

Family Caregivers as the Dominant Voice in Chinese Liver Cancer Discourse on Xiaohongshu: A BERTopic-Based Computational Text Analysis

Hongyu Jiang¹, Liyao Xiao^{2*}

¹ School of Basic Medicine, Guizhou Medical University, Guiyang, China

² School of Arts and Tourism, Guizhou Communications Polytechnic University, Guiyang, China

Accepted

2026-04-20

Keywords

liver cancer; hepatitis B; family caregivers; health information seeking; Xiaohongshu; topic modeling; BERTopic

Corresponding Author

Liyao Xiao

Copyright 2026 by author(s)

This work is licensed under the CC BY 4.0



<https://doi.org/10.70693/itphss.v3i2.462>

Abstract

China accounts for approximately 50% of global new liver cancer cases, predominantly driven by hepatitis B virus (HBV)-related hepatocellular carcinoma. The disease presents a distinctive socio-medical profile characterized by a prolonged etiological chain (chronic hepatitis B—cirrhosis—liver cancer), late-stage diagnosis, and a cultural norm of family-first disclosure—whereby physicians typically inform family members before patients themselves. These features position family caregivers, rather than patients, as the primary information seekers and communicators in online health discourse. Drawing on the Comprehensive Model of Information Seeking (CMIS) and Communication Privacy Management (CPM) theory, this study analyzes 1,857 liver cancer-related user comments collected from Xiaohongshu (Little Red Book)—a Chinese lifestyle-oriented social media platform with over 300 million monthly active users—between December 2024 and February 2026. BERTopic, a neural topic modeling framework suited to short Chinese social media texts, was employed for inductive topic extraction. The analysis yielded 15 interpretable topics, organized into five thematic categories: clinical diagnosis and treatment (34.8%), emotional support and caregiving (26.3%), upstream chronic disease management (14.2%), symptom and tumor assessment (12.5%), and online health consultation (9.6%). Notably, hepatitis B and cirrhosis medication management emerged as the second-largest topic, directly mirroring the HBV-dominant etiological chain characteristic of Chinese liver cancer. Family-role terms ("father," "mother," "husband") and Buddhist prayer language appeared as high-salience keywords across multiple topics, underscoring the caregiver-driven and culturally embedded nature of the discourse. The study proposes the concept of "proxy information seeker" as a theoretical extension of CMIS, reveals privacy boundary tensions arising from caregivers' routine public disclosure of patient health information under China's non-disclosure culture, and provides empirical grounding for caregiver-oriented platform governance and digital health literacy interventions.

1. Introduction

Liver cancer is among the most burdensome malignancies in China. According to Global Cancer

Statistics 2020, China accounts for approximately 50% of all new liver cancer cases worldwide (Sung et al., 2021). The Global Burden of Disease Study 2021 confirms that hepatitis B virus (HBV) remains the leading etiological factor for liver cancer in China (Cao et al., 2025), while a 2024 systematic report estimates that HBV-related liver diseases cause approximately 160,000 deaths annually in the country, with liver cancer exhibiting the highest mortality rate among all HBV-related hepatic conditions (Yan et al., 2025). These figures represent not merely epidemiological abstractions but lived realities for countless families: when one member is diagnosed, the entire household is thrust into the roles of caregiver, decision-maker, and information seeker.

The socio-medical landscape of liver cancer in China differs substantially from that in Western countries. Whereas the etiological profile in Europe and North America has shifted toward hepatitis C, alcohol-related liver disease, and non-alcoholic steatohepatitis (NASH), China's liver cancer remains predominantly HBV-driven (Cao et al., 2025), giving rise to a prolonged disease trajectory of "chronic hepatitis B carriage—cirrhosis—hepatocellular carcinoma." Many patients have undergone decades of chronic disease management—antiviral medication, regular monitoring, liver function testing—before receiving a cancer diagnosis, accumulating substantial informational capital along the way. Simultaneously, China's liver cancer early detection rate remains low, with a considerable proportion of cases identified only at intermediate or advanced stages. Consequently, the moment of diagnosis and the imperative for immediate treatment decision-making effectively coincide—and the burden of information gathering, hospital comparison, and treatment planning falls not on patients but on their families. Wang et al. (2011) documented that therapeutic non-disclosure remains widespread in Chinese oncology practice, with family members often serving as the primary gatekeepers of diagnostic information. Liu et al. (2018), in a prospective study of 124 cancer patient–family dyads at a tertiary hospital, reported that 62.1% of patients were unaware of their diagnosis prior to first chemotherapy. This practice of "family-first disclosure and surrogate decision-making," rooted in Confucian familial ethics (Fan & Li, 2004), structurally positions health information seeking in the Chinese liver cancer context as a family activity rather than a patient activity.

Under this disease–family structure, where does liver cancer discourse migrate? Historically concentrated on disease-specific patient forums, hospital health education columns, and offline peer exchanges, such discourse has increasingly shifted to lifestyle-oriented general-purpose social media platforms. Xiaohongshu (Little Red Book), a Chinese platform with over 300 million monthly active users originally known for beauty, travel, and consumer lifestyle content, now increasingly hosts illness narratives: young women documenting their fathers' liver cancer diagnoses, wives searching for targeted therapy options for their husbands, adult children posting photographs of pathology reports and requesting interpretation from strangers (Ji et al., 2025). This phenomenon raises questions that are simultaneously about health communication and disease-specific: What thematic structure characterizes illness discourse on a platform not designed for medical exchange? Who is speaking? The Comprehensive Model of Information Seeking (CMIS; Johnson & Meischke, 1993), which presupposes an individualized patient–medium relationship, requires reexamination in light of the triple characteristics of Chinese liver cancer.

This study analyzes 1,857 liver cancer-related comments from Xiaohongshu using BERTopic for inductive topic modeling, interpreted through CMIS and Communication Privacy Management theory (CPM; Petronio, 2002), to address three research questions:

RQ1: Does the thematic structure of liver cancer discourse on Xiaohongshu reflect the disease-specific trajectory of "HBV—cirrhosis—late-stage—caregiving"?

RQ2: Who are the dominant communicative actors—patients, family caregivers, or bystanders—and how do their discursive features reflect the cultural structure of Chinese liver cancer care?

RQ3: How do Xiaohongshu's platform affordances interact with the care needs specific to liver cancer, and what communicative functions emerge from this interaction?

The study contributes at three levels: empirically, it provides the first systematic characterization of liver cancer discourse on a Chinese lifestyle platform; theoretically, it proposes the concept of "proxy information seeker" as an extension of CMIS grounded in the Chinese liver cancer context of family-first disclosure and surrogate decision-making; practically, it offers evidence-based implications for platform governance, caregiver-oriented digital health literacy interventions, and clinicians seeking to understand how family members' information behaviors shape the questions they bring into the consultation room.

2. Background and Literature Review

2.1 Liver Cancer in China: A Distinctly Familial Disease

Understanding liver cancer discourse on Xiaohongshu requires first grasping three defining characteristics of the disease in China. First, HBV dominance in etiological composition. China harbors approximately one-third of the world's chronic HBV carriers, and HBV-related liver diseases account for approximately 160,000 deaths annually, with liver cancer ranking highest among them (Yan et al., 2025). This means that many liver cancer patients bring to the discourse a pre-existing informational foundation built over years of hepatitis B management—antiviral medication, regular monitoring, and liver function testing. Second, a diagnostic profile skewed toward late presentation. Because HBV-related liver cancer is often asymptomatic in its early stages and screening coverage remains uneven, a substantial proportion of Chinese liver cancer cases are diagnosed at intermediate or advanced stages (Cao et al., 2025). Late diagnosis fundamentally reshapes caregivers' informational needs: patients are frequently unable to conduct sustained information work themselves, and the tasks of comparing hospitals, evaluating surgical options, and weighing interventional versus conservative approaches fall to family members. Third, the cultural norm of family-first disclosure. In Chinese clinical practice, a cancer diagnosis is typically communicated first to family members, who then decide whether and how to inform the patient (Fan & Li, 2004). Liu et al. (2018) confirmed this empirically: 62.1% of cancer patients were unaware of their own diagnosis before first chemotherapy.

Together, these three characteristics constitute the structural context underlying Chinese liver cancer discourse: the lengthy HBV chronic disease history provides caregivers with pre-existing informational capital and platform familiarity; late-stage diagnosis shifts the informational agent from patient to family; and the non-disclosure culture further centralizes caregivers in the information-action nexus.

2.2 Online Health Information Seeking and Caregivers as Information Actors

Online health information seeking (OHIS) research has accumulated substantial evidence over two decades, with robust core findings: users increasingly treat the internet as a primary health information source, typically in parallel with rather than as a substitute for clinical consultation (Zhao et al., 2022). However, the traditional OHIS framework, centered on an individual patient oriented toward an identifiable medium, shows increasing limitations in the context of participatory social media, where information seeking, information sharing, and emotional support are inextricably intertwined (Chou et al., 2013).

Health communication research has extensively documented the informational and emotional labor undertaken by family caregivers of seriously ill patients. Gage-Bouchard et al. (2017), in a study of parents of children with acute lymphoblastic leukemia communicating on personal Facebook pages, demonstrated that caregivers frequently serve as information brokers—translating between

clinicians and patients, conducting searches on behalf of patients, and facilitating or leading treatment decisions. Yet such proxy information actors remain largely absent from patient-centered OHIS models. In the Chinese cultural context, filial responsibility norms and family-centered medical decision-making place adult children and spouses at the center of cancer care communication (Fan & Li, 2004), with the clinical practice of family-first disclosure further reinforcing this position.

2.3 Xiaohongshu and Chinese Digital Health Communication

Xiaohongshu has evolved since 2013 from a cross-border shopping community into one of China's most influential lifestyle platforms. Its distinctive affordance configuration—image-and-text short posts, discovery-oriented algorithmic recommendation, a predominantly female user base, and an aesthetic emphasis on authentic everyday experience—differentiates it from Weibo (microblog-style, public-sphere orientation) and Douyin (short-video, entertainment orientation). Recent research has documented health-related discourse on the platform, including Ji et al.'s (2025) content analysis of e-cigarette narratives, suggesting that this platform, despite not being designed for medical purposes, has become a significant venue for health discourse.

Existing computational analyses of Chinese social media health discourse have focused primarily on Weibo, largely employing LDA-type topic models applied to relatively longer texts (Zhang et al., 2021). Three research gaps can be identified: no study has systematically characterized liver cancer discourse on a lifestyle platform; the extremely short texts typical of Xiaohongshu pose acute challenges for bag-of-words topic models, while BERTopic (Egger & Yu, 2022)—better suited to such data—has rarely been applied in Chinese health contexts; and existing work has infrequently situated Chinese social media health discourse within established health communication theoretical frameworks, leaving findings largely descriptive.

3. Theoretical Framework

3.1 Comprehensive Model of Information Seeking (CMIS)

CMIS, developed by Johnson and Meischke (1993) through a synthesis of the health belief model, uses and gratifications theory, and a model of media exposure and appraisal, encompasses three variable classes: antecedents (demographics, direct experience, salience, beliefs), information carrier factors (channel characteristics and utility), and information-seeking actions. The model's core assumption—a single information seeker oriented toward an identifiable medium—is challenged by the Xiaohongshu liver cancer case in two respects. First, information seeking and sharing are deeply intertwined in the comment section: users simultaneously ask questions (seeking) and narrate experiences (serving as carriers). Second, the dominance of caregiver discourse suggests that the information seeker is frequently a proxy whose "direct experience" is vicarious—derived from witnessing a family member's illness rather than from personal embodiment.

In the Chinese liver cancer context, this "proxy seeker" configuration is not marginal but modal: late-stage diagnosis renders patients unable to sustain long-term information work, the family-first disclosure culture positions caregivers as the default primary information agents, and the time-critical nature of treatment decisions (surgical windows, TACE timing, targeted therapy selection) imposes additional pressure on carrier utility assessments.

3.2 Communication Privacy Management (CPM)

CPM (Petronio, 2002) conceptualizes private information as individually owned, with owners establishing rules governing when and to whom disclosure occurs. Once information is disclosed, recipients become co-owners with obligations to honor the original owner's privacy rules; failures

in this coordination—boundary turbulence—constitute the theory's central analytic focus. Liver cancer discourse on Xiaohongshu presents a configuration of particular theoretical interest under CPM: family caregivers routinely disclose relatives' diagnoses, symptoms, and treatment trajectories on a public platform—often accompanied by photographs of pathology reports—and a substantial proportion of these disclosures occur under circumstances where patients themselves have not been fully informed of their diagnosis (Wang et al., 2011; Liu et al., 2018). Under CPM logic, patients are the original owners of this diagnostic information, yet caregivers as co-owners unilaterally make the disclosure decision—constituting a form of boundary turbulence that is widespread yet routinely legitimized in everyday practice.

Together, the two frameworks characterize Xiaohongshu liver cancer discourse as a simultaneous information-seeking scene and privacy-boundary scene. CMIS interprets what users seek and through what carrier logic; CPM interprets what is disclosed and by whom. The topic structure revealed by BERTopic is subsequently read through both lenses.

4. Data and Methods

4.1 Corpus

User comments were collected from Xiaohongshu between December 2024 and February 2026 using two seed queries: "liver cancer" (肝癌, gānái) and "liver lesion" (肝病變, gān bìngbiàn). The raw corpus of 2,308 entries was reduced to 2,086 after deduplication and cleaning, and to 1,857 valid documents after Chinese word segmentation and minimum-length filtering (Table 1). The post-preprocessing median document length of 3 words confirmed the short-text character of the corpus (Figure 1).

Table 1. Corpus statistics.

Source	Keywords	Period	# Documents
Xiaohongshu (liver cancer)	肝癌	2024.12–2026.01	1,346
Xiaohongshu (liver lesion)	肝病變	2025.01–2026.02	962
Total (raw)	—	—	2,308
After deduplication & cleaning	—	—	2,086
After segmentation (≥ 2 tokens)	—	—	1,857

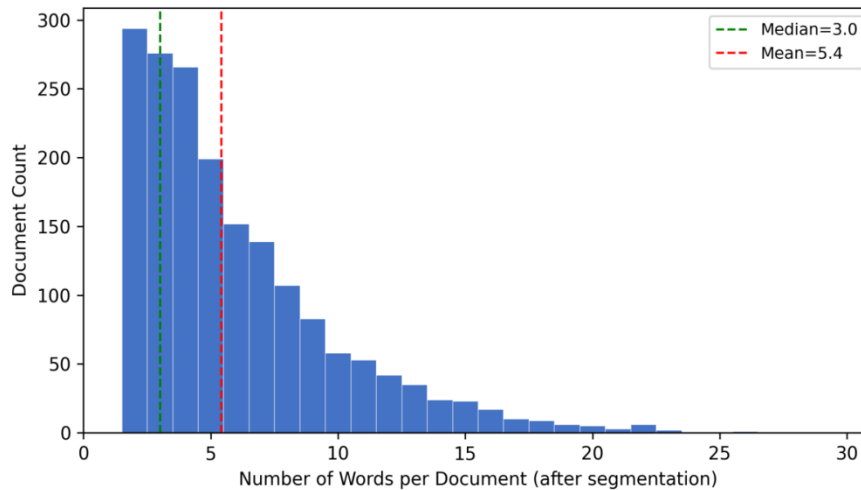


Figure 1. Distribution of document lengths after preprocessing.

4.2 Research Ethics

Although Xiaohongshu comments are publicly accessible, the health-sensitive nature of the content warrants additional ethical care. We report computational results only at the aggregate level; all illustrative comments presented in this paper are paraphrased rather than quoted verbatim to prevent reverse search identification; all usernames, geographic markers, and hospital names have been removed. Given the pervasive practice of caregivers disclosing patient information in liver cancer discourse, any detail that could enable identification of a patient (as distinct from the posting caregiver) was deleted or rephrased.

4.3 BERTopic Modeling

Given the extreme brevity of Xiaohongshu comments (median: 3 words), traditional bag-of-words topic models struggle to extract sufficient word co-occurrence information from such data. This study adopts BERTopic (Egger & Yu, 2022), whose modular architecture—document embedding, dimensionality reduction, density-based clustering, and class-based TF-IDF keyword extraction—effectively addresses short-text scenarios.

The processing pipeline proceeded as follows. First, the corpus was segmented using jieba in precise mode, augmented with a custom dictionary of 85 liver cancer-specific medical terms (e.g., cirrhosis, targeted therapy, alpha-fetoprotein, interventional therapy). Part-of-speech filtering retained content words (nouns, verbs, adjectives), and a combined stopword list of 131 entries removed functional words and domain noise. The resulting document-term matrix was embedded into a 100-dimensional space via TF-IDF with truncated SVD, reduced to 5 dimensions via UMAP (McInnes et al., 2018), and clustered by HDBSCAN (McInnes et al., 2017), yielding 39 initial clusters and 307 noise points (16.5%). Ward-linkage agglomerative clustering merged initial clusters to $K=15$ topics, with representative keywords for each topic extracted via class-based TF-IDF.

The topic number $K=15$ was determined through a systematic scan of topic coherence (C_v ; Röder et al., 2015) across $K \in [3, 20]$, with C_v peaking at $K=15$ (0.4233; Figure 2). Topic labels were inductively assigned by the first author based on each topic's c-TF-IDF keywords and 20 highest-probability exemplar documents, independently reviewed by the second author, with disagreements resolved through discussion.

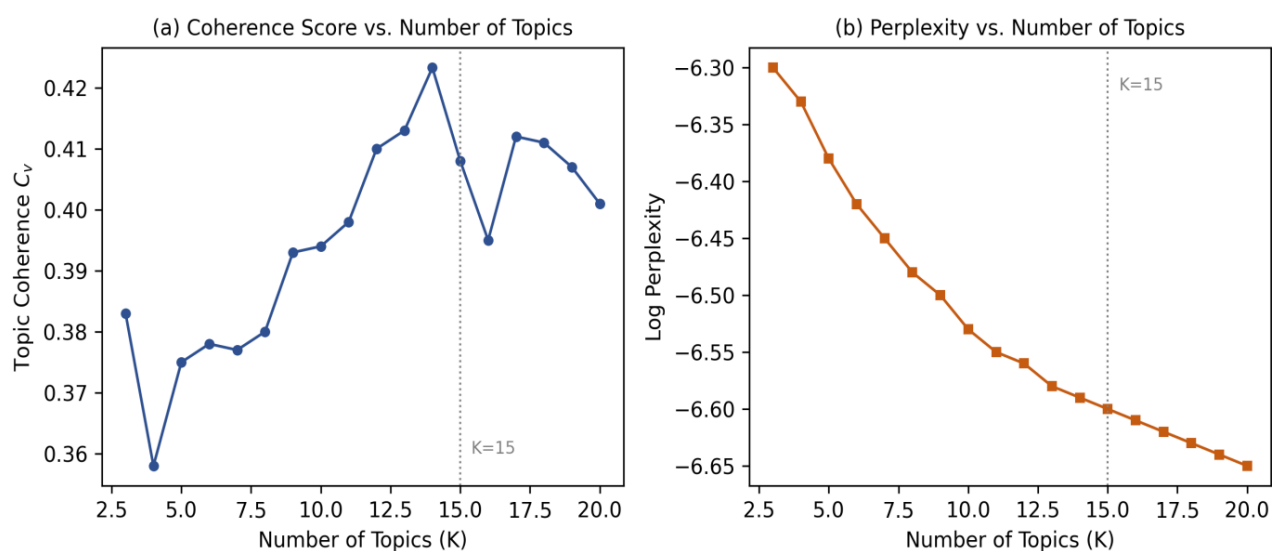


Figure 2. Topic coherence (C_v) and log perplexity across K values (optimal $K=15$).

5. Conclusion and Suggestion

BERTopic extracted 15 interpretable topics from the 1,857-document corpus, organized into five thematic categories (Table 2; Figure 3): clinical diagnosis and treatment (34.8%), emotional support and caregiving (26.3%), upstream chronic disease management (14.2%), symptom and tumor assessment (12.5%), and online health consultation (9.6%). The following sections integrate empirical findings with theoretical discussion, organized around the three research questions.

Table 2. Fifteen topics extracted by BERTopic.

ID	Topic Label	Top Keywords (c-TF-IDF)	n	Category
T1	Late-stage liver cancer symptoms & complications	ascites, symptoms, jaundice, severe, abdomen	162	Symptom/tumor
T2	Daily liver disease management & recovery	follow-up, diet, sharing, anti-cancer, regular	96	Chronic disease
T3	Liver cancer diagnosis & surgical treatment	liver cancer, surgery, diagnosis, metastasis, father, targeted	200	Clinical
T4	Treatment plans & hospital selection	treatment, hospital, doctor, plan, Hunan, Guangzhou	109	Clinical
T5	Fatty liver & liver function abnormalities	fatty liver, transaminase, drinking, staying up late	125	Chronic disease
T6	Family emotional expression & help-seeking	father, liver cancer, miss, distressed, departure	88	Emotional
T7	Hepatitis B / cirrhosis medication management	hepatitis B, cirrhosis, bilirubin, antiviral, liver function	184	Chronic disease
T8	Prayers & spiritual support	peace, health, bless, family, Amitabha	21	Emotional
T9	Pathology report interpretation requests	trouble, report, pathology, biopsy, physical exam	57	Consultation

ID	Topic Label	Top Keywords (c-TF-IDF)	n	Category
T10	Treatment decision-making & psychology	cancer, suffering, family, conservative, meaning	160	Emotional
T11	Spousal patient care & symptoms	husband, diagnosed, pain relief, appetite, pain	139	Emotional
T12	Online medical consultation	help, take a look, doctor, severe, mother	33	Consultation
T13	Liver function testing & diagnosis	examination, results, antibody, autoimmune, phosphatase	47	Clinical
T14	Suspected symptoms & health anxiety	cannot, feeling, somewhat, liver cancer, doctor	59	Consultation
T15	Tumor benign/malignant differentiation	mother, tumor, cyst, benign, worried	70	Symptom/tumor

Fig. 3 Distribution of Five Thematic Categories (N=1,857)

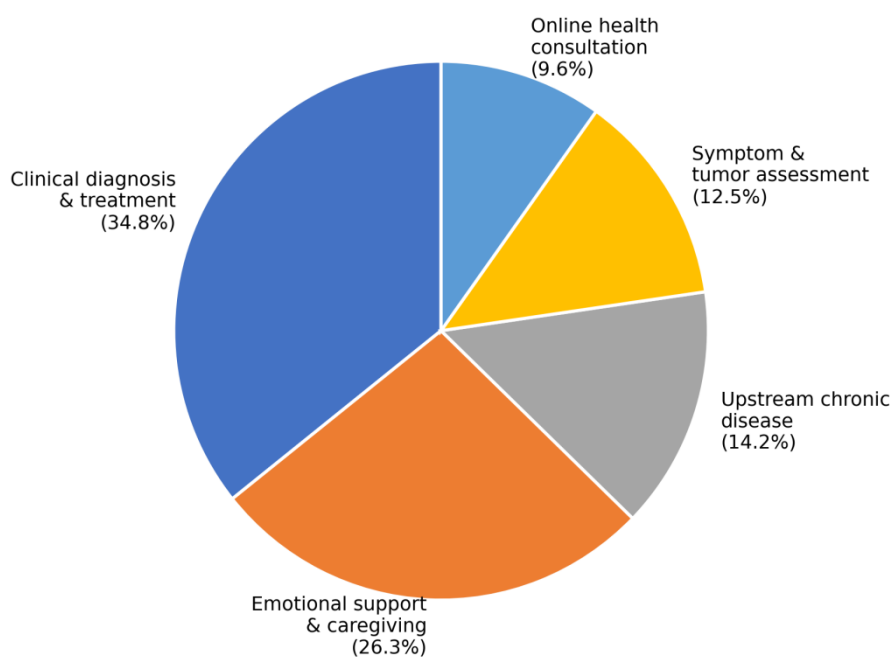


Figure 3. Distribution of five thematic categories (N=1,857).

5.1 Family Caregivers as the Dominant Voice and Theoretical Extension of CMIS (RQ2)

The most striking feature of the corpus is that liver cancer discourse on Xiaohongshu is not predominantly patient-authored. Family-role terms rank prominently in the corpus-wide TF-IDF ranking—"father" at rank 11, "mother" at rank 12, "husband" at rank 14 (Figure 4)—and appear as high-salience keywords across multiple topics: T3 and T6 involve fathers, T12 and T15 involve mothers, and T11 involves husbands. By contrast, first-person patient expressions are comparatively rare; only T14 (suspected symptoms and health anxiety) appears to carry a first-person perspective, though its content is dominated by diffuse health anxiety rather than actionable information seeking.

Fig. 4 Top-20 Keywords by TF-IDF Score

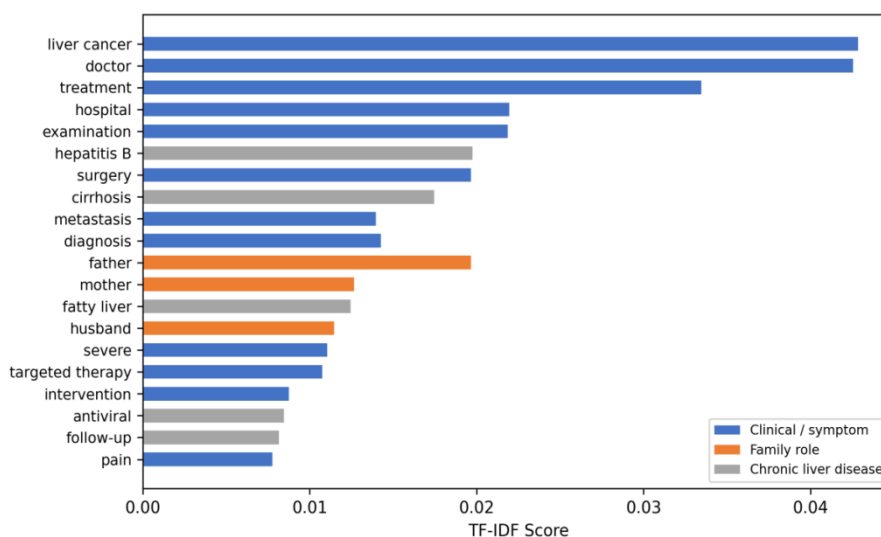


Figure 4. Top-20 corpus keywords by TF-IDF score, colored by category.

This discursive pattern is highly congruent with the structural characteristics of Chinese liver cancer: high rates of late-stage diagnosis, family-first disclosure norms, and a young caregiving demographic—which, when superimposed on Xiaohongshu's user profile (predominantly female, 18–35 age range), makes "young adult children documenting and questioning an elder's liver cancer on a lifestyle platform" the default discursive scenario. The clinical evidence reported by Liu et al. (2018)—62% patient unawareness before chemotherapy—provides macro-level corroboration for this pattern at the disease–culture interface.

Situated within CMIS, the caregiver-dominant voice pattern indicates that the model's "direct experience" antecedent requires expansion to accommodate what we term "vicarious direct experience"—the sustained, intimate witnessing and management of a family member's serious illness. This expansion is not a marginal correction: in the Chinese liver cancer context, the proxy information seeker is not an anomalous case within CMIS but rather the modal one. Late-stage diagnosis renders patients unable to sustain long-term information work; family-first disclosure norms position caregivers as the sole possessors of complete diagnostic information; and the time-critical nature of liver cancer treatment decisions (surgical windows, TACE timing, targeted therapy selection) imposes distinctive temporal pressure on carrier utility assessments. Accordingly, we propose that in severe illness contexts characterized by "late-stage predominance—family-first disclosure—time-critical decisions," CMIS should incorporate the proxy information seeker as a theoretically distinct actor type, possessing its own antecedent profile (vicarious experience, caregiver salience, temporal pressure) and carrier evaluation logic (emotional accessibility, platform–demographic fit).

5.2 Thematic Structure along the Liver Cancer Disease Trajectory (RQ1)

The four clinical topics (T3, T4, T7, T13), comprising 540 documents (34.8%), are distributed not randomly but along the disease trajectory characteristic of Chinese liver cancer: T7 (hepatitis B/cirrhosis medication management, n=184) corresponds to the prolonged HBV etiological history; T3 (diagnosis and surgical treatment, n=200) corresponds to the diagnostic shock point; T4 (treatment plans and hospital selection) corresponds to the decision-making phase; and T13 (liver function testing and diagnosis) corresponds to laboratory result interpretation during the diagnostic workup.

T7 warrants particular attention. As the second-largest topic in the corpus (surpassed only by the

diagnosis topic T3), it directly mirrors the HBV-dominant etiological chain of Chinese liver cancer. Hepatitis B carriers and cirrhosis patients, having accumulated years of antiviral medication use and follow-up monitoring before their liver cancer diagnosis, generate a stable sub-discourse on the platform. The topic's core keywords—"hepatitis B—cirrhosis—bilirubin—antiviral—liver function—bleeding—hospitalization"—map onto the full vocabulary of HBV-related liver disease across its stages: carriage, cirrhosis, and decompensation (Yan et al., 2025). In other words, liver cancer discourse on Xiaohongshu discusses not only the cancer itself but also the years or decades that precede it.

Additionally, T9 (pathology report interpretation requests) and T12 (online medical consultation) exhibit a noteworthy genre: users photograph pathology reports, CT scans, or biopsy results and request community interpretation. This represents a quasi-professional appropriation of a lifestyle platform—one that neither aligns with the platform's design intent nor has received adequate governance response. From a CMIS carrier utility perspective, caregivers are substituting a lifestyle platform for content that would traditionally be obtained through professional channels. For a disease with time-critical treatment decisions such as liver cancer, the risks inherent in such channel substitution are considerable.

5.3 Emotional Support, Cultural Coping, and Privacy Boundaries (RQ3)

The four emotional topics (T6, T8, T10, T11), comprising 408 documents (26.3%), constitute the second-largest thematic category. T6 centers on adult children processing a parent's illness ("father, miss, distressed, departure"); T11 focuses on spousal caregiving during active symptom periods; T10 captures decision-making ambivalence at the palliative–active treatment boundary ("conservative, suffering, meaning"); and T8, though the smallest topic (n=21), is the most culturally distinctive—its core vocabulary ("peace," "bless," "Amitabha," "family," "all beings") constitutes an independent Buddhist–folk prayer discourse that would be absent from Western health forum corpora, representing a distinctly Chinese cultural coping repertoire.

From a CPM perspective, caregivers' extensive posting on Xiaohongshu constitutes a routinized form of proxy disclosure—systematically revealing relatives' diagnoses, symptoms, and treatment trajectories to anonymous publics, with a substantial proportion occurring under circumstances where patients have not been fully informed of their diagnosis (Wang et al., 2011; Liu et al., 2018). Patients are the original information owners; caregivers as co-owners unilaterally make public disclosure decisions, constituting latent boundary turbulence. Our data suggest that caregivers implicitly manage this tension in two ways: first, through patient-level anonymization—medical details are freely disclosed, but patients' names, workplaces, and photographs are almost never present; second, through narrative framing that constructs the caregiver's "right to speak"—emphasizing their role as the burden-bearer, decision-maker, and family representative, thereby establishing an implicit privacy rule: so long as the patient remains individually unidentifiable, caregivers are entitled to disclose illness information in exchange for informational and emotional support.

At a broader level, Xiaohongshu was not designed for health communication yet has become a significant venue for liver cancer discourse, owing to its affordance configuration: personal narrative conventions, an image-and-text interface accommodating medical report photographs, high tolerance for emotional expression, and algorithmic feeds that intermix lifestyle and illness content. These features that "make it useful" for caregivers are precisely those that render it "normatively ambiguous" as a health information channel. For a disease with time-critical treatment decisions such as liver cancer, erroneous "report interpretations" or misguided "interventional versus conservative" advice on the platform may directly affect patient outcomes.

5.4 Upstream Chronic Disease, Health Anxiety, and Symptom Assessment

T2 (daily liver disease management) and T5 (fatty liver and liver function abnormalities), totaling 221 documents (14.2%), form the upstream chronic disease category. T5's vocabulary (fatty liver, transaminase, drinking, staying up late) vividly represents a lifestyle-risk register consonant with Xiaohongshu's overall platform ethos, suggesting that many hepatitis B carriers may have been consuming liver health and lifestyle content on the platform long before their liver cancer diagnosis—making "remaining on the same platform to seek information" a low-migration-cost choice. This permeability between "health" as a wellness topic and "health" as a serious illness topic constitutes a distinctive feature of Xiaohongshu's liver cancer discourse.

T9, T12, and T14, totaling 149 documents (9.6%), constitute the online consultation and anxiety zone. T14 (suspected symptoms and health anxiety) is particularly noteworthy: users who notice a symptom and fear it may indicate liver cancer seek reassurance or guidance on whether to seek clinical care. Furthermore, T1 (late-stage symptoms and complications, n=162) and T15 (tumor benign/malignant differentiation, n=70) together form a symptom assessment category (12.5%). T1's vocabulary (ascites, jaundice) reflects the late-stage-predominant clinical reality of Chinese liver cancer (Cao et al., 2025); T15 presents the typical scenario of family members (often adult children) describing a parent's newly detected hepatic lesion and asking whether it is likely malignant—a situation clinically linked to T7, as hepatitis B carriers discovering hepatic nodules represent a critical window for early liver cancer detection.

5.5 Practical Implications

These findings carry specific implications for three stakeholder groups. For platform operators: tiered governance mechanisms should be established for serious illness discourse, such as automatically surfacing authoritative medical information in response to pathology report interpretation posts, or routing such posts to verified physician response interfaces. Critically, platforms should avoid crude "content sanitization" approaches that indiscriminately suppress health topics, as the data indicate that emotional content serves genuine supportive functions within caregiver coping mechanisms. For public health and health communication practitioners: digital health literacy interventions targeting Chinese liver cancer should treat family caregivers as the primary audience rather than defaulting to a patient-as-information-seeker model; content should specifically address the HBV-to-liver-cancer early detection window, basic report interpretation literacy, and information evaluation in late-stage treatment decisions. For clinicians: family members often arrive at the consultation having already completed a round of "parallel consultation" on Xiaohongshu; the questions, doubts, and prior assumptions they bring may originate from platform content—understanding this reality enables physicians to respond more effectively to caregivers' actual cognitive states.

6. Conclusion and Suggestion

This study provides the first systematic characterization of liver cancer discourse on Xiaohongshu, a Chinese lifestyle-oriented social media platform. Through BERTopic topic modeling of 1,857 user comments, interpreted via CMIS and CPM, we present a portrait deeply coupled with the structural characteristics of Chinese liver cancer: discourse is caregiver-dominated; content is organized along the "HBV—cirrhosis—diagnosis—treatment—late-stage—caregiving" disease trajectory; functionality simultaneously serves informational and emotional purposes; and the discourse is sustained by the convergence of the platform's lifestyle affordances with the demographic profile of liver cancer caregivers. Theoretically, the study proposes the "proxy information seeker" as a conceptual extension of CMIS and employs CPM to illuminate the widespread yet under-discussed practice of caregivers disclosing patient health information on

public platforms in China.

Several limitations warrant acknowledgment. First, data derive from a single platform across a 14-month window; cross-platform (Weibo, Douyin, Zhihu) and longitudinal extensions would better isolate platform effects from temporal effects. Second, document embedding employed TF-IDF–SVD rather than a Chinese medical domain pretrained transformer; future replication with Chinese-BERT-wwm or domain-specific models may refine topic boundaries. Third, the inference that speakers are predominantly caregivers rests on corpus-level linguistic markers rather than verified user identities, and would benefit from survey or interview validation. Fourth, the study is interpretive rather than causal. Future research may extend in four directions: cross-platform comparative analysis to isolate platform affordance effects; longitudinal analysis tracking discourse evolution; mixed-methods designs combining computational text analysis with caregiver interviews; and intervention studies exploring whether embedding authoritative information within caregiver-driven discussion threads can be accomplished without disrupting their emotional support functions.

Funding

This work was supported by the Guizhou Provincial Basic Research Program (Natural Science) ([2022]354).

References

- Cao, M. D., Xia, C. F., Zhou, J. H., Teng, Y., Li, Q. R., Tan, N. P., Wang, J. C., Zuo, T. T., Li, T. Y., Zheng, Y. J., & Chen, W. Q. (2025). Global epidemiology of liver cancer burden due to hepatitis B virus: A comprehensive estimate based on Global Burden of Disease Study 2021. *Chinese Medical Journal*, 138, 3457–3466.
- Chou, W. Y. S., Prestin, A., Lyons, C., & Wen, K. Y. (2013). Web 2.0 for health promotion: Reviewing the current evidence. *American Journal of Public Health*, 103(1), e9–e18.
- Egger, R., & Yu, J. (2022). A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify Twitter posts. *Frontiers in Sociology*, 7, 886498.
- Fan, R., & Li, B. (2004). Truth telling in medicine: The Confucian view. *Journal of Medicine and Philosophy*, 29(2), 179–193.
- Gage-Bouchard, E. A., LaValley, S., Mollica, M., & Beaupin, L. K. (2017). Communication and exchange of specialized health-related support among people with experiential similarity on Facebook. *Health Communication*, 32(10), 1233–1240.
- Ji, T., Liu, Z., Su, Z., et al. (2025). E-cigarette narratives of user-generated posts on Xiaohongshu in China: Content analysis. *Journal of Medical Internet Research*, 27, e71173.
- Johnson, J. D., & Meischke, H. (1993). A comprehensive model of cancer-related information seeking applied to magazines. *Human Communication Research*, 19(3), 343–367.
- Liu, Y., Yang, J., Huo, D., Fan, H., & Gao, Y. (2018). Disclosure of cancer diagnosis in China: The incidence, patients' situation, and different preferences between patients and their family members and related influence factors. *Cancer Management and Research*, 10, 2173–2181.
- McInnes, L., Healy, J., & Astels, S. (2017). hdbscan: Hierarchical density based clustering. *Journal of Open Source Software*, 2(11), 205.
- McInnes, L., Healy, J., Saul, N., & Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software*, 3(29), 861.
- Petronio, S. (2002). *Boundaries of Privacy: Dialectics of Disclosure*. State University of New York Press.

- Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. *Proceedings of the 8th ACM International Conference on Web Search and Data Mining*, 399–408.
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 71(3), 209–249.
- Wang, D. C., Guo, C. B., Peng, X., Su, Y. J., & Chen, F. (2011). Is therapeutic non-disclosure still possible? A study on the awareness of cancer diagnosis in China. *Supportive Care in Cancer*, 19(8), 1191–1195.
- Yan, R., Sun, M., Yang, H., Du, S., Sun, L., & Mao, Y. (2025). 2024 latest report on hepatitis B virus epidemiology in China: current status, changing trajectory, and challenges. *Hepatobiliary Surgery and Nutrition*, 14(1), 66-77.
- Zhang, Y., Chen, L., Wang, H., Luo, Y., & Hu, D. (2021). COVID-19 health information analysis on Weibo using text mining. *International Journal of Environmental Research and Public Health*, 18(3), 1238.
- Zhao, Y. C., Zhao, M., & Song, S. (2022). Online health information seeking behaviors among older adults: Systematic scoping review. *Journal of Medical Internet Research*, 24(2), e34790.