ORIGINAL ARTICLE DOI: 10.20515/otd.1971695

Osmangazi Journal of Med, 2026;48(5):888-902

Osmangazi Journal of Medicine

e-ISSN: 2587-1579

Neonatal Fever as a Digital Emergency: A Clinical Safety Audit of Turkish-Language YouTube Health Information

Dijital Bir Acil Durum Olarak Yenidoğan Ateşi: Türkçe YouTube Sağlık İçeriklerinin Klinik Güvenlik Açısından Denetimi

Elif Keles Gülnerman[1], Ibrahim Murat Hirfanoglu[2]

[1]Department of Radiology, Northwestern University Feinberg School of Medicine, Machine and Hybrid Intelligence Lab, Chicago, IL, USA
[2]Division of Neonatology, Department of Pediatrics, Gazi University School of Medicine, Ankara, Türkiye

Correspondence / Sorumlu yazar
Elif KELEŞ GÜLNERMAN

Department of Radiology, Northwestern University Feinberg School of Medicine, Machine and Hybrid Intelligence Lab, Chicago, IL, USA

e-mail: elifkeles.dr@gmail.com

Received : 15.06.2026

Accepted : 07.07.2026

Abstract: To evaluate the clinical accuracy, information literacy, and public health safety of Turkish-language YouTube videos addressing neonatal fever, a cross-sectional content audit was conducted using three Turkish-language search terms. Of 112 screened videos, 24 met neonatal-specific inclusion criteria and were independently evaluated by two neonatologists. Video quality was assessed using GQS, JAMA criteria, mDISCERN, CRAAP, PEMAT-U/A, Newborn Fever Safety and Quality Score (NFSQS), and E-GRADE classification. Correlation and multivariable regression analyses identified factors associated with guideline adherence. The 24 videos accumulated 935,427 views. Inter-rater reliability was excellent (κ=0.88). Mean NFSQS was 8.38 ± 5.58 out of 24. Fifteen videos were classified as "Not Differentiate Newborn," whereas only three were rated "Safe and Useful." CRAAP score demonstrated a strong positive correlation with NFSQS (r_s = 0.762, p < 0.001), while view count was negatively associated with the Clinician Safety Index (r_s = −0.521, p = 0.009). Short-form videos (<1 minute) achieved significantly lower NFSQS scores than longer videos (3.88 ± 3.64 vs 10.62 ± 5.03, p = 0.004). In multivariable analysis, CRAAP score (β = 1.770, 95% CI 1.105–2.435, p < 0.001) and video duration (β = 0.182, p = 0.010) were independently associated with NFSQS. Most Turkish-language YouTube videos on neonatal fever provide inadequate neonatal-specific guidance, potentially contributing to unsafe caregiver decision-making. Information literacy quality and content depth, rather than popularity, determined guideline adherence. The predominance of incomplete short-form content and poor alignment between visibility and clinical quality raise important public health concerns for high-risk neonatal conditions.
Keywords: neonatal fever, youtube, neonatal sepsis, digital health literacy, patient education, guideline adherence

Özet: Yenidoğan ateşini ele alan Türkçe YouTube videolarının klinik doğruluğunu, bilgi okuryazarlığını ve halk sağlığı güvenliğini değerlendirmek amacıyla, üç Türkçe anahtar kelime kullanılarak kesitsel bir içerik denetimi gerçekleştirildi. Taranan 112 videodan 24'ü yenidoğana özgü dahil edilme kriterlerini karşıladı ve iki neonatolog tarafından değerlendirildi. Video kalitesi; GQS, JAMA Kriterleri, mDISCERN, CRAAP, PEMAT-U/A, Yenidoğan Ateşi Güvenlik ve Kalite Skoru (NFSQS) ve E-GRADE sınıflaması kullanılarak değerlendirildi. Kılavuz uyumuyla ilişkili faktörleri belirlemek amacıyla korelasyon ve çok değişkenli regresyon analizleri gerçekleştirildi. Analize dahil edilen 24 video, 935.427 görüntülenmeye ulaştı. Gözlemciler arası uyum mükemmeldi (κ = 0,88). Ortalama NFSQS skoru 24 üzerinden 8,38 ± 5,58 olarak bulundu. On beş video "Yenidoğanı Ayırt Etmiyor", yalnızca üç video ise "Güvenli ve Yararlı" olarak sınıflandırıldı. CRAAP skoru ile NFSQS arasında güçlü pozitif korelasyon saptandı (r_s = 0,762; p < 0,001), görüntülenme sayısı ise Klinisyen Güvenlik İndeksi ile negatif ilişki gösterdi (r_s = −0,521; p = 0,009). Kısa videoların (<1 dakika) NFSQS skorları, uzun videolara göre anlamlı derecede düşüktü (3,88 ± 3,64; 10,62 ± 5,03; p = 0,004). Çok değişkenli analizde CRAAP skoru (β = 1,770; %95 p < 0,001) ve video süresi (β = 0,182; p = 0,010) NFSQS ile bağımsız olarak ilişkili bulundu. Yenidoğan ateşi hakkında Türkçe YouTube videolarının çoğu, yenidoğana özgü yeterli rehberlik sunmamakta ve bakım verenlerin güvenli olmayan kararlar almasına neden olmaktadır. Bilgi okuryazarlığı kalitesi ve içerik derinliği, kılavuz uyumunun belirleyicileri olarak öne çıkmıştır. Eksik bilgi içeren kısa videoların yaygınlığı ve görünürlüğüyle klinik kalitesi arasındaki uyumsuzluk, yüksek riskli yenidoğan durumları açısından önemli halk sağlığı sorunlarına işaret etmektedir.
Anahtar Kelimeler: yenidogan, ates, YouTube, saglik okuryazarligi, hasta guvenligi, dijital saglik

Ethics Committee Approval: This study analyzed publicly available YouTube video content and did not involve human participants, patient data, or animal subjects. No personal identifying information was collected or recorded. In accordance with national regulations and institutional policies, studies using publicly available online content without human participant involvement do not require formal ethics committee approval.

Informed Consent: The authors stated that informed consent was not applicable due to the nature of the study.

Authorship Contributions: Concept:E.K.G. , I.M.H. Data Collection or Processin: E.K.G. , I.M.H.Analysis or Interpretation: E.K.G. , I.M.H Literature Search: E.K.G. , I.M.H. Writing: E.K.G. , I.M.H

Copyright Transfer Form: Copyright Transfer Form was signed by author.

Peer-review: Externally peer-reviewed.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The author declared that this study received no financial support

How to cite/ Atıf için: Keles Gülnerman E , Hirfanoglu IM, Neonatal Fever as a Digital Emergency: A Clinical Safety Audit of Turkish-Language YouTube Health Information, Osmangazi Journal of Medicine, 2026;48(5):888-902

1. Introduction

The digital transformation of healthcare communication has fundamentally altered the information-seeking behaviors of parents and caregivers (1). YouTube, with more than two billion monthly users worldwide, has become one of the most widely used platforms for obtaining health-related information, providing free, accessible, and visually engaging educational content that transcends language, literacy, and socioeconomic barriers (2). However, the platform's open-access structure allows medical information to be disseminated without peer review, editorial oversight, or formal professional vetting, creating an environment susceptible to health misinformation and the broader digital health infodemic(3-5).

The clinical implications of this phenomenon are particularly important in neonatal medicine. The first 28 days of life represent the most physiologically vulnerable period of infancy due to the immaturity of both innate and adaptive immune responses (6). Unlike older infants, in whom fever is frequently associated with self-limiting viral illnesses, fever in a neonate may be the earliest or only manifestation of a serious bacterial infection, including neonatal sepsis, meningitis, or urinary tract infection(7). Delayed recognition and management of these conditions can result in significant morbidity and mortality (8).

Current international guidelines from the American Academy of Pediatrics (AAP) and the National Institute for Health and Care Excellence (NICE) emphasize that a rectal temperature of ≥38.0°C in an infant younger than 28 days should be considered a medical emergency requiring immediate clinical evaluation(9-11). These guidelines do not recommend routine home antipyretic treatment as a substitute for urgent clinical evaluation. Failure to recognize these principles may delay diagnosis and increase the risk of adverse outcomes.

Despite these recommendations, online parenting content frequently conflates neonatal fever management with general pediatric fever advice, often recommending home monitoring, hydration, or antipyretic use without emphasizing the need for urgent neonatal evaluation. Furthermore, videos describing fever measurement frequently demonstrate non-rectal methods without clarifying their limitations in newborns. While numerous studies have evaluated the quality and reliability of YouTube health information using instruments such as the Global Quality Score (GQS), JAMA Benchmark Criteria, and DISCERN little is known about the clinical safety, evidence quality, information literacy, and actionability of neonatal fever content(12-14).

Recent studies have highlighted important limitations of traditional video-quality assessment tools. Kang et al. demonstrated a substantial "credibility–evidence gap" in physician-generated medical videos, showing that professionally presented content frequently relied on low levels of supporting evidence despite appearing authoritative and achieving favorable quality scores (15). These findings suggest that evaluating digital health content requires not only assessment of presentation quality but also appraisal of the underlying evidence supporting medical recommendations. Accordingly, the present study incorporated evidence-level assessment (E-GRADE), information-literacy evaluation (CRAAP), and neonatal-specific clinical assessment (NFSQS) to capture complementary dimensions of digital health information quality(16,17).

To address this gap, we conducted a comprehensive audit of Turkish-language YouTube videos related to neonatal fever management. In addition to established quality assessment instruments, videos were evaluated using the Newborn Fever Safety and Quality Score (NFSQS), a neonatal fever–specific assessment tool previously developed and internally validated by our group. The full development and validation study is currently under separate consideration for publication. Videos were also assessed using the CRAAP Test, the Patient Education Materials Assessment Tool (PEMAT), and the Evidence-GRADE (E-GRADE) framework (16,18). These instruments enabled a multidimensional evaluation of clinical safety, guideline adherence, evidence quality, information literacy, understandability, and actionability.

We hypothesized that clinically accurate neonatal fever content would be uncommon, that short-form videos would demonstrate lower guideline adherence than conventional-length videos, and that video popularity would not reliably reflect clinical quality. Therefore, the primary objectives of this study were to evaluate the clinical safety, guideline adherence, evidence quality, information literacy, and health-literacy characteristics of Turkish-language YouTube videos addressing neonatal fever

management, and to examine whether video visibility and popularity reflect clinically appropriate neonatal fever guidance.

2. Materials and Methods

2.1. Study Design and Search Strategy

This cross-sectional content audit was conducted systematically on YouTube (http://www.youtube.com/) using three Turkish keyword strings reflecting common parental search patterns: "Bebeklerde Ateş" (Fever in Babies), "Yenidoğan Bebeklerde Ateş" (Fever in Newborn Babies), and "Bebeklerde Ateş Nasıl Düşürülür" (How to Lower Fever in Babies). Searches were performed between March 15 and March 17, 2026. To eliminate personalized-search, geographic, and cookie-based bias, all searches were conducted in Incognito Mode with cleared browser cache and deactivated location data. Videos were categorized according to the primary individual or professional group delivering the educational content rather than the ownership of the YouTube channel itself. Thus, classification reflected the principal source of clinical information presented in the video and not necessarily the entity hosting or publishing the content.

2.2. Screening and Exclusion

The search yielded 112 unique videos. Two board-certified neonatologists independently reviewed and audited all 112 videos in their entirety, logging structural and thematic characteristics against international clinical criteria. Inter-observer agreement was quantified using Cohen's Kappa (κ); discrepancies were resolved by consensus.

Videos were excluded if they: (i) used the term "newborn" as a search engine optimization (SEO) keyword without providing neonatal-specific clinical thresholds, physiological guidance, or safety warnings; (ii) presented general childhood antipyretic dosing without neonatal qualification; (iii) failed to indicate the requirement for emergency evaluation in infants ≤ 28 days; (iv) demonstrated fever measurement techniques without specifying rectal thermometry as the neonatal gold standard; or (v) consisted of commercial product placement, news clips, or duplicate segments. Following this protocol, 88 videos (78.6%) were excluded, yielding a final sample of 24 videos for full evaluation.

2.3. Evaluation Instruments

To provide a multidimensional evaluation of neonatal fever videos, complementary assessment instruments were selected to measure distinct domains of information quality. Disease-specific clinical accuracy and neonatal safety recommendations were evaluated using the Newborn Fever Safety and Quality Score (NFSQS). Evidence quality underlying medical claims was assessed using the Evidence-GRADE (E-GRADE) framework (15). General source credibility and information literacy were evaluated using the CRAAP framework, whereas reliability and transparency were assessed using the modified DISCERN instrument and JAMA benchmark criteria(13,14,16). Educational value was evaluated using the Global Quality Scale (GQS), and patient-centered communication characteristics were assessed using the Patient Education Materials Assessment Tool for Audiovisual Materials (PEMAT-A/V), including understandability (PEMAT-U) and actionability (PEMAT-A) domains (12,18). Together, these instruments were selected to capture complementary aspects of digital health communication rather than overlapping constructs.

Seven validated or structured instruments were applied to each video:

GQS: The Global Quality Scale (GQS) is a five-point Likert instrument that evaluates the overall educational quality, flow, and usefulness of health information materials, with higher scores indicating greater educational value (12).

JAMA Benchmark Criteria: The JAMA benchmark criteria assess four dimensions of online health information quality, including authorship, attribution, disclosure, and currency. One point is assigned for each criterion met, yielding a total score ranging from 0 to 4 (13).

mDISCERN: The modified DISCERN instrument consisted of five items evaluating clarity of aims, use of reliable information sources, balance and objectivity, provision of additional sources, and discussion of uncertainties (14). Each item was scored as present (1) or absent (0), yielding a total score ranging from 0 to 5, with higher scores indicating greater reliability and transparency of the information provided (14).

CRAAP Test (0–15): The CRAAP framework (Currency, Relevance, Authority, Accuracy, and Purpose) was used to assess the reliability and credibility of video content. Originally developed as an information-literacy instrument for evaluating online sources, CRAAP has subsequently been applied in medical YouTube research (16,17). Because disease-specific instruments such as the Newborn Fever Safety and Quality Score (NFSQS)

primarily evaluate clinical content and guideline adherence, CRAAP was included to provide an independent assessment of source credibility, transparency, information currency, authorship, and potential commercial bias. This dual-assessment approach is consistent with previous YouTube-based medical education studies that combined disease-specific content evaluation with broader information-quality assessment. Each domain was scored from 0 to 3, yielding a total score ranging from 0 to 15, with higher scores indicating greater source reliability and information quality.

PEMAT-U and PEMAT-A: Patient-centered communication characteristics were evaluated using the Patient Education Materials Assessment Tool for Audiovisual Materials (PEMAT-A/V). Two complementary domains were assessed: Understandability (PEMAT-U), which measures the extent to which consumers of varying health-literacy levels can process and explain key messages, and Actionability (PEMAT-A), which evaluates whether viewers can identify specific actions to take based on the information presented (18). Items were scored according to the original PEMAT methodology and converted to percentage scores, with higher values indicating greater understandability and actionability.

NFSQS (0–24): The Newborn Fever Safety and Quality Score (NFSQS) is a neonatal fever–specific assessment instrument previously developed and internally validated by our group based on recommendations from the American Academy of Pediatrics (AAP), the National Institute for Health and Care Excellence (NICE), and expert consensus(9-11). Similar disease-specific assessment tools have previously been proposed in digital health research when generic quality instruments were insufficient to evaluate clinically relevant condition-specific content. The instrument consists of six sub-domains scored from 0 to 4: Fever Definition (including specification of rectal thermometry), Etiology, Red Flags, Emergency Care Awareness, Home Care and Monitoring, and Medication Safety (N_Med). Total scores range from 0 to 24, with higher scores indicating greater adherence to evidence-based neonatal fever management principles. A concise description of the NFSQS domains and their clinical rationale is provided in Supplementary Table S1. The Fever Definition sub-domain (FeverDef) was specifically designed to assess whether videos identified a rectal temperature ≥38.0°C as the diagnostic threshold for neonatal fever.

E-GRADE (Evidence-GRADE; 4 levels, A–D): Videos were classified according to an adapted version of the Evidence-GRADE framework proposed by Kang et al., which operationalizes the GRADE (Grading of Recommendations Assessment, Development and Evaluation) methodology for evaluating health-related video claims(15). In the original framework, individual medical claims are graded according to the strength of supporting evidence. Because neonatal fever videos typically contain a limited number of closely related recommendations, a single overall E-GRADE level was assigned to each video. Grade A indicated recommendations directly supported by clinical practice guidelines or systematic reviews; Grade B indicated support from randomized or high-quality observational studies; Grade C indicated limited observational evidence or physiological reasoning; and Grade D indicated anecdotal, unsupported, or potentially misleading recommendations. Classification was performed independently by two neonatologists following comprehensive review of video content and its consistency with contemporary neonatal fever management evidence.

2.4. Composite Indices

To facilitate exploratory comparisons across multiple quality domains, two composite metrics were calculated.

HKGE: Hybrid Clinical Safety Index (HKGE, acronym retained from the original Turkish designation Hibrit Klinik Güvenlik Endeksi). A composite measure integrating overall educational quality, information literacy, and neonatal guideline adherence. Instruments were normalized to a 0–100 scale with predefined clinical weighting. Weighting coefficients were determined a priori according to perceived clinical relevance and should be interpreted as exploratory rather than empirically derived parameters. Greater weight was assigned to NFSQS because neonatal fever represents a time-sensitive clinical emergency, and accurate recognition of neonatal-specific warning signs and guideline-concordant caregiver actions are considered the most critical determinants of patient safety. Therefore, neonatal clinical safety and guideline adherence were prioritized over general information quality measures in the composite index:

Overall Quality (OQ) = (GQS + JAMA + DISCERN) / 3

$$HKGE = [(OQ \times 0.20) + ((CRAAP/15 \times 5) \times 0.20) + ((NFSQS/24 \times 5) \times 0.60)] \times 20$$

CSI (Clinician Safety Index): An exploratory exposure-adjusted metric intended to estimate clinical quality relative to viewer traffic:

$$CSI = HKGE / [\log_{10}(\text{View Count}) \times 20]$$

Because neither HKGE nor CSI has undergone external validation, both measures were interpreted as hypothesis-generating composite indicators rather than standalone validated instruments.

2.5. Video Format Classification

Duration-based subgroup analysis included comparison of short-form videos (<1 minute) versus conventional-length videos (≥1 minute), as well as Kruskal–Wallis testing across duration quartiles to evaluate the association between video format and neonatal guideline adherence

2.6. Safety Classification Approach

Safety_Flag classification was assigned independently by two board-certified neonatologists through structured review of full video content, blinded to NFSQS sub-scores at the time of classification. Although higher NFSQS scores generally corresponded to safer content, Safety_Flag classification incorporated qualitative clinical review and therefore was not determined solely by numerical score thresholds. Each video was assigned to one of three categories based on whether it provided sufficient information to safely identify and manage fever in a neonate (≤ 28 days of age) as distinct from general pediatric fever:

Safe and Useful: the video explicitly identifies neonates as a distinct emergency population, specifies a rectal temperature ≥38.0°C as the diagnostic fever threshold, and provides accurate, actionable guidance for emergency referral.

Mostly Safe but Incomplete: the video contains clinically accurate fever-related information that does not contradict neonatal safety principles but does not explicitly differentiate neonatal-specific thresholds, red flags, or emergency referral criteria from general infant/child fever advice.

Not Differentiate Newborn: the video presents fever management guidance applicable to older infants or children without acknowledging that neonates require a fundamentally different, emergency-level approach, or contains content that could mislead a caregiver of a febrile neonate into delaying emergency evaluation.

Inter-rater reliability for Safety_Flag assignment was $\kappa = 0.88$ (95% CI: 0.81–0.95), indicating almost perfect agreement between assessors. NFSQS sub-domain scores, including the Fever Definition sub-domain (FeverDef, NFSQS domain 1; 0–4) and the Medication Safety sub-domain (N_Med, NFSQS domain 6; 0–4), were calculated independently of, and subsequent to, Safety_Flag assignment, and were used to characterize rather than determine the resulting categories (Section 3.2).

2.7. Statistical Analysis

Statistical analyses were performed using Python 3.10 (pandas, numpy, scipy.stats, and statsmodels). Normality was assessed using the Shapiro–Wilk test. Continuous variables are reported as mean ± standard deviation or median (range), as appropriate. Inter-rater reliability was assessed using Cohen's kappa (κ). Associations between variables were evaluated using Spearman's rank correlation coefficient (rs). Group comparisons were performed using Mann–Whitney U and Kruskal–Wallis tests. Given the limited sample size (n = 24), a parsimonious regression strategy was adopted. CRAAP score, video duration, and log10-transformed view count were selected a priori because they represented three conceptually distinct domains: information literacy, content depth, and popularity. Other quality instruments (GQS, JAMA, DISCERN, and PEMAT) were evaluated descriptively and through correlation analyses but were not entered into the multivariable model to avoid overfitting and redundancy among closely related quality measures. Because only three videos were classified as "Safe and Useful," logistic regression analyses were not performed. Statistical significance was defined as $p < 0.05$.

2. Results

The final sample consisted of 24 videos, which together accumulated 935,427 views. Inter-observer reliability was high ($\kappa = 0.88$, 95% CI: 0.81–0.95), indicating almost perfect agreement between evaluators. Shapiro–Wilk testing demonstrated non-normal distributions for GQS ($p < 0.001$), JAMA ($p < 0.001$), mDISCERN ($p = 0.010$), PEMAT-U ($p = 0.009$), PEMAT-A ($p = 0.003$), and CSI ($p < 0.001$). CRAAP ($p = 0.096$), NFSQS ($p = 0.175$), and HKGE ($p = 0.428$) were normally distributed. Because several key variables were non-normally distributed, Spearman rank correlation was used for all correlation analyses. Descriptive statistics across all instruments are summarized in Table 1.

Table 1. Descriptive Statistics of All Informatics Instruments (n = 24)

Instrument	Max Score	Mean ± SD	Median	Range	SW p-value	Distribution
GQS	/5	2.71 ± 0.69	3.00	1–4	<0.001	Non-normal
JAMA	/4	1.67 ± 0.48	2.00	1–2	<0.001	Non-normal
mDISCERN	/5	2.54 ± 0.98	3.00	1–4	0.010	Non-normal
CRAAP	/15	6.50 ± 2.15	6.00	2–10	0.096	Normal
NFSQS	/24	8.38 ± 5.58	7.50	1–20	0.175	Normal
PEMAT-U (%)	100%	65.97 ± 24.32	66.67	33–100	0.009	Non-normal
PEMAT-A (%)	100%	67.50 ± 29.38	60.00	20–100	0.003	Non-normal
Final HKGE	/100	38.83 ± 18.34	36.08	10–77	0.428	Normal
CSI	Ratio	0.68 ± 0.54	0.56	0.12–2.19	<0.001	Non-normal

Safety_Flag distribution is illustrated in Figure 1. The most prevalent classification was "Not Differentiate Newborn" (n = 15; 62.5%), followed by "Mostly Safe but Incomplete" (n = 6; 25.0%) and "Safe and Useful" (n = 3; 12.5%). Combined, 21 of 24 videos (87.5%) failed to provide complete, neonatal-specific fever guidance.

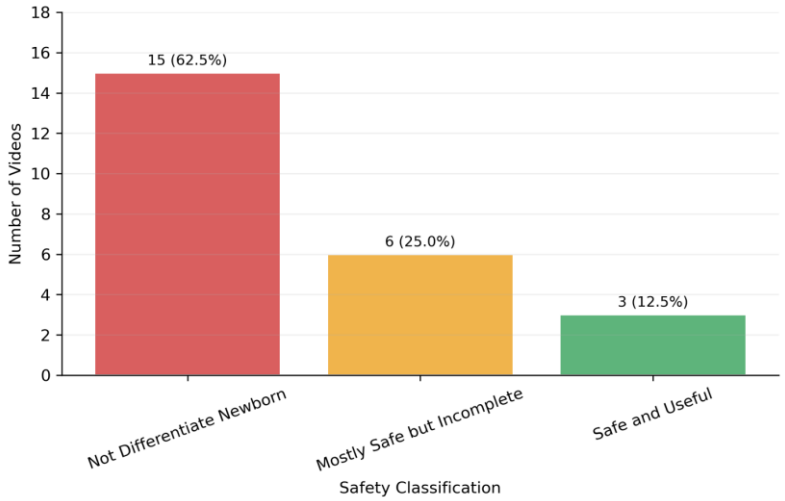


Figure 1. Safety classification distribution of neonatal fever management videos (n = 24).

Videos were classified into three clinically defined safety categories: "Not Differentiate Newborn," "Mostly Safe but Incomplete," and "Safe and Useful." The distribution demonstrates that most videos failed to provide complete neonatal-specific fever guidance.

The three "Safe and Useful" videos were all pediatrician-led. The highest-scoring video (NFSQS = 20, GQS = 4, HKGE = 76.67) had accumulated only 56 views, a second (NFSQS = 18, GQS = 4, HKGE = 70.33) had accumulated 595,618 views, and a third (NFSQS = 18, GQS = 3, HKGE = 67.67) had accumulated 102 views. Among "Not Differentiate Newborn" videos, FeverDef ≤ 1 was present in 14 of 15 cases, confirming that failure to specify neonatal fever definition and rectal thermometry was the single most prevalent clinical safety gap in the dataset.

Table 2. Full Score and Safety Classification Matrix (n = 24)

ID	Primary Source Category	Views	GQS	JAMA	DISCERN	CRAAP	NFSQS	PU%	PA%	HKGE	CSI	Safety Flag	E-GRADE
V1	Ped. Individual Expert Ch.	595,618	4	2	4	9	18	100	100	70.33	0.61	Safe and Useful	B
V6	Academic Ped.	75,000	3	2	3	9	12	67	60	52.67	0.54	Mostly Safe but Incomplete	C
V7	AHP Nurse	77,000	3	2	2	5	8	67	80	36.00	0.37	Not Differentiate Newborn	C
V10	AHP Nurse	1,900	3	2	3	5	10	83	100	42.33	0.65	Not Differentiate Newborn	D
V15	Ped. Individual Expert	102	3	2	4	8	18	100	100	67.67	1.68	Safe and Useful	B
V42	Ped. Individual Expert	56	4	2	4	10	20	83	100	76.67	2.19	Safe and Useful	B
V49	Ped. Individual Expert	26	3	2	4	8	10	100	60	47.67	1.68	Mostly Safe but Incomplete	C
V55	Ped. Individual Expert	56	2	2	3	6	13	33	100	49.83	1.43	Mostly Safe but Incomplete	C
V66	Ped. Individual Expert	52,000	3	2	3	9	13	83	80	55.17	0.58	Mostly Safe but Incomplete	C
V70	Ped. Individual Expert	4,400	3	2	3	10	11	67	60	51.50	0.71	Mostly Safe but Incomplete	C
V2	AHP Midwife	9,200	3	2	2	9	13	100	100	53.83	0.68	Mostly Safe but Incomplete	D
V11	Ped. Corporate	16,115	3	1	1	8	6	67	60	32.33	0.38	Not Differentiate Newborn	D
V34	Indep. Healthcare.	2,948	3	1	2	5	6	83	100	29.67	0.43	Not Differentiate Newborn	C
V12	Ped. Individual Expert	52,350	3	2	2	7	7	83	80	36.17	0.38	Not Differentiate Newborn	D
V3	Ped. Corporate	4,991	2	2	2	6	1	33	60	18.50	0.25	Not Differentiate Newborn	C
V16	AHP Nurse	15,975	2	1	1	2	1	83	100	10.50	0.12	Not Differentiate Newborn	D
V21	AHP Nurse	3,063	2	1	2	5	1	50	20	15.83	0.23	Not Differentiate Newborn	C
V25	Ped. Individual Expert	12,816	2	1	2	4	3	50	60	19.50	0.24	Not Differentiate Newborn	C

V26	Nurse Corporate Ch.	5,929	2	1	3	5	4	50	60	24.67	0.33	Not Differentiate Newborn	D
V27	Ped. Individual Expert	2,825	1	1	1	4	4	67	20	19.33	0.28	Not Differentiate Newborn	D
V28	Layperson / Parent	45	3	1	3	5	7	33	40	33.50	1.01	Not Differentiate Newborn	D
V40	Ped. Individual Expert	360	3	2	3	6	6	33	40	33.67	0.66	Not Differentiate Newborn	C
V46	Ped. Individual Expert	552	2	2	1	5	1	33	20	15.83	0.29	Not Differentiate Newborn	D
V47	Ped. Individual Expert	2,100	3	2	3	6	8	33	20	38.67	0.58	Not Differentiate Newborn	C

DIS = mDISCERN; PU = PEMAT Understandability; PA = PEMAT Actionability. Green = safe/high performers; red = safety concerns. PEMAT calculated from verified sub-items for all 24 videos.

Spearman correlation analysis demonstrated several significant associations among the evaluated metrics (Table 3; Figure 2). CRAAP information literacy showed strong positive correlations with NFSQS ($r_s = +0.762$, $p < 0.001$) and HKGE ($r_s = +0.828$, $p < 0.001$), indicating that videos with higher source quality, transparency, and informational credibility also demonstrated greater adherence to neonatal fever management guidelines. Similarly, GQS was positively associated with NFSQS ($r_s = +0.735$, $p < 0.001$), CRAAP ($r_s = +0.699$, $p < 0.001$), and HKGE ($r_s = +0.773$, $p < 0.001$). mDISCERN also correlated significantly with NFSQS ($r_s = +0.741$, $p < 0.001$), suggesting that more reliable and transparent health information sources provided more clinically appropriate neonatal guidance.

NFSQS demonstrated a moderate positive association with PEMAT-A ($r_s = +0.564$, $p = 0.004$), indicating that videos with greater guideline adherence were also more actionable for caregivers. A significant negative correlation was observed between view count and CSI ($r_s = −0.521$, $p = 0.009$; Figure 3), suggesting that highly viewed videos tended to demonstrate lower exposure-adjusted clinical safety. No significant association was identified between view count and NFSQS directly ($r_s = −0.063$, $p = 0.772$).

Table 3. Key Spearman Correlation Coefficients

Variable Pair	rs	p-value
CRAAP vs NFSQS	+0.762	<0.001
CRAAP vs HKGE	+0.828	<0.001
GQS vs NFSQS	+0.735	<0.001
GQS vs HKGE	+0.773	<0.001
mDISCERN vs NFSQS	+0.741	<0.001
NFSQS vs PEMAT-A	+0.564	0.004
HKGE vs CSI	+0.817	<0.001
Views vs CSI	-0.521	0.009

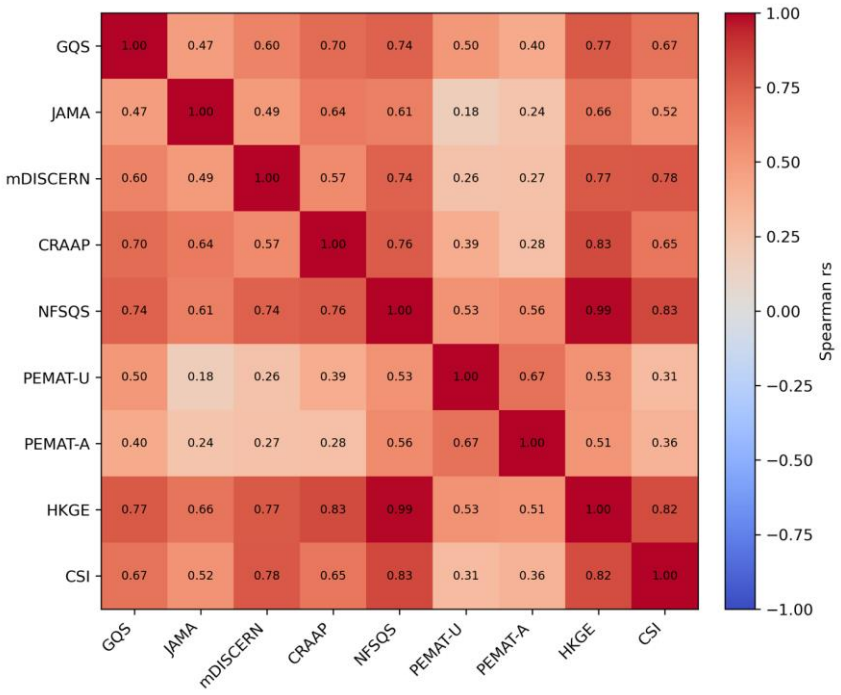


Figure 2. *Spearman correlation heatmap of video quality, safety, health-literacy, and visibility metrics.*

The heatmap displays Spearman rank correlation coefficients among GQS, JAMA, mDISCERN, CRAAP, NFSQS, PEMAT-U, PEMAT-A, HKGE, and CSI. Warmer colors indicate positive correlations, and cooler colors indicate negative correlations. This figure visually complements the key pairwise correlations reported in Table 3.

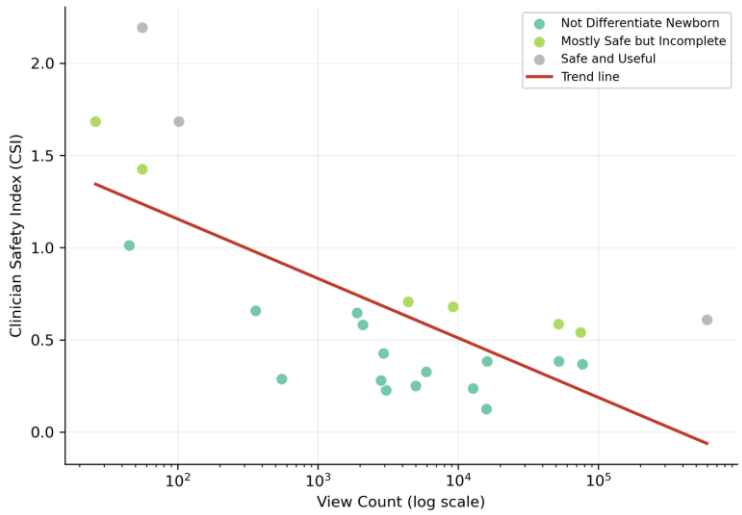


Figure 3. *Relationship between view count and Clinician Safety Index (CSI), stratified by safety classification. Each point represents one video. The x-axis displays view count on a logarithmic scale, and the y-axis displays CSI. Point color indicates Safety Flag classification. The fitted trend line demonstrates an inverse relationship between audience reach and exposure-adjusted clinical safety, indicating that higher visibility did not correspond to safer neonatal fever guidance.*

Grouping by primary content source suggested variation in NFSQS scores across presenter types. Pediatrician-led and academic pediatric specialist videos generally achieved higher NFSQS scores than non-physician or parent-led videos, although interpretation should be cautious because several categories contained only one or two videos and formal subgroup inference was not performed (Table 2).

E-GRADE clinical evidence classification, adapted from Kang et al.'s framework, was assigned by the two evaluating neonatologists based on overall evidence quality, consistency with contemporary neonatal fever recommendations, and clinical framing(15). No video achieved Grade A status. Grade B included 3 videos (12.5%), all of which were classified as "Safe and Useful" and achieved high NFSQS scores (mean 18.67 ± 1.15; range 18–20). Grade C included 12 videos (50.0%) with a mean NFSQS of 7.67 ± 4.36 (range 1–13), comprising both "Not Differentiate Newborn" (n = 7) and "Mostly Safe but Incomplete" (n = 5) classifications. Grade D included 9 videos (37.5%) with a mean NFSQS of 5.89 ± 3.95 (range 1–13), predominantly "Not Differentiate Newborn" videos (n = 8), with one "Mostly Safe but Incomplete" case. These findings are summarized in Table 4.

Table 4. NFSQS Distribution by E-GRADE Clinical Evidence Level (n = 24)

E-GRADE	n (%)	NFSQS Mean ± SD	NFSQS Range	Predominant Safety Classification
A	0 (0.0%)	N/A	N/A	N/A
B	3 (12.5%)	18.67 ± 1.15	18–20	Safe and Useful (n = 3)
C	12 (50.0%)	7.67 ± 4.36	1–13	Not Differentiate Newborn (n = 7); Mostly Safe but Incomplete (n = 5)
D	9 (37.5%)	5.89 ± 3.95	1–13	Not Differentiate Newborn (n = 8); Mostly Safe but Incomplete (n = 1)

Safety classifications within each E-GRADE level are summarized in Table 4. Kruskal–Wallis analysis demonstrated a statistically significant difference in NFSQS scores across E-GRADE levels (H = 8.352, p = 0.015). Post-hoc pairwise Mann–Whitney U tests confirmed that Grade B videos scored significantly higher than Grade C (U = 36.0, p = 0.011) and Grade D (U = 27.0, p = 0.016). No significant difference was found between Grades C and D (U = 67.5, p = 0.352), indicating that, despite differing E-GRADE assignments, both groups corresponded predominantly to the lower NFSQS ranges associated with incomplete or absent neonatal-specific guidance ("Mostly Safe but Incomplete" and "Not Differentiate Newborn" Safety_Flag categories), and their NFSQS distributions were statistically indistinguishable. Spearman correlation between E-GRADE ordinal rank and NFSQS yielded r_s = +0.501 (p = 0.013), confirming a moderate positive association between evidence grade and guideline adherence score.

Mean video duration was 5.67 minutes (median 1.73 min; range 0.38–51.36 min), reflecting a highly right-skewed distribution dominated by shorter videos. Duration showed a strong positive Spearman correlation with NFSQS (r_s = +0.725, p < 0.001) and with HKGE (r_s = +0.712, p < 0.001).

Short-form videos (duration <1 minute, n = 8) achieved significantly lower NFSQS scores than conventional-length videos (≥1 minute, n = 16; 3.88 ± 3.64 vs 10.62 ± 5.03; Mann–Whitney U = 17.0, p = 0.004). Kruskal–Wallis analysis across duration quartiles also demonstrated a significant difference in NFSQS (H = 12.611, p = 0.006), with scores increasing from Q1 to Q4 (Table 5). These findings suggest that shorter video duration was associated with lower neonatal guideline adherence, likely reflecting limitations in the amount of clinical detail that can be communicated within very brief formats.

Table 5. NFSQS by Video Duration Quartile Group

Quartile	Duration Range (min)	n	NFSQS Mean ± SD
Q1	0–0.91	6	2.83 ± 3.60
Q2	0.91–1.73	6	7.50 ± 3.45
Q3	1.73–4.57	6	8.83 ± 3.31
Q4	> 4.57	6	14.33 ± 5.16

To identify factors independently associated with neonatal guideline adherence, a multivariable linear regression model was constructed with NFSQS as the dependent variable and video duration, log10-transformed view count, and CRAAP score as explanatory variables. Multicollinearity diagnostics demonstrated negligible collinearity among predictors (all VIF values <2), supporting the stability of the regression estimates.

The overall model explained 72.1% of the variance in NFSQS ($R^2 = 0.721$). CRAAP score demonstrated the strongest independent association with NFSQS ($\beta = 1.770$, $p < 0.001$), indicating that videos with greater source credibility and information quality had higher neonatal guideline adherence scores. Video duration was also independently associated with NFSQS ($\beta = 0.182$ per minute, $p = 0.010$), whereas log10-transformed view count was not significantly associated with NFSQS after adjustment for other variables ($\beta = -0.296$, $p = 0.615$).

These findings suggest that information quality and content depth, rather than popularity, are the primary determinants of clinically appropriate neonatal fever guidance on YouTube.

Table 6. Multivariable Linear Regression Predicting NFSQS

Variable	β	95% CI	p
CRAAP	1.770	1.105–2.435	<0.001
Duration (min)	0.182	0.048–0.316	0.010
Log10(Views)	−0.296	−1.505–0.913	0.615

Model $R^2 = 0.721$.

4. Discussion

Neonatal fever represents one of the few pediatric emergencies in which a single symptom may trigger immediate hospital evaluation. In this context, we found that most Turkish-language YouTube videos failed to distinguish neonatal fever from general pediatric fever advice. The headline finding, that 62.5% of evaluated videos (15/24) were classified as "Not Differentiate Newborn" while only 12.5% (3/24) were classified as "Safe and Useful," reflects a major gap in the translation of clinical evidence into parental digital education.

The most prevalent driver of the "Not Differentiate Newborn" classification was the absence of neonatal-specific fever definition and rectal thermometry guidance. A rectal temperature ≥38.0°C is the recommended diagnostic threshold for neonatal fever, and rectal measurement remains the reference method for neonatal temperature assessment (9,10). Yet 14 of 15 videos in the "Not Differentiate Newborn" category had Fever Definition sub-scores ≤1, indicating either no fever definition or reliance on general, non-neonatal thresholds. Videos demonstrating axillary or forehead temperature measurement without clarifying their limitations in newborns may therefore contribute to delayed recognition of serious infection (7,8).

The E-GRADE findings provide additional insight into the quality of neonatal fever information available on YouTube. Importantly, no video

achieved Grade A classification despite the availability of well-established international neonatal fever guidelines. This finding suggests that the principal challenge is not a lack of evidence, but rather inadequate translation of existing evidence into caregiver-directed digital educational content.

The absence of Grade A videos is particularly noteworthy because neonatal fever management is supported by clear and widely disseminated clinical guidelines. Unlike many areas of medicine characterized by substantial uncertainty, the core safety messages surrounding neonatal fever are relatively straightforward: fever should be measured using rectal thermometry, a rectal temperature of ≥38.0°C constitutes a medical emergency, and prompt clinical evaluation is required. Therefore, the observed deficiencies likely reflect incomplete translation of established evidence into caregiver-facing educational content rather than a lack of available evidence. This interpretation is further supported by the predominance of Grade C and Grade D videos within our sample.

Beyond the absence of the neonatal rectal threshold, a clinically concerning content pattern was identified across multiple videos: attribution of neonatal fever to benign causes, most commonly dehydration, without the essential caveat that this distinction cannot be made safely at home. Current neonatal fever guidance is unambiguous: in an infant under 28 days of age, fever should be treated as a sepsis alert until proven otherwise through clinical and laboratory evaluation. This distinction is particularly important because the clinical appearance of a febrile neonate may be deceptively reassuring. Unlike older children, neonates with invasive bacterial infection may initially present with isolated fever and minimal accompanying symptoms. Consequently, contemporary neonatal fever pathways emphasize age-based risk stratification rather than symptom severity alone. Failure to communicate this principle may create a false sense of security among caregivers and contribute to delays in medical evaluation. Unlike many pediatric conditions in which home observation may be appropriate, neonatal fever pathways are intentionally designed to prioritize sensitivity over specificity because the consequences of missed invasive bacterial infection may be severe. Consequently, even well-appearing febrile neonates generally require prompt medical assessment.

Caregiver observation or non-rectal temperature measurement cannot exclude bacteremia or meningitis in this age group. Similarly, content implying that only higher fever values warrant urgent care is potentially misleading, because serious bacterial infection can occur across the fever spectrum in neonates. The clinically correct message, that any rectal temperature ≥38.0°C in a neonate requires urgent medical evaluation, was consistently absent from the lowest-scoring videos.

While information quality and caregiver actionability appeared to align, a different pattern emerged when content visibility was examined. The relationship between popularity and clinical quality represents one of the most important findings of this study. Similar observations have been reported in evaluations of neonatal sepsis content on YouTube, where highly reliable physician-generated videos did not necessarily achieve greater audience reach (19). Although view count was negatively correlated with CSI ($r_s = -0.521$, $p = 0.009$), multivariable regression demonstrated that view count was not independently associated with NFSQS after adjustment for information literacy and video duration ($\beta = -0.296$, $p = 0.615$). In contrast, CRAAP score and video duration remained significant independent predictors of guideline adherence. In our dataset, highly viewed videos were not necessarily the videos that provided the safest neonatal fever advice. The highest-scoring video in the dataset (NFSQS = 20) had accumulated only 56 views, whereas several lower-quality videos achieved substantially greater audience reach. Collectively, these results support a quality–visibility mismatch in neonatal health information dissemination.

The association between information literacy and clinical quality was particularly noteworthy. CRAAP score demonstrated both a strong correlation with NFSQS ($r_s = +0.762$, $p < 0.001$) and the strongest independent association in multivariable analysis ($\beta = 1.770$, $p < 0.001$). This suggests that source transparency, authority, currency, and purpose are closely linked to delivery of guideline-concordant neonatal fever information. Although CRAAP was not originally designed as a health-specific assessment tool, its strong association with NFSQS suggests that broader indicators of source credibility may serve as useful surrogate markers of clinical information quality in digital neonatal education. Interestingly, the relationship between CRAAP and NFSQS suggests that information literacy and clinical quality may be more closely linked than previously appreciated. Kang et al. described a “credibility–evidence gap,” whereby professionally presented medical videos

frequently appeared trustworthy despite limited evidentiary support. Our findings extend this observation to neonatal health communication by demonstrating that videos with higher source credibility were also more likely to provide guideline-concordant neonatal fever recommendations. Together, these findings suggest that credibility, evidence quality, and clinical safety represent complementary but distinct dimensions of digital health information. This finding is consistent with the rationale proposed by McNamee et al., who combined disease-specific content evaluation with broader information-quality assessment in YouTube-based medical education research (17).

Although information-literacy tools such as CRAAP and reliability instruments such as mDISCERN provide important insights into source quality, they do not directly address whether caregivers can understand and act upon the information presented. For neonatal fever education, where delayed action may have important clinical consequences, actionability represents a critical component of effective health communication. The NFSQS versus PEMAT-A correlation (r_s = +0.564, p = 0.004) reveals an additional dimension: videos that adhered more closely to clinical guidelines were also more actionable for caregivers. Importantly, the highest-scoring videos were also among the most actionable, suggesting that accurate neonatal fever guidance can be delivered in a caregiver-friendly manner. A common concern in caregiver education is that clinically rigorous information may be difficult to translate into practical actions. In contrast, our findings suggest that neonatal guideline adherence and caregiver actionability are not competing objectives. The highest-quality videos successfully communicated both the urgency of neonatal fever, and the specific actions caregivers should take, indicating that evidence-based messaging can remain understandable and actionable without oversimplification. These findings address an important gap identified in previous reviews of YouTube health-information research, which have largely focused on content quality and reliability while providing limited assessment of caregiver actionability (20,21). Our results suggest that evidence-based neonatal fever education can remain both clinically rigorous and practically actionable.

Given that caregivers frequently seek online information before contacting healthcare services, omission of neonatal-specific safety messages may influence initial decision-making at a critical point in the care pathway.

Exploratory analysis according to primary content source suggested substantial heterogeneity within professional categories. Some videos presented by healthcare professionals achieved very low NFSQS scores despite otherwise acceptable production quality. This finding highlights that professional affiliation alone does not guarantee neonatal guideline adherence. Rather, inclusion of neonatal-specific concepts, particularly fever definition, emergency referral criteria, and medication safety, appears to be more important than presenter credentials alone.

The duration analysis revealed an additional structural dimension. The YouTube Shorts format (<60 seconds) was associated with "Not Differentiate Newborn" classification in 7 of 8 Shorts (87.5%), consistent with the hypothesis that format constraints may limit inclusion of the informational components required for safe neonatal fever guidance. The strong positive correlation between video duration and HKGE (r_s = +0.712, p < 0.001) may partly reflect this format-driven ceiling effect. However, unsafe classification was not exclusive to Shorts: 8 of 16 videos lasting ≥1 minute (50.0%) were also classified as "Not Differentiate Newborn," indicating that adequate duration alone does not ensure neonatal-specific content. Platform-level interventions focused only on short-form videos would therefore address only part of the safety problem; broader content-level correction by producers, clinicians, and professional organizations is also needed.

The duration findings further support the importance of content depth in neonatal health communication. Video duration remained independently associated with NFSQS in multivariable analysis (β = 0.182, p = 0.010), suggesting that the observed relationship is not solely attributable to publisher type or information literacy. Neonatal fever management requires communication of multiple interrelated concepts, including fever definition, temperature measurement technique, red flags, emergency referral, and medication safety. Extremely short videos may face inherent challenges in delivering the minimum informational content required for guideline-concordant caregiver education.

Importantly, the presence of several high-performing videos demonstrates that accurate neonatal fever education is achievable within the YouTube environment. These videos simultaneously achieved high NFSQS, CRAAP, and PEMAT-A scores, indicating that evidence-based neonatal recommendations can be communicated clearly, accurately, and in a caregiver-centered manner. The

challenge therefore appears to be dissemination and visibility rather than feasibility of content creation.

The practical implications of these findings extend beyond digital media evaluation. Neonatologists, pediatricians, and professional organizations increasingly serve as indirect educators through online platforms. Given that many caregivers seek information before contacting healthcare services, digital educational materials should consistently reinforce three core messages: neonatal fever should be measured using rectal thermometry, a rectal temperature of ≥38.0°C constitutes a medical emergency, and urgent clinical assessment is required regardless of the infant's apparent well-being. Standardization of these safety messages may improve the consistency of online neonatal fever education and reduce potentially harmful delays in care-seeking.

4.1. Limitations

Several limitations should be considered. First, this was a cross-sectional audit performed at a single time point, whereas YouTube content, engagement metrics, and search rankings change dynamically over time. Second, although the study included a broader audit than many comparable neonatal YouTube analyses, the final sample size was modest, limiting statistical power for subgroup analyses and precluding more complex multivariable modeling. Third, although the NFSQS was previously developed and internally validated by our group, external validation in independent datasets remains limited and warrants further investigation. In contrast, HKGE and CSI are exploratory composite metrics that have not undergone formal validation and should therefore be interpreted as hypothesis-generating measures rather than validated clinical instruments. Fourth, the study was restricted to Turkish-language content and may not reflect the quality of neonatal fever information in other linguistic or cultural settings. Fifth, viewer behavior, watch time, recommendation pathways, and algorithmic promotion mechanisms could not be directly assessed, limiting causal interpretation of the observed quality–visibility mismatch. Finally, although inter-rater agreement was high, some degree of subjective interpretation is unavoidable when applying content-based assessment instruments, and video content may have changed after the audit period.

5. Conclusion

Turkish-language YouTube videos addressing neonatal fever management frequently fail to communicate the emergency nature of neonatal fever, with 62.5% of evaluated videos classified as "Not Differentiate Newborn" and only 12.5% meeting criteria for "Safe and Useful" content. The most common deficiencies were failure to specify the rectal fever threshold (≥38.0°C), inadequate differentiation of neonatal fever from general pediatric fever, and omission of urgent referral recommendations.

Information literacy quality, measured by the CRAAP framework, emerged as the strongest independent predictor of neonatal guideline adherence, while video duration was also positively associated with clinical quality. In contrast, viewer traffic was not independently associated with guideline adherence, indicating that popularity is a poor surrogate for clinical accuracy. The predominance of incomplete short-form content and the observed mismatch between visibility and clinical quality raise important concerns regarding digital health communication for high-risk neonatal conditions.

Given the vulnerability of febrile neonates, pediatric societies and public health organizations should actively support dissemination of evidence-based neonatal fever education on digital platforms. Future efforts should focus on integrating information-literacy indicators, clinical guideline adherence, and health-literacy principles into content certification and recommendation systems to improve the safety of online health information available to caregivers.

REFERENCES

1. Madathil KC, Rivera-Rodriguez AJ, Greenstein JS, Gramopadhye AK. Healthcare information on YouTube: A systematic review. Health Informatics Journal 2014; 21: 173-94.
2. Osman W, Mohamed F, Elhassan M, Shoufan A. Is YouTube a reliable source of health-related information? A systematic review. BMC medical education 2022; 22: 382.
3. Borges do Nascimento IJ, Beatriz Pizarro A, Almeida J, Azzopardi-Muscat N, André Gonçalves M, Björklund M, et al. Infodemics and health misinformation: a systematic review of reviews.

Bulletin of the World Health Organization 2022; 100: 544-61.
4. Briand SC, Cinelli M, Nguyen T, Lewis R, Prybylski D, Valensise CM, et al. Infodemics: A new challenge for public health. Cell 2021; 184: 6010-4.
5. Suarez-Lledo V, Alvarez-Galvez J. Prevalence of Health Misinformation on Social Media: Systematic Review. J Med Internet Res 2021; 23: e17187.
6. Kollmann TR, Kampmann B, Mazmanian SK, Marchant A, Levy O. Protecting the newborn and young infant from infectious diseases: lessons from immune ontogeny. Immunity 2017; 46: 350-63.
7. Kim F, Polin RA, Hooven TA. Neonatal sepsis. bmj 2020; 371.
8. Celik IH, Hanna M, Canpolat FE, Mohan P. Diagnosis of neonatal sepsis: the past, present and future. Pediatric Research 2022; 91: 337-50.
9. Leazer RC. Evaluation and Management of Young Febrile Infants: An Overview of the New AAP Guideline. Pediatrics In Review 2023; 44: 127-38.
10. Pantell RH, Roberts KB, Adams WG, Dreyer BP, Kuppermann N, O'Leary ST, et al. Evaluation and Management of Well-Appearing Febrile Infants 8 to 60 Days Old. Pediatrics 2021; 148.
11. Paul SP, Khattak H, Kini PK, Heaton PA, Goel N. NICE guideline review: neonatal infection: antibiotics for prevention and treatment (NG195). Archives of disease in childhood - Education & practice edition 2022; 107: 292-7.
12. Bernard A, Langille M, Hughes S, Rose C, Leddin D, van Zanten SV. A Systematic Review of Patient Inflammatory Bowel Disease Information Resources on the World Wide Web. Official journal of the American College of Gastroenterology | ACG 2007; 102: 2070-7.
13. Silberg WM, Lundberg GD, Musacchio RA. Assessing, Controlling, and Assuring the Quality of Medical Information on the Internet: Caveant Lector et Viewor—Let the Reader and Viewer Beware. JAMA 1997; 277: 1244-5.
14. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. Journal of Epidemiology and Community Health 1999; 53: 105-11.
15. Kang E, Lee H, Choi J, Ju H. The Quality of Evidence of and Engagement With Video Medical Claims. JAMA Network Open 2026; 9.
16. Blakeslee S. The CRAAP test. Loex Quarterly 2004; 31: 4.
17. McNamee LA, Shakartzi H, Wasser TE, Li Y, Kim C-S. Assessing the accuracy and quality of YouTube videos on early pregnancy loss. Heliyon 2022; 8.
18. Shoemaker SJ, Wolf MS, Brach C. Development of the Patient Education Materials Assessment Tool (PEMAT): a new measure of understandability and actionability for print and audiovisual patient information. Patient education and counseling 2014; 96: 395-403.
19. Hakyemez Toptan H, Kizildemir A. Quality and Reliability Analysis of YouTube Videos Related to Neonatal Sepsis. Cureus 2023.
20. Drozd B, Couvillon E, Suarez A. Medical YouTube Videos and Methods of Evaluation: Literature Review. JMIR Medical Education 2018; 4.
21. Gabarron E, Fernandez-Luque L, Armayones M, Lau AYS. Identifying Measures Used for Assessing Quality of YouTube Videos with Patient Health Information: A Review of Current Literature. Interactive Journal of Medical Research 2013; 2.

©Copyright 2026 by Osmangazi Tıp Dergisi - Available online at tip.ogu.edu.tr©Telif Hakkı 2026ESOGÜ Tıp Fakültesi - Makale metnine dergipark.org.tr/otdweb sayfasından ulaşılabilir.