication age and quality scores.Generalised estimating-equation(GEE)-based robust Poisson regression was performed to identify independent predictors of video-quality outcomes.Cohen's κ quantified inter-rater reliability.All statistical tests were two-tailed,with $\alpha = 0.05$ for significance threshold.SPSS 27.0 software was used for all statistical computations.

3 Results

3.1 Video inclusion and basic sample characteristics

Following the systematic screening procedure, a total of 52 short-videos were eligible for final analysis, including 38 videos from TikTok and 14 videos from Bilibili. Inter-rater reliability between the two independent raters was substantial: the

Cohen's κ value was 0.708 for the GQS scale and 0.741 for the mDISCERN scale.

Of the 52 included Chinese-language short-videos related to spinal navigation robots, 14 (26.9%) were from Bilibili and 38 (73.1%) were from TikTok. Regarding uploader types, clinician accounts constituted the largest subgroup ($n = 27$, 51.9%), followed by institutional accounts ($n = 10$, 19.2%), commercial marketing accounts ($n = 9$, 17.3%), and science-popularization bloggers ($n = 6$, 11.5%). For content themes, intraoperative surgical footage was the most frequent category ($n = 15$, 28.8%), followed by Pros & Cons analysis ($n = 13$, 25.0%), Technology & Equipment introduction ($n = 12$, 23.1%), Promotional & Other content ($n = 7$, 13.5%), and Indications & Contraindications, which represented the smallest proportion ($n = 5$, 9.6%). (Figure 2, Figure 3, Figure4)

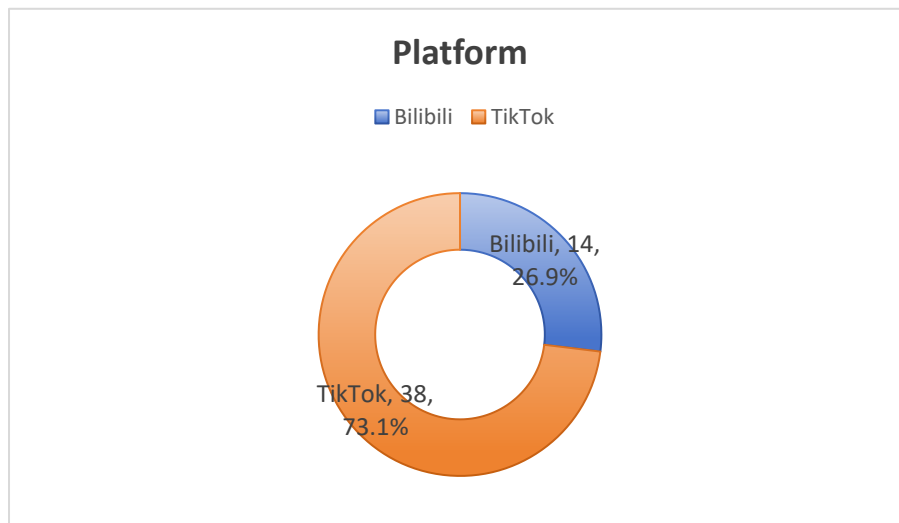


Figure 2. Distribution of included spinal navigation robot-related short videos across Bilibili and TikTok platforms.

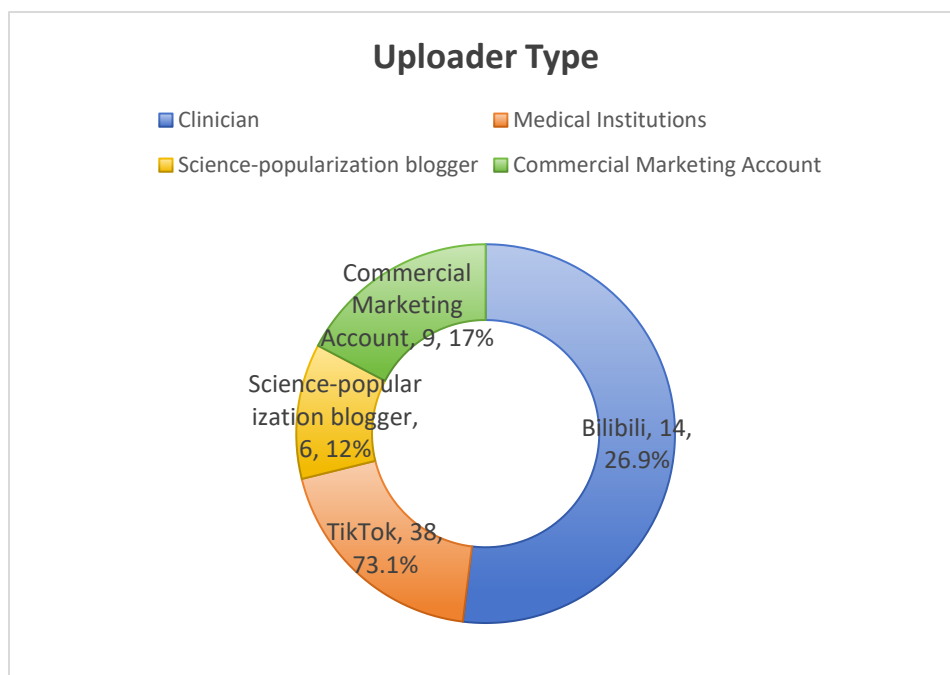


Figure 3. Composition of uploader types for the included spinal navigation robot-related

short-video samples.

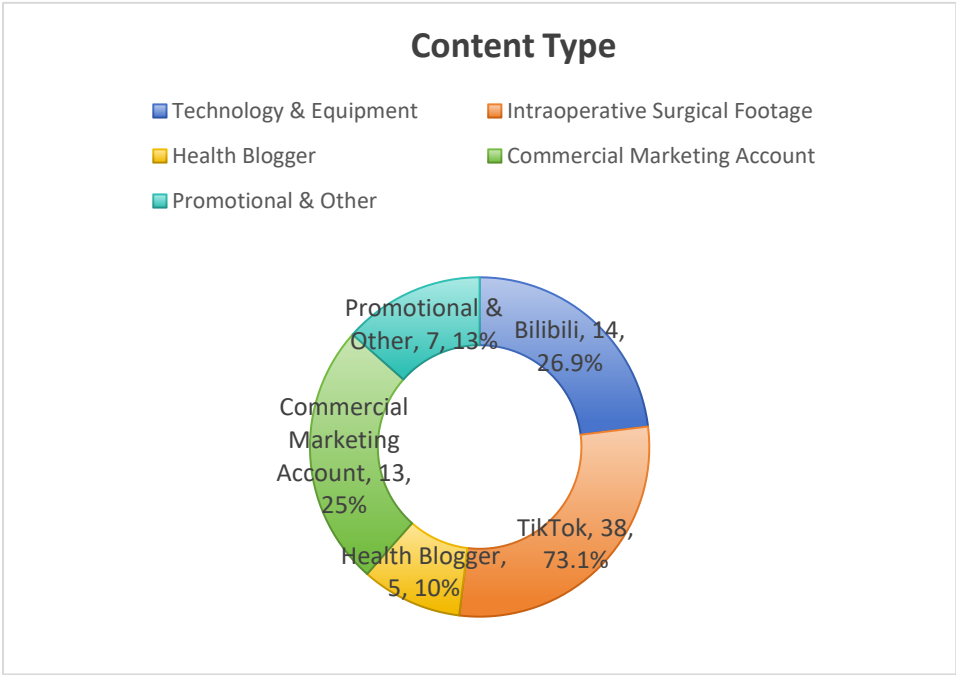


Figure 4. Classification and proportion of content themes among the included short-video samples.

3.2 Video characteristics, engagement metrics, uploader and content-theme distributions across Bilibili and TikTok

Inter-platform differences in video metadata, engagement metrics, quality scores, uploader composition and content profiles are summarised in **Table 1** and **Table 2**. A total of 52 eligible short-videos were included, with 14 videos from Bilibili and 38 from TikTok. Continuous variables were compared using the Mann-Whitney U (Wilcoxon rank-sum) test, while categorical variables were analysed via Fisher's exact test, with $P < 0.05$ defined as statistically significant.

Bilibili videos had a significantly longer median time since publication [1234 (762.5, 1507.75) days] and longer median duration [112.0 (91.0, 174.0) seconds] relative to TikTok [332 (194.75, 1260.75) days; 56.0 (32.75, 118.75) seconds; $P = 0.024$ and $P = 0.002$, respectively]. For user engagement, TikTok obtained significantly higher median likes [110.5 (39.5, 289.5)] and comments [10.5 (5.0, 52.0)] (both $P < 0.001$), whereas Bilibili showed higher median collection counts [30.0 (15.0, 116.75), $P = 0.006$]. No significant inter-platform difference was observed for shares ($P = 0.243$). In terms of information quality, TikTok videos achieved higher median GQS [3.0 (2.75, 4.0)] and modified DISCERN scores [2.0 (2.0, 2.25)] compared with Bilibili (GQS: $P = 0.021$; DISCERN: $P = 0.012$).

The distribution of uploader types differed substantially across platforms ($P = 0.001$). Clinician-authored videos constituted only 14.3% (2/14) of Bilibili samples versus 65.8% (25/38) on TikTok. Bilibili featured larger proportions of commercial marketing accounts (28.6%) and science-popularization bloggers (35.7%), while the proportion of institutional accounts was comparable between the two groups. Content-theme distribution also differed significantly ($P = 0.001$). Bilibili content was predominantly technology-and-equipment-oriented (57.1%), with zero videos addressing indications-contraindications or benefit-risk analysis. By contrast, TikTok delivered more balanced and diverse content, including 13.2% videos on indications-contraindications and 34.2% videos analysing pros-and-cons; intraoperative surgical footage was prevalent on both platforms.

Table 1. Comparison of video characteristics, engagement metrics and quality scores between Bilibili and TikTok for spinal navigation robot-related short videos.

Variable	Total, $n = 52$ median (IQR)	Bilibili, $n = 14$ median (IQR)	TikTok, $n = 38$ median (IQR)	Wilcoxon rank-sum test	
				Z score	P score
Days since published	635.0 (221.75, 1363.50)	1234 (762.5, 1507.75)	332 (194.75, 1260.75)	-2.2259	0.024
Likes	63.5 (23.5, 230.75)	15.0 (6.50, 42.75)	110.5 (39.5, 289.5)	-3.724	0.00
Comments	8.0 (1.25, 28.0)	0 (0, 1.25)	10.5 (5.0, 52.0)	-4.502	0.00
Collections	14.5 (3.75, 38.5)	30.0 (15.0, 116.75)	8.5 (3.0, 27.75)	-2.736	0.006
Shares	15.5 (5.0, 60.25)	32.5 (7.25, 97.25)	14.5 (4.75, 33.25)	-1.166	0.243
Duration	80.0 (37.5, 130.0)	112.0 (91.0, 174.0)	56.0 (32.75, 118.75)	-3.064	0.002
GQS	3.0 (2.0, 4.0)	2.0 (1.75, 3.0)	3 (2.75, 4.0)	-2.302	0.021
DISCERN score	2.0 (1.0, 2.0)	1.0 (1.0, 2.0)	2.0 (2.0, 2.25)	-2.509	0.012

Note: Data are presented as median (interquartile range, IQR). Mann-Whitney U (Wilcoxon rank-sum) test was used for inter-platform comparison. $P < 0.05$ was considered statistically significant. GQS = Global Quality Score; DISCERN = modified DISCERN score.

Table 2. Distribution of uploader types and content themes of spinal navigation robot-related short videos between Bilibili and TikTok.

Variable	Bilibili (n=14)	TikTok (n=38)	P 值
Uploader Type			0.001 ^a
Clinician	2 (14.3)	25 (65.8)	
Institutional account	3 (21.4)	7 (18.4)	
Commercial marketing account	4 (28.6)	2 (5.3)	
Science-popularization blogger	5 (35.7)	4 (10.5)	

Content Type			0.001 ^a
Technology & Equipment	8 (57.1)	4 (10.5)	
Intraoperative Surgical Footage	5 (35.7)	10 (26.3)	
Indications & Contraindications	0 (0.0)	5 (13.2)	
Pros & Cons	0 (0.0)	13 (34.2)	
Promotional & Other	1 (7.1)	6 (15.8)	

Note: Data are expressed as n (%). Fisher's exact test was applied to compare the categorical variables between platforms. $P < 0.05$ was considered statistically significant.

3.3 Quality scores stratified by uploader type and video content theme

Presents quality-score stratification according to uploader type and content theme. Kruskal-Wallis H tests revealed significant overall differences across four uploader subgroups for both GQS ($H = 15.145$, $df = 3$, $P = 0.002$) and modified DISCERN scores ($H = 17.061$, $df = 3$, $P = 0.001$). Clinician-generated videos achieved the highest median GQS [3 (3, 4)] and mDISCERN scores [2 (2, 3)], followed by institutional-account videos. Commercial marketing accounts obtained the lowest quality metrics (**Table 3**). Dunn-Bonferroni pairwise comparisons demonstrated that videos released by clinicians had significantly higher GQS and mDISCERN scores compared with those from commercial marketing accounts (adjusted $P = 0.013$; adjusted $P = 0.017$) and science-popularization bloggers (adjusted $P = 0.019$; adjusted $P = 0.004$). No significant pairwise differences were observed between clinicians and institutional accounts, nor between commercial marketing accounts and science-popularization bloggers (**Table 4**).

When stratified by content theme, overall GQS differed significantly among five content categories ($H = 18.267$, $df = 4$, $P = 0.01$), whereas overall mDISCERN scores did not reach statistical significance ($H = 13.510$, $df = 4$, $P = 0.09$). Videos addressing pros-and-cons analysis attained the highest median GQS [4 (3, 4.5)], while pure technology-and-equipment videos obtained the lowest GQS [2 (1.25, 2.0)] (**Table 5**). Post-hoc tests showed that pros-and-cons videos had significantly higher GQS than technology-and-equipment-focused videos (adjusted $P = 0.023$). Intraoperative surgical-footage videos exhibited significantly higher GQS and mDISCERN relative to indications-and-contraindications-themed videos (adjusted $P < 0.001$; adjusted $P = 0.006$). Most remaining pairwise comparisons across content subgroups were non-significant (**Table 6**).

Table 3. Comparison of GQS and mDISCERN scores among different uploader types (Kruskal–Wallis test)

Uploader Type	N (52)	GQS Score	mDISCERN Score
Clinician	27	3 (3, 4)	2 (2, 3)
Institutional account	10	3 (2, 4)	2 (1, 3)
Commercial marketing account	9	2 (1.75, 2.25)	1 (0.75, 2)
Science-popularization blogger	6	2 (1.5, 3)	1 (1, 2)

Kruskal-Wallis H	H=15.145, df=3	H=17.061, df=3
Overall P-value	0.002	0.001

Note: Data are presented as median (interquartile range). Between-group comparisons were performed using the Kruskal-Wallis H test. df = degree of freedom.

Table 4. Post-hoc pairwise comparisons of GQS and mDISCERN scores among uploader types (Dunn's test with adjusted P)

Comparison	GQS Score Adjusted P	mDISCERN Score Adjusted P
Commercial marketing vs Science-popularization blogger	1.000	1.000
Commercial marketing vs Institutional account	0.721	0.417
Commercial marketing vs Clinician	0.013*	0.017*
Science-popularization blogger vs Institutional account (4 vs 2)	1.000	0.287
Science-popularization blogger vs Clinician	0.019*	0.004*
Institutional account vs Clinician	0.681	1.000

Note: Data are presented as median (interquartile range), M(P₂₅, P₇₅). Between-group comparisons were performed using the Kruskal-Wallis H test. When the overall comparison was statistically significant, Dunn-Bonferroni post-hoc test was used for pairwise multiple comparisons. *Indicates adjusted P < 0.05.

Table 5. Comparison of GQS and mDISCERN scores among different content types (Kruskal-Wallis test)

Content Type	N (52)	GQS Score	mDISCERN Score
Technology & Equipment	12	2 (1.25, 2.0)	1 (1, 2)
Intraoperative Surgical Footage	15	3 (3, 4)	2 (2, 2)
Indications & Contraindications	5	2 (2.5, 3)	1 (1.5, 2)
Pros & Cons	13	4 (3, 4.5)	2 (2, 4)
Promotional & Other	7	3 (3, 4)	2 (1, 2)
Kruskal-Wallis H		H=18.267, df=4	H=13.510, df=4
Overall P-value		0.01	0.09

Table 6. Post-hoc pairwise comparisons of GQS and mDISCERN scores among content types (Dunn's test with adjusted P)

Comparison	GQS Score Adjusted P	mDISCERN Score Adjusted P
Technology & Equipment vs Intraoperative Surgical Footage	1.000	1.000
Technology & Equipment vs Indications & Contraindications	0.721	0.417
Technology & Equipment vs Pros & Cons	0.023*	0.106
Technology & Equipment vs Promotional & Other	1.000	1.000
Intraoperative Surgical Footage vs Indications & Contraindications	<0.001*	0.006*
Intraoperative Surgical Footage vs Pros & Cons	0.334	1.000
Indications & Contraindications vs Promotional & Other	1.000	1.000
Indications & Contraindications vs Pros & Cons	1.000	1.000
Indications & Contraindications vs Promotional & Other	1.000	0.454
Pros & Cons vs Promotional & Other	1.000	1.000

Note: Pairwise multiple comparisons were conducted using Dunn-Bonferroni post-hoc test. * Indicates adjusted $P < 0.05$, statistically significant.

4. Spearman's rank-order correlation analysis of video-related indicators

illustrates Spearman's rank-order correlations among engagement metrics, video duration, days since publication and quality scores. Among engagement indicators, likes were positively correlated with comments ($r_s = 0.832$, $P < 0.01$), saves ($r_s = 0.445$, $P < 0.01$) and shares ($r_s = 0.576$, $P < 0.01$). Comments were positively correlated with saves ($r_s = 0.225$, $P = 0.108$) and shares ($r_s = 0.446$, $P < 0.01$). Saves exhibited strong positive correlation with shares ($r_s = 0.705$, $P < 0.01$). Video duration showed weak negative correlation with likes ($r_s = -0.284$, $P = 0.041$), while no significant correlation was detected between duration and other engagement indicators. Days since publication was positively associated with saves ($r_s = 0.352$, $P = 0.010$) and shares ($r_s = 0.655$, $P < 0.01$). (Table 7)

Regarding correlation between engagement metrics and quality scores: **Likes** ($r_s = 0.311$, $P = 0.025$) and **comments** ($r_s = 0.379$, $P = 0.006$) were positively and significantly correlated with GQS. For mDISCERN score, only comments demonstrated significant positive correlation ($r_s = 0.319$, $P = 0.021$). Saves, shares, video duration and days since publication had no statistically significant correlation with either GQS or mDISCERN scores. (Table 8)

Table 7. Spearman's rank-order correlation among video engagement metrics, duration and days since publication

Variable	Likes	Comments	Saves	Shares	Duration	Days since published
Likes						
r	1					
P value						
Comments						
r	0.832	1				
P value	<0.01					
Saves						
r	<0.445	0.225	1			
P value	<0.01	0.108				
Shares						
r	0.576	0.446	0.705	1		
P value	<0.01	<0.01	<0.01			
Duration						
r	-0.284	-0.272	0.078	0.060	1	
P value	0.041	0.051	0.582	0.673		
Days since published						
r	0.124	-0.002	0.352	0.655		1
P value	0.381	0.987	0.010	<0.01		

Note: r_s represents Spearman's rank-order correlation coefficient. $P < 0.05$ was considered statistically significant.

Table 8. Spearman's rank-order correlation of engagement metrics with GQS and modified DISCERN quality scores

Variable	Global quality scores	DISCERN
Likes		
r	0.311	0.212
P value	0.025*	0.132
Comments		
r	0.379	0.319
P value	0.006**	0.021*
Saves		

r	-0.093	-0.150
P value	0.514	0.287
Shares		
r	-0.119	-0.181
P value	0.402	0.199
Duration		
r	-0.180	-0.162
P value	0.202	0.252
Days since published		
r	-0.243	-0.148
P value	0.083	0.295

Note. Spearman's rank-order correlation coefficient (r_s) was presented. GQS = Global Quality Score; mDISCERN = modified DISCERN score. * $P < 0.05$ was considered statistically significant.

5. GEE-based robust Poisson regression was performed to identify predictors of GQS and mDISCERN total scores

For the **GQS score (Table 9)**, in the adjusted multivariable model, content type showed overall significant inter-group difference ($P=0.001$). Compared with the reference category Promotional & Other, videos of Technology & Equipment were associated with lower GQS scores (adjusted IRR = 0.614, 95%CI 0.442-0.853, $P=0.004$). Among interaction metrics, higher likes (adjusted IRR = 1.000, $P=0.025$), fewer comments (adjusted IRR = 0.997, 95%CI 0.994-0.999, $P=0.011$) and longer video duration (adjusted IRR = 1.001, $P=0.003$) were independently associated with GQS score. Platform and uploader type were not independent predictors of GQS after adjustment.

For the **mDISCERN total score (Table 10)**, the overall difference across content types remained statistically significant ($P=0.002$). Clinician-released videos exhibited significantly higher mDISCERN scores relative to science-popularization bloggers (adjusted IRR = 1.713, 95%CI 1.014-2.892, $P=0.044$). No video engagement metrics or video duration were independently correlated with mDISCERN total score in the adjusted model.

Table 9. Results of GEE-based robust Poisson regression for predictors of GQS score

Variable	Crude IRR (95%CI)	P-value	Adjusted IRR (95%CI)	P-value
Uploader Type				0.687
Bilibili			0.945(0.716-1.246)	0.687
TikTok			1.000	
UploaderType				0.049

Clinician			1.271(0.900-1.795)	0.173
Institutional account			1.032(0.693-1.536)	0.876
Commercial marketing account			0.894(0.675-1.185)	0.437
Science-popularization blogger			1.000	
Content Type				0.001
Technology & Equipment			0.614(0.442-0.853)	0.004
Intraoperative Surgical Footage			0.935(0.687-1.273)	0.671
Indications & Contraindications			0.758(0.551-1.044)	0.090
Pros & Cons			1.044(0.795-1.370)	0.758
Promotional & Other			1.000	
Likes	1.000(1.000-1.000)	0.045	1.000(1.000-1.000)	0.025
Comments	0.998(0.995-1.001)	0.246	0.997(0.994-0.999)	0.011
Collections	0.998(0.996-1.000)	0.100	0.999(0.997-1.001)	0.326
Shares	1.001(1.000-1.002)	0.136	1.001(1.000-1.002)	0.111
Duration	1.000(0.999-1.001)	0.521	1.001(1.000-1.002)	0.003

*Note.*GEE-robust Poisson regression model. IRR = incidence rate ratio; 95%CI = 95% confidence interval. Crude IRR: univariate model; Adjusted IRR: multivariable model.

Table10.Results of GEE-based robust Poisson regression for predictors of mDISCERN total score

Variable	Crude IRR (95%CI)	P-value	Adjusted IRR (95%CI)	P-value
Uploader Type				0.191
Bilibili			0.796(0.566-1.120)	0.191
TikTok			1.000	
UploaderType				0.116
Clinician			1.713(1.014-2.892)	0.044
Institutional account			1.585(0.868-2.892)	0.134
Commercial marketing account			0.987(0.567-1.719)	0.963
Science-popularization blogger			1.000	
Content Type				0.002
Technology & Equipment			0.985(0.567-1.711)	0.957
Intraoperative Surgical Footage			1.227(0.801-1.878)	0.347
Indications &			0.822(0.541-1.247)	0.356

Contraindications				
Pros & Cons			1.284(0.856-1.928)	0.227
Promotional & Other			1.000	
Likes	1.000(1.000-1.000)	0.045	1.000(1.000-1.000)	0.285
Comments	0.998(0.995-1.001)	0.246	0.996(0.991-1.002)	0.199
Collections	0.998(0.996-1.000)	0.100	0.998(0.995-1.001)	0.268
Shares	1.001(1.000-1.002)	0.136	1.001(0.999-1.003)	0.272
Duration	1.000(0.999-1.001)	0.521	1.001(1.000-1.003)	0.088

Note. GEE-robust Poisson regression model. IRR = incidence rate ratio; 95%CI = 95% confidence interval. Crude IRR: univariate model; Adjusted IRR: multivariable model. mDISCERN = modified DISCERN score.

4. Discussion

4.1 Principal findings

This cross-sectional content-analysis study systematically characterised 52 Chinese-language short-videos regarding spinal-navigation-robot surgery collected from TikTok and Bilibili. Overall, the quality and reliability of included videos remained suboptimal. Inter-platform heterogeneity existed in video-duration characteristics, engagement-metric distributions, uploader-composition patterns and thematic coverage. Physician-produced videos yielded significantly higher GQS and mDISCERN scores than content generated by commercial-marketing accounts and non-medical science-popularisation bloggers. Content addressing balanced benefit-risk analysis achieved the highest GQS values, whereas pure hardware-technology-oriented videos exhibited poorer overall quality.

Spearman-rank-order correlation revealed partial weak-to-moderate correlations between engagement indicators and quality scores: likes and comments correlated positively with GQS, while only comment volume was positively associated with mDISCERN. Nevertheless, correlation magnitudes were limited, suggesting high-traffic metrics cannot reliably signal high-reliability medical information. GEE-robust Poisson regression further demonstrated that content type constituted an independent predictor for both GQS and mDISCERN outcomes. Technology-and-equipment-focused videos independently predicted lower GQS after multivariable adjustment. Clinician-authored content was an independent positive predictor for mDISCERN reliability scores relative to science-popularisation bloggers, whereas platform origin lost independent predictive value after adjusting for confounding variables. Collectively, these findings highlight that **creator background and content theme exert more decisive impacts on information quality than platform or user-interaction performance.

4.2 Comparison with existing literature

Consistent with previous orthopaedic-short-video studies focusing on knee-arthroplasty and other musculoskeletal-surgery topics^[8,9,18-20], physician-generated videos demonstrated superior quality and reliability compared with non-medical commercial or blogger-produced materials. Unlike joint-surgery-related short-video datasets, however, spinal-navigation-robot-related

videos exhibited distinct thematic imbalance. Videos covering indications, contraindications and balanced risk-benefit discussions were scarce on Bilibili in our sample, which may reflect platform-specific user-content preferences: Bilibili favoured hardware-principle-oriented explanatory videos, while TikTok hosted more diverse clinical-decision-oriented content.

Similar to prior Chinese dual-platform comparative investigations^[21-23], Bilibili videos in our cohort were longer, whereas TikTok videos obtained higher likes and comment volumes, reflecting the divergent product-logic of the two platforms. Notably, higher user-engagement metrics were only weakly correlated with quality indicators in the present work. This echoes repeated observations across social-media-health-information research: public popularity does not equate to medical accuracy or information completeness^[24-26]. Although correlation existed between partial interaction indices and GQS, no independent predictive effect of engagement metrics emerged for mDISCERN scores in adjusted Poisson-regression models. This distinction between GQS and mDISCERN is clinically meaningful: GQS mainly reflects comprehensibility and presentation quality, whereas mDISCERN emphasises objectivity, risk disclosure and evidence-based balance. Our results implied that viral, high-comment-volume videos might be easy-to-understand yet remain potentially biased regarding risk-and-benefit disclosure.

4.3 Mechanistic interpretation and clinical implications

Spinal-robot-assisted surgery represents high-risk, high-cost advanced-technology orthopaedic care^[27]. Patients searching for preoperative information are highly vulnerable to one-sided promotion. If social-media content overemphasises technical advantages while concealing complications, contraindications and learning-curve limitations, patients may form unrealistic expectations for surgical outcomes, which may subsequently amplify doctor-patient communication barriers^[28].

4.4 Our study underscores the value of physician-participated online popular-science.

Clinician uploaders achieved superior mDISCERN reliability scores, indicating that medical-professional participation improves risk-disclosure and evidence-balance performance^[29]. However, physician-produced videos were scarce on Bilibili within our sample. Platform-side medical-creator-certification mechanisms could be optimised to encourage spinal-surgeon-produced evidence-based content on Bilibili. Moreover, pure technical-equipment-oriented videos showed relatively low GQS scores. Mere hardware-introduction material often omits practical patient-oriented decision-making information, even if technically accurate^[30]. Therefore, future medical-popular-science production should move beyond device-feature demonstration and strengthen interpretation of indications, contraindications, risks and cost-related considerations for patient audiences.

4.5 Recommendations for clinical practice and platform governance

For spinal-surgeon clinicians, active participation in standardised short-video popular-science is encouraged. When producing short-video content for public education, clinicians should prioritise balanced reporting of both benefits and risks instead of merely demonstrating surgical-robot-technique highlights.

For short-video-platform regulators, multiple measures could be implemented. First, medical-creator-qualification verification should be strengthened for spinal-surgery-related high-risk-theme content. Second, recommendation-algorithms should appropriately increase exposure for physician-generated balanced-risk-benefit-analysis content, instead of purely prioritising high-play-volume or high-interaction outputs. Third, warning-label reminders could be attached for high-risk surgical-technology-related videos, advising viewers to consult in-person clinical specialists for individualised treatment-decision advice rather than making surgical choices solely according to online-video resources.

Limitations

Several limitations of this study should be acknowledged. First, only Chinese-language public short-videos from TikTok and Bilibili were analysed; other social-media platforms were not covered. Second, GQS and modified DISCERN scales are widely adopted for orthopaedic-video-evaluation in Chinese-context studies, yet these tools were originally adapted from written-health-information-assessment instruments and have not obtained formal full-validation for short-video-multimedia-scenarios. Third, the sample size of the research is relatively limited, which stems from the fundamental constraint that the overall output volume of Chinese short videos related to spinal navigation robots on mainstream short-video platforms is inherently scarce, which may restrict statistical power for subgroup-stratified analyses. Future multi-platform larger-sample-size investigations are required to validate our conclusions.

Conclusions

The overall quality and reliability of spinal-navigation-robot-related short-videos on TikTok and Bilibili remain unsatisfactory. Uploader-type and content theme strongly predict video-information quality, while platform-level effects diminish after multivariable adjustment. High-user-engagement metrics cannot serve as reliable proxies for medically balanced, trustworthy information. Increased participation of spinal-surgery clinicians, optimised platform-qualification-audit workflows and priority-promotion of balanced risk-benefit-analysis content are urgently required to improve the quality of publicly-available online educational resources for patients considering robotic spinal-surgery interventions.

Author contributions

Qidi Geng participated in the conception and design of the study, drafted the original manuscript, and performed research-related investigation. Mingzhe Li was responsible for study methodology and formal statistical analysis. Yaxuan Su completed data coding and relevant investigation. Wei Guo provided overall supervision and revised the manuscript. All authors reviewed and approved the final version of the manuscript.

Funding

This research received no external funding.

Data availability

The datasets generated and analysed during the current study are available from the corresponding author upon reasonable request.

参考文献

- [1].Gulec A, Eravsar E, Ciftci S, Beyoglu A, Aydin BK. Ocular Safety of Unilateral Biportal Endoscopic Spinal Surgery: An Optical Coherence Tomography Angiography-Based Analysis. *J Clin Med*. 2026 Feb 26;15(5):1774. doi: 10.3390/jcm15051774. PMID: 41827193; PMCID: PMC12985841.
- [2].Qi Z, Da H, Yanming F, Mingxing F. Current status and prospects of robot-assisted spine surgery. *Expert Rev Med Devices*. 2025 Mar;22(3):187-192. doi: 10.1080/17434440.2025.2467779. Epub 2025 Feb 17. PMID: 39953843.
- [3].Yeo QY, Pham MH, Oh JY. Single-Position Robotic-Assisted Prone Lateral Fusion: Technical Description and Feasibility. *Asian Spine J*. 2024 Feb;18(1):118-123. doi: 10.31616/asj.2023.0215. Epub 2024 Feb 21. PMID: 38379151; PMCID: PMC10910140.
- [4].Kent R. Minimally Invasive Robotic-Assisted Complex Adult Spinal Deformity Correction in a Surgical Specialty Hospital: Bringing Adult Spinal Deformity Care Closer to Home. *J Clin Med*. 2026 Apr 11;15(8):2913. doi: 10.3390/jcm15082913. PMID: 42074715; PMCID: PMC13116422.
- [5].Zheng X, Li Q, Jin L, Shi K, Deng M. Quality, reliability, and dissemination of lung cancer information on short-video platforms in China: a cross-platform content analysis of TikTok, Kwai, and Rednote. *Front Public Health*. 2025 Dec 1;13:1683561. doi: 10.3389/fpubh.2025.1683561. PMID: 41404572; PMCID: PMC12702867.
- [6].Faraj D, Galgano M. Redefining the applications of navigation in spine surgery. *Neurosurg Focus Video*. 2026 Jul 1;15(1):V2. doi: 10.3171/2026.4.FOCVID2631. PMID: 42529674; PMCID: PMC13418630.
- [7].Xu R, Ren Y, Li X, Su L, Su J. The quality and reliability of short videos about premature ovarian failure on Bilibili and TikTok: Cross-sectional study. *Digit Health*. 2025 Jun 17;11:20552076251351077. doi: 10.1177/20552076251351077. PMID: 40534890; PMCID: PMC12174690.
- [8].Li K, Su W, Yang Y, Dai H, Wu Y. Engagement-quality mismatch in robot-assisted total knee arthroplasty short videos on TikTok and Bilibili. *J Robot Surg*. 2026 Apr 21;20(1):447. doi: 10.1007/s11701-026-03411-8. PMID: 42010150.
- [9].Wei B, Zhang X, Chen L, Ren X, Li Y, Chen F, Zhang L, Fan Y, Xie Z, Li J, Chen Y, Wang S. A comparative study of medical information quality and dissemination efficacy of knee arthroplasty videos in Bilibili/TikTok short video platforms. *Digit Health*. 2026 Feb 4;12:20552076261421072. doi: 10.1177/20552076261421072. PMID: 41659053; PMCID: PMC12873082.
- [10].Tu J, Zhang C, Zhang H, Liang L, He J. Evaluating the reliability and quality of knee osteoarthritis educational content on TikTok and Bilibili: A cross-sectional content analysis. *Digit Health*. 2025 Aug 12;11:20552076251366390. doi: 10.1177/20552076251366390. PMID: 40808711; PMCID: PMC12344336.
- [11].Gezmiş CT, Çakıroğlu B. YouTube as a source of patient information on robot-assisted radical prostatectomy: a cross-sectional study of quality, reliability, transparency, and content gaps. *Ther Adv Urol*. 2026 Jul 24;18:17562872261469665. doi: 10.1177/17562872261469665. PMID: 42502269; PMCID: PMC13400912.
- [12].Zhu Y, Cui L, Zhou T, Xu H. The quality and reliability of short videos about pulmonary hypertension on Bilibili and TikTok: a cross-sectional study. *BMC Cardiovasc Disord*. 2026 Feb 12;26(1):236. doi: 10.1186/s12872-026-05597-z. PMID: 41680686; PMCID: PMC12998168.
- [13].Ding R, Zhou J, Tang H, Shi L, Wang Z, Gu M, Deng Z, Gui Z, Tang Z. Evaluating TikTok and YouTube as patient-education resources on kidney transplantation: a comparative analysis. *Front*

Public Health. 2026 Mar 13;14:1764270. doi: 10.3389/fpubh.2026.1764270. PMID: 41908743; PMCID: PMC13021784.

[14].Ma T, Li X, Yu Z, Qu W, Xie W, Cong H. Quality and reliability of anterior cruciate ligament injury-related short Chinese videos on Bilibili and TikTok: Cross-sectional study. Digit Health. 2026 Jan 27;12:20552076261418829. doi: 10.1177/20552076261418829. PMID: 41613991; PMCID: PMC12847663.

[15].Bernard A, Langille M, Hughes S, Rose C, Leddin D, Veldhuyzen van Zanten S. A systematic review of patient inflammatory bowel disease information resources on the World Wide Web. Am J Gastroenterol. 2007 Sep;102(9):2070-7. doi: 10.1111/j.1572-0241.2007.01325.x. Epub 2007 May 19. PMID: 17511753.

[16].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 Feb;53(2):105-11. doi: 10.1136/jech.53.2.105. PMID: 10396471; PMCID: PMC1756830.

[17].Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977 Mar;33(1):159-74. PMID: 843571.

[18].Cai F, Yang F, Zhang R. Analysis of content, quality, and reliability of temporomandibular disorder-related Chinese videos on TikTok and Bilibili: A cross-sectional study. Medicine (Baltimore). 2026 Jun 19;105(25):e49340. doi: 10.1097/MD.00000000000049340. PMID: 42332477; PMCID: PMC13286456.

[19].Zhang S, Yang X, Bai H, Wang H, Chen Y, Liu P, Zhou C, Li H, Zhang M. Evaluating the reliability and quality of osteoporosis content on TikTok and Bilibili: A cross-sectional content analysis. Digit Health. 2026 May 20;12:20552076261455205. doi: 10.1177/20552076261455205. PMID: 42183307; PMCID: PMC13191130.

[20].Zhang Y, Huang C, Tong Y, Teng Y, Wan B, Huang J, Zeng G, Zong S. Quality assessment of spinal cord injury-related health information on short-form video platforms: Cross-sectional content analysis of TikTok, Kwai, and Bilibili. Digit Health. 2025 Sep 2;11:20552076251374226. doi: 10.1177/20552076251374226. PMID: 40918080; PMCID: PMC12409039.

[21].Xia C, Rong T, Wang B. A cross sectional content analysis evaluates chemotherapy health information quality and reliability on TikTok and Bilibili. Sci Rep. 2026 Jun 15;16(1):27570. doi: 10.1038/s41598-026-56880-0. PMID: 42297889; PMCID: PMC13538549.

[22].Shen N, Xu C, Yang Y, Zhang D. Evaluating the quality and reliability of Kawasaki disease-related content on TikTok and Bilibili: a cross-sectional study. Front Public Health. 2026 Jan 22;13:1664542. doi: 10.3389/fpubh.2025.1664542. PMID: 41658290; PMCID: PMC12872735.

[23].Wang Y, Xu Q, Lei Y, Ma D. Assessment of information quality and reliability on plantar fasciitis in short video platforms: A cross-sectional study of TikTok and Bilibili. Digit Health. 2026 Apr 16;12:20552076261444451. doi: 10.1177/20552076261444451. PMID: 42011434; PMCID: PMC13091981.

[24].Dong H, Gong Y, Zhao Z, Wang K, Zhang X, Ji F. Quality and reliability of sarcopenia-related videos on Bilibili and TikTok: a cross-sectional content analysis study. BMC Public Health. 2026 Jan 12;26(1):517. doi: 10.1186/s12889-025-26154-x. PMID: 41526888; PMCID: PMC12888518.

[25].Xu R, Wu Z, Chen S. Quality of rib fracture-related video information on TikTok and Bilibili: a cross-sectional study. J Thorac Dis. 2026 Apr 30;18(4):344. doi: 10.21037/jtd-2025-1-2502. Epub 2026 Mar 27. PMID: 42182739; PMCID: PMC13190219.

[26].Zhou Q, Shen J, Cao D, Yu W, Chen J. Quality and accuracy of scoliosis-related short videos on

TikTok reviewed in China in August 2025. *Sci Rep.* 2025 Dec 17;15(1):44007. doi: 10.1038/s41598-025-27684-5. PMID: 41408068; PMCID: PMC12711883.

[27].Li T, Yan J, Li J, Shang Y, Tang X. Robot-assisted technique versus freehand technique for spine surgery: an umbrella review. *Ann Med.* 2025 Dec;57(1):2523564. doi: 10.1080/07853890.2025.2523564. Epub 2025 Jul 9. PMID: 40631354; PMCID: PMC12243020.

[28].Tu J, Xie S, Zheng J, Lin Z, Ren W. Quality and reliability of femoral neck fracture educational short videos: a cross-sectional study. *Sci Rep.* 2026 Mar 30;16(1):10652. doi: 10.1038/s41598-026-46431-y. PMID: 41912793; PMCID: PMC13040079.

[29].Fan RC, Chen QB, Zheng XX, Huang LJ, Hong CL. Quality and reliability of chest Pain-Related short-form health videos on social media: A cross-sectional content analysis. *Digit Health.* 2026 May 12;12:20552076261450409. doi: 10.1177/20552076261450409. PMID: 42145323; PMCID: PMC13172721.

[30].Zhou M, Pan Y, Zhang Y, Song X, Zhou Y. Evaluating AI-generated patient education materials for spinal surgeries: Comparative analysis of readability and DISCERN quality across ChatGPT and deepseek models. *Int J Med Inform.* 2025 Jun;198:105871. doi: 10.1016/j.ijmedinf.2025.105871. Epub 2025 Mar 13. PMID: 40107040.