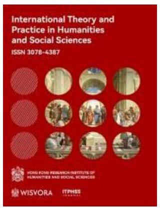
WISVORA

International Theory and Practice in Humanities and Social Sciences
2026 Volume3,Issue2 ISSN 3078-4387



Seeking Diagnosis or Seeking Support? A Thematic Analysis of Diagnostic Health Information Seeking in Liver Cancer–Related Content

Hongyu Jiang[1], 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

2026-04-19

Keywords

health information-seeking behaviour; diagnostic health information seeking; Xiaohongshu; liver cancer; BERTopic; platform medicalisation

Corresponding Author

Liyao Xiao

https://doi.org/10.70693/itphss.v3i2.463

Abstract

China carries approximately half of the global new burden of primary liver cancer, and the extended disease trajectory of "hepatitis B – cirrhosis – hepatocellular carcinoma" places patients and family caregivers in a state of sustained, stage-specific health information-seeking needs. With the popularisation of user-generated-content (UGC) platforms such as Xiaohongshu (also known as Rednote), a growing number of users bypass or supplement formal medical consultation by treating these platforms as informal "clinics". However, existing research has largely focused on post-diagnosis caregiver discourse and emotional support, while paying insufficient attention to the pre-diagnosis behaviour of users uploading imaging reports and soliciting lay diagnostic judgement—what we term diagnostic health information seeking (DHIS)—and to the quality of the replies they receive. Drawing on health information-seeking behaviour (HISB) models and information quality theory, this study integrates two independently scraped Xiaohongshu datasets ("liver cancer" and "liver lesion"), yielding 2,761 texts (2,218 question-oriented comments and 543 platform-side replies). A c-TF-IDF–based BERTopic-style pipeline combined with rule-based coding is applied to a three-track design. Three findings emerge. First, diagnostic seeking accounts for 31.5% of pre-diagnosis comments (liver-lesion context) but only 3.4% of post-diagnosis comments (liver-cancer context), a nearly tenfold gap ($\chi^2(1) = 333.84$, $p < .001$). Second, emotional seeking shows the reverse pattern (11.5% vs. 20.7%, $\chi^2(1) = 31.68$, $p < .001$). Third, among 543 question–reply dyads, 40.1% of diagnostic questions received no response; among those that were answered, only 13.6% received professional medical advice, while 15.2% received low-quality reassurance or deflection. We propose the concept of platform medicalisation and offer implications for platform governance, credential verification for content creators, and digital health-literacy interventions.

1.Introduction

Liver cancer has become a major global public-health challenge. According to GLOBOCAN 2022 data, liver cancer ranks sixth in global incidence and third in global cancer-related mortality, and China accounts for approximately 42% of new cases and 44% of deaths worldwide (Li et al., 2024; Diao et al., 2025). Unlike many other cancer sites, the epidemiology of liver cancer in China is characterised by an extended disease trajectory: of individuals infected with hepatitis B virus (HBV), approximately 15%–25% develop cirrhosis within 20–40 years, and 5%–10% of those with cirrhosis progress to hepatocellular carcinoma each year (McGlynn et al., 2021; Chen et al., 2025). This long causal pathway means that, long before a formal diagnosis, a large latent population must cope with abnormal physical-examination results, fluctuating liver-function indicators and equivocal imaging findings—a "prolonged period of uncertainty".

This period of uncertainty provides a strong motivational basis for information-seeking behaviour. Classical studies in health information-seeking behaviour (HISB) have established that individuals actively seek information to reduce uncertainty when they perceive a health threat and feel under-informed (Wilson, 1999; Longo, 2005; Lambert & Loiselle, 2007). Traditionally, the targets of such seeking have been clinicians, hospitals, and authoritative medical websites. Over the past decade, however, social-media platforms such as WeChat, Weibo, Zhihu and Xiaohongshu have become increasingly important supplementary channels for health information among Chinese users (Xiong et al., 2021; Zhang et al., 2022; Wang et al., 2021).

Xiaohongshu offers a particularly valuable case. Centred on a "UGC + algorithmic recommendation" mechanism and with more than 300 million monthly active users—predominantly young women—the platform has extended its content ecology from beauty and travel into parenting, medical, and health-related lifestyle topics (Huang & Miao, 2024). Because notes on Xiaohongshu are routinely image-plus-text, users can directly upload ultrasound reports, CT images and lab test screenshots, and make concrete diagnostic requests in the comment section directly to bloggers or to other users. This pattern differs fundamentally from "seeking general health knowledge"; it approximates informal remote consultation.

Yet, this emerging form of health information-seeking has not been systematically examined. Existing research on social-media health behaviours in the Chinese context has followed three broad lines: (1) the quality and misinformation of health content on WeChat public accounts and Weibo (Zhang et al., 2017; Chang et al., 2024); (2) information-seeking and caregiver-mediated disclosure among older adults and chronic-disease families (Wang et al., 2023); and (3) online social support in disease-specific contexts such as diabetes and mental health (Wu et al., 2018; Chen & Wang, 2021). These studies have built a necessary foundation, but three gaps remain.

First, conceptually, there is no dedicated construct for the pre-diagnostic behaviour of directly soliciting others' diagnostic judgement. It is neither general information seeking nor emotional support seeking, but rather a request for informal diagnosis; existing HISB taxonomies fail to capture it. Second, methodologically, most studies focus on the question side and pay little attention to the supply side (bloggers' and top commenters' replies), and even fewer examine question–reply dyads together. Third, thematically, there is no systematic comparison of how information seeking shifts across disease stages within a single health condition. Liver cancer and its pre-cursor state—"liver lesion"—straddle the pre-diagnosis and post-diagnosis nodes, offering a natural site for stage-wise comparison.

Against this backdrop, the present study integrates two separately scraped Xiaohongshu datasets ("liver cancer" and "liver lesion"), covering 2,218 question-oriented texts and 543 platform replies. Using a c-TF-IDF BERTopic-style pipeline (Grootendorst, 2022) combined with rule-based coding, we run three parallel tracks to address three research questions:

RQ1: What thematic categories of health information-seeking behaviour appear in liver-cancer–related content on Xiaohongshu?

RQ2: How does the distribution of diagnostic versus emotional seeking differ across disease stages (pre- vs. post-diagnosis)?

RQ3: What patterns of match or mismatch exist between user requests and platform replies?

The contributions of this study are threefold. First, we propose the construct of diagnostic health information seeking (DHIS), extending the HISB taxonomy to capture a previously under-theorised user behaviour in social-media contexts. Second, we propose the concept of platform medicalisation to describe the structural phenomenon by which social-media platforms take on informal medical-diagnostic functions. Third, we offer the first systematic empirical portrait of stage-wise differentiation and supply–demand mismatch along the full liver-cancer trajectory in the Chinese social-media ecosystem.

2. Literature Review and Theoretical Framework

2.1 Health Information-Seeking Behaviour (HISB): Theoretical Evolution

HISB is a core concern at the intersection of information science, health communication, and medical sociology. Wilson's (1999) macro model of information behaviour treats information seeking as an agentive activity aimed at reducing uncertainty in a specific context, and distinguishes four types of behaviour: active seeking, ongoing seeking, passive search, and passive attention. Longo (2005) extended the model into the health context and proposed an integrative HISB framework that emphasises two dimensions: instrumental versus affective needs, and source selection. Instrumental needs concern concrete medical decision support (e.g., whether to see a doctor, which treatment to choose), while affective needs concern comfort, belonging, and meaning-making.

Lambert and Loiselle's (2007) systematic review of cancer-context HISB argued that information seeking is not merely a cognitive act, but a process embedded in social networks and institutional structures. They further posited that HISB varies substantially across the illness trajectory: the diagnostic phase is dominated by instrumental needs, the treatment phase combines treatment-option and side-effect information, and the palliative phase is dominated by emotional support and existential meaning. This "stage-wise transition" hypothesis directly informs the present study.

In the social-media context, the scope and boundaries of HISB have shifted substantially. Drawing on a cross-sectional survey across four Chinese sites (N = 18,144), Xiong et al. (2021) show that online HISB in China has moved from a portal-dominated to a social-media-centred, channel-diversified landscape, with young and urban users showing the highest platform reliance and with the content sought expanding from general knowledge to concrete symptom interpretation and clinical decision support. Wang et al.'s (2021) meta-analysis further indicates that online HISB is positively associated with self-efficacy and health literacy, but is also linked to indiscriminate self-diagnosis and delayed care.

2.2 Diagnostic Health Information Seeking (DHIS): Toward a New Construct

The strand of literature most adjacent to our concern is research on online symptom checking, which examines users' self-diagnostic use of symptom checkers such as WebMD and Chunyu Doctor (Chambers et al., 2019). These tools provide rule- or ML-based standardised judgements, which differ fundamentally in nature from users asking unspecified others for a diagnostic judgement.

A second adjacent strand concerns paid online consultation. Wu et al. (2018), for example, examine users' intentions to use Chinese paid consultation platforms (e.g., Haodf), showing that doctor–patient interaction quality and physician credential signals substantially shape usage intention. Yet, platforms such as Haodf explicitly verify credentials and operate on paid consultation models, placing them within the formal telemedicine ecology. By contrast, "consultations" on Xiaohongshu or Weibo come with no credential verification mechanism; information suppliers may be physicians, nurses, or equally anxious lay users.

Building on this gap, we define diagnostic health information seeking (DHIS) as: the behaviour by which users actively solicit diagnostic interpretation of specific symptoms, imaging reports, or pathology results from others on informal, non-credential-guaranteed social-media platforms. DHIS has three defining features: (1) its content is concrete and directive, seeking judgements such as "what is it", "is it serious", and "do I need urgent care"; (2) the channel is informal, with no credentialing requirement for information providers; and (3) its boundaries are fluid, frequently overlapping with emotional support seeking and general information seeking in the same text. Distinguished from general HIS, DHIS focuses on verdicts about an individual's condition rather than knowledge acquisition; distinguished from emotional seeking, DHIS is more instrumental; distinguished from formal telemedicine, DHIS operates outside credentialing safeguards.

2.3 Information Quality Theory and Supply–Demand Mismatch

Wang and Strong's (1996) four-dimensional framework of information quality—intrinsic, contextual, representational, and accessibility—remains a classic analytic tool. Eysenbach et al. (2002) extended this framework to the health-information domain, proposing accuracy, completeness, timeliness and comprehensibility as the core evaluative dimensions for online health content. Sbaffi and Rowley's (2017) systematic review argues that users' credibility assessments rely heavily on heuristic signals such as source authority, specificity, and social cues (likes, comments, blogger identity markers).

Applied to DHIS, a central issue is supply–demand mismatch: when users articulate highly instrumental requests (e.g., asking for a CT report interpretation), does the supply side provide a matching high-quality response, or does it defer to emotional comfort, vague advice, or silence? This mismatch is accentuated on UGC platforms where commercial and social logics intersect—bloggers' core incentive is maintaining follower engagement rather than providing accurate medical advice, and top commenters are more inclined to express empathy than to render careful judgements (Lederman et al., 2014).

2.4 The Chinese Medical Context and the "Platform Medicalisation" Hypothesis

Several structural features of China's healthcare system have created fertile ground for the "medicalisation" of social-media platforms. First, outpatient encounters are short: consultation time in rural and tertiary Chinese hospitals averages only a few minutes (Wang et al., 2022), leaving patients without adequate interpretation of their diagnostic reports. Second, physician–patient trust remains fragile, with a long history of recurrent disputes and information asymmetry fuelling dissatisfaction with formal channels (Li & Khan, 2023). Third, primary-care capacity is uneven, with many users doubting local physicians and seeking second opinions online. Fourth, the platforms themselves offer accessibility and immediacy—whether for night-time anxiety or out-of-town consultation, Xiaohongshu provides a near-instant response channel.

These factors together give rise to what we term platform medicalisation: a structural pattern of use in which users systematically treat social-media platforms as informal medical consultation channels and integrate platform responses into personal health decision-making. The concept draws on the medical-sociology concept of medicalisation (Conrad, 2007) but shifts the analytic lens from "everyday life being brought into medical discourse" to "non-medical platforms being brought into diagnostic function"—extending medicalisation theory into the era of digital platforms.

3. Methods

3.1 Data Collection and Pre-processing

The study integrates two independently scraped datasets that together span the key nodes of the liver-cancer trajectory and exhibit strong semantic continuity. Specifically, in the liver-lesion dataset, the term "liver cancer" occurs 252 times, and 88 comments involve direct discussion of a liver-cancer diagnosis—supporting the continuity of the two datasets. Data were scraped using Houyi Caiji on 8 February 2026 and, after de-duplication and cleaning of advertisements and non-informative characters, 2,761 valid texts were retained (see Table 1).

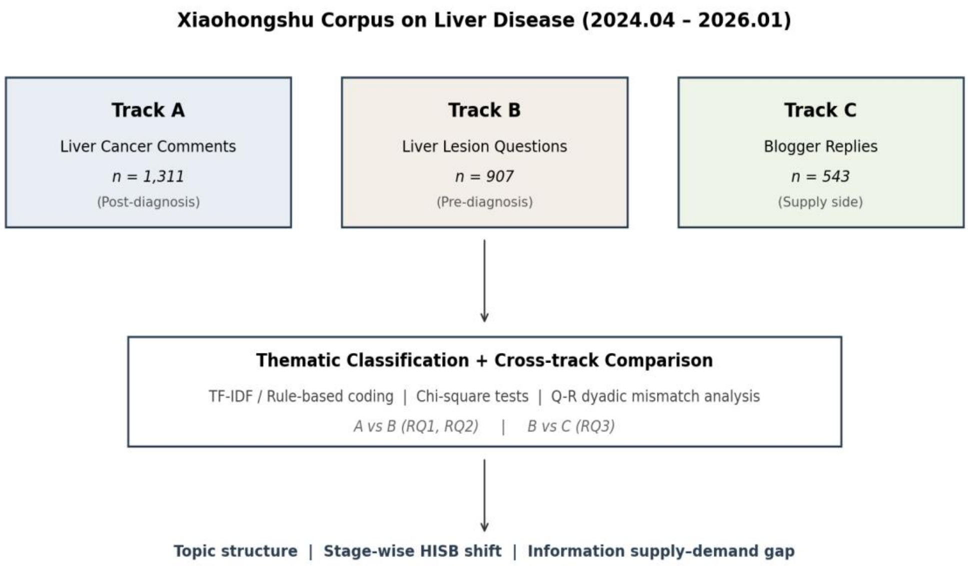


Figure 1. Research design and three-track analytical framework.

Table 1. Corpus Description and Analytical Tracks

Dimension	Track A	Track B	Track C
	Liver Cancer Comments	*Liver Lesion Questions*	*Blogger Replies*
Keyword	"Liver cancer"	"Liver lesion"	(same as Track B)
Text type	Comments on posts	Question comments	Replies to questions
Posts (n)	19	112	(same as Track B)
Valid texts	1,311	907	543
Avg. length (char)	30.4	23.7	17.7
Time span	2024.04 – 2026.01	2024.05 – 2026.01	2024.05 – 2026.01
Stage location	Post-diagnosis	Pre-diagnosis	Information supply side

3.2 Three-track Parallel Topic Modelling

Given that our research questions require two-sided comparison across questions and replies, we adopt a three-track parallel modelling design. Track A draws on the 1,311 comments in the liver-cancer dataset, capturing post-diagnosis information-seeking themes; Track B uses the 907 question-oriented texts from the liver-lesion dataset, capturing pre-diagnosis diagnostic seeking; Track C uses the 543 blogger replies, capturing supply-side response patterns.

The core steps of the pipeline follow the BERTopic framework (Grootendorst, 2022). Because Xiaohongshu comments are short (mean lengths of 30.4, 23.7 and 17.7 characters for Tracks A, B and C, respectively), highly colloquial and irregularly punctuated—conditions under which the default BERTopic pipeline has stability limitations on extremely short Chinese text—we used a BERTopic-style variant: (1) jieba Chinese segmentation with a customised stop-word list (generic function words plus high-frequency non-informative domain tokens such as "now", "is", "not"); (2) TF-IDF vectorisation of document representations (max_features = 3,000, min_df = 2, max_df = 0.85); (3) truncated SVD dimensionality reduction (n_components = 50, random_state = 42); (4) K-means clustering (n_clusters = 8, 8 and 6 for Tracks A, B and C, determined via the elbow method); and (5) top-keyword extraction via mean c-TF-IDF per cluster.

On top of TF-IDF/cluster-based topics, we conducted domain-rule coding to produce cleaner semantic categories. Two coders with health-communication training drew on the modelled topics and representative texts to construct three rule dictionaries—one per track—and multi-label-coded each text. To assess reliability, 10% of each track (221 texts) were independently re-coded by a second coder; Cohen's κ values were 0.81, 0.84 and 0.78 for Tracks A, B and C respectively, reaching acceptable thresholds.

3.3 Cross-track Analysis: Stage Comparison and Supply–Demand Mismatch

To address RQ2, two aggregate indicators were constructed: "diagnostic seeking" (combining the DHIS and medical-inquiry topics from Track A with the report-interpretation, diagnostic-anxiety, exam-inquiry and clinical-indicator topics from Track B) and "emotional seeking" (combining emotional-support and bereavement topics from Track A with family-narrative and ritual-expression topics from Track B). Chi-square tests compared the distribution of these two behaviours across the pre- (Track B) and post- (Track A) diagnosis stages.

To address RQ3, the liver-lesion dataset was restructured into 543 question–reply (Q–R) dyads, with 364 unanswered questions (40.1% of the total—itself a finding of interest) excluded from this specific analysis. Within the dyadic matrix, we cross-tabulated question-type × reply-type, computed the distribution of reply types received by DHIS-type questions, and compared them against non-DHIS questions as a control.

4. Results

4.1 Three-Track Topic Structure (RQ1)

The results of topic modelling and rule coding across the three tracks are summarised in Figure 2 and Table 2. The three tracks exhibit markedly different topic structures.

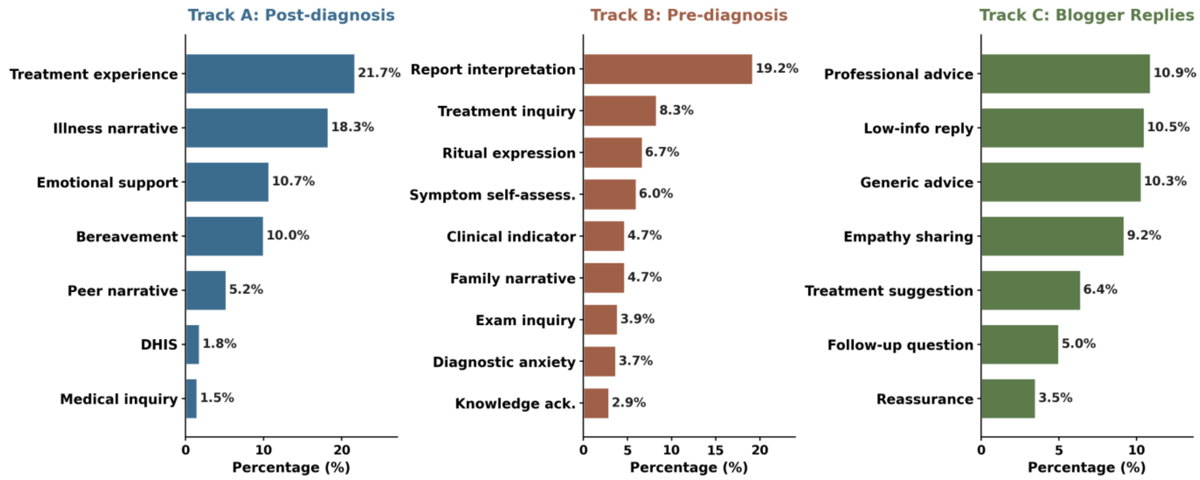


Figure 2. Topic distribution across three tracks (horizontal bars show share of texts per topic category).

Track A (n = 1,311) is dominated by "treatment experience" (21.7%) and "illness/family narrative" (18.3%), which together account for nearly 40% of content. Representative texts include "Lenvatinib stopped working after a while; my dad's coagulation and liver function are slightly better, so we're waiting until after Spring Festival to reassess for transplant" (sharing targeted-drug experience) and "My dad, late-stage liver cancer, came for gastroscopy on Aug 27; by Sept 1 his jaundice was at 213" (documenting rapid disease progression). These are consistent with the expectation that post-diagnosis users in the liver-cancer context centre on concrete medical decisions and caregiver narratives. "Emotional support" (10.7%) and "bereavement" (10.0%) form a second tier, indicating the persistent role of affective seeking in the post-diagnosis context. Strikingly, direct diagnostic queries (DHIS) constitute only 1.8% of Track A—reasonable, since the diagnosis has already been made and what users care about shifts from "what is it" to "how to cope".

Track B (n = 907) presents a very different structure. "Report interpretation" is the largest category at 19.2%, and, excluding "Other", represents roughly one-third of clearly themed texts. Typical examples include "Please help me interpret this near-isoechoic region" and "A chest CT incidentally showed a sub-5 mm hypoechoic area in my liver—should I get a follow-up?" Such requests are often accompanied by user-uploaded ultrasound or CT images, constituting a signature "upload + request interpretation" discourse pattern. "Treatment inquiry" (8.3%), "symptom self-assessment" (6.0%), "clinical-indicator inquiry" (4.7%), "exam inquiry" (3.9%), and "diagnostic anxiety" (3.7%) together account for about 26.6% of Track B and collectively constitute the core of DHIS. Also notable is "ritual expression" (6.7%)—for example, the "tuituitui (go away, go away)" prayer-style comment often invoked when discussing illness—reflecting a collective emotional-relief pattern distinctive to Xiaohongshu comment sections.

Track C (n = 543) reveals the stratified quality of the supply side. "Professional medical advice" (10.9%) is the largest substantive reply category, exemplified by "Get a contrast MRI as soon as possible" or "ALT twice the upper limit means hepatocellular inflammation; you should see a paediatric specialist"—responses that provide specific examination recommendations or indicator interpretation. "Low-information replies" (10.5%)—single-character responses such as "reply" or short interjections like "how are things now?"—occupy a near-equal share. More broadly, "generic advice" (10.3%, such as "just go get a recheck") offers only slightly more information. "Empathy sharing" (9.2%, e.g., "Same here—did you ever find the cause?") represents a distinct pattern in which responders substitute peer solidarity for professional judgement—highlighting the dual role of "seeker-as-responder" on the platform.

Table 2. Theme Categories across Three Tracks

ID	Topic	Top keywords	%	Example
Track B: Pre-diagnosis Questions (n = 907)				
B1	Report interpretation	help me look, what does it mean, image	19.2%	*"Help me check this near-isoechoic area."*
B2	Treatment inquiry	treatment, surgery, targeted, antiviral	8.3%	*"How to intervene at stage I?"*
B3	Symptom self-assess.	liver pain, itching, fatigue, weight loss	6.0%	*"Occasional pain on right rib. Why?"*
B4	Clinical indicator	ALT, bilirubin, AFP, nodule, lesion	4.7%	*"ALT twice the normal — what does it mean?"*
B5	Family narrative	my dad, my mom, father, husband	4.7%	*"My dad is also in decompensated cirrhosis."*
B6	Exam inquiry	recheck, contrast CT, MRI, ultrasound	3.9%	*"Do I still need a contrast CT after MRI?"*

ID	Topic	Top keywords	%	Example
B7	Diagnostic anxiety	could it be, serious, scared, cancer	3.7%	*"Nodule grew 2.5→3.5cm in a year. Cancer?"*
Track A: Post-diagnosis Comments (n = 1,311)				
A1	Treatment experience	surgery, targeted, lenvatinib, TACE, TCM	21.7%	*"Lenvatinib lost effect; waiting for transplant."*
A2	Illness narrative	late-stage, metastasis, jaundice, HBV	18.3%	*"Jaundice shot from 213 to 434 in 8 days."*
A3	Emotional support	keep going, bless, safe, health	10.7%	*"Bless all the family with health."*
A4	Bereavement	passed away, gone, left us, condolence	10.0%	*"My dad left us on Feb 10."*
A5	Peer narrative	my uncle, colleague, neighbor, grandpa	5.2%	*"My uncle was diagnosed in June, gone by Dec."*
A6	DHIS	help me look, doctor, is this, serious	1.8%	*"Could this be liver palm?"*
Track C: Blogger Replies (n = 543)				
C1	Professional advice	contrast CT/MRI, biopsy, specialist	10.9%	*"Get a contrast MRI as soon as possible."*
C2	Generic advice	recheck, go to hospital, see doctor	10.3%	*"Go get a recheck."*
C3	Low-info reply	OK, hmm, reply (≤3 characters)	10.5%	*"Reply."*
C4	Empathy sharing	me too, same here, also my family	9.2%	*"Same here — did you ever find the cause?"*
C5	Treatment suggestion	antiviral, liver protection, stop drinking	6.4%	*"Watch diet, exercise, recheck regularly."*
C6	Follow-up question	how is it now, any update	5.0%	*"How are things going now?"*
C7	Reassurance	nothing serious, don't worry, minor	3.5%	*"Mild elevation, nothing serious."*

4.2 Cross-stage Comparison: Diagnostic and Emotional Seeking as Mirror Distributions (RQ2)

Comparing the two tracks along the aggregate dimensions of diagnostic and emotional seeking yields the central finding of this study (Figure 3). In the pre-diagnosis stage (Track B), diagnostic seeking accounts for 31.5% and emotional seeking for 11.5%; in the post-diagnosis stage (Track A), diagnostic seeking drops to 3.4% while emotional seeking rises to 20.7%. Chi-square tests confirm that both distributions differ substantially across the two stages: for diagnostic seeking, $\chi^2(1) = 333.84$, $p < .001$; for emotional seeking, $\chi^2(1) = 31.68$, $p < .001$.

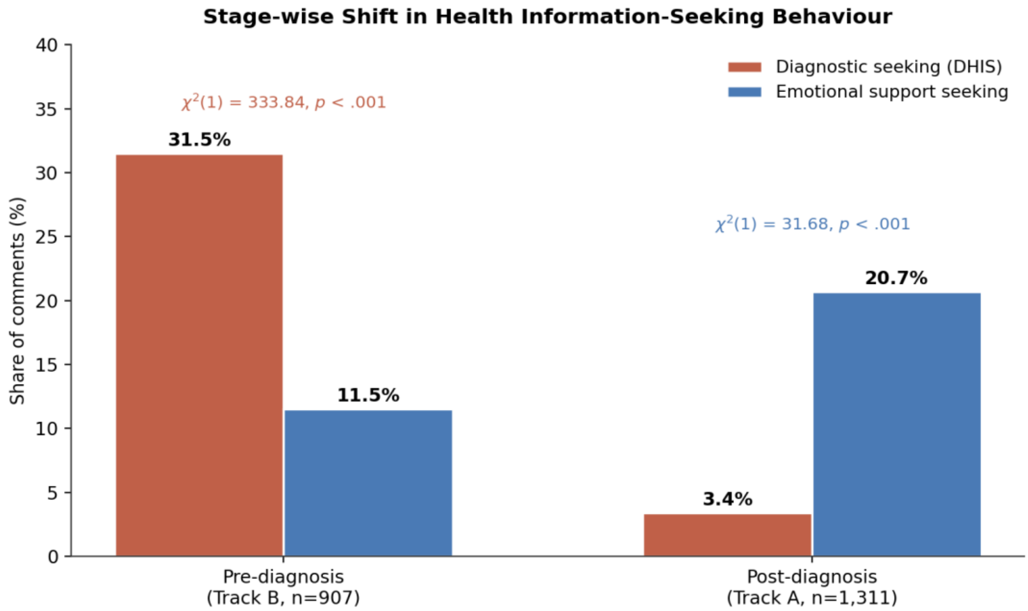


Figure 3. Stage-wise shift in health information-seeking behaviour between pre- and post-diagnosis contexts.

This "mirror-image" distribution has clear theoretical significance. First, it directly validates the illness-trajectory/information-need hypothesis advanced by Lambert and Loiselle (2007): pre-diagnosis users face the foundational uncertainty of "what is it", so diagnostic needs dominate; post-diagnosis users pivot to "how to cope", where instrumental and affective needs rebalance and affective seeking becomes more prominent. Second, the pattern reveals a previously overlooked fact: DHIS on Xiaohongshu is most active not in comment sections of the confirmed disease ("liver cancer") but in those of the undetermined liminal state ("liver lesion"). This has important implications for HISB research, since frameworks based on already-diagnosed patients may systematically under-estimate pre-diagnosis demand for informal diagnostic information.

A closer look at DHIS sub-categories within Track B reveals that "report interpretation" alone accounts for 19.2%—meaning that nearly one in five pre-diagnosis comments is a direct request for interpretation of a user's imaging or laboratory report. This behaviour sits awkwardly within existing HISB taxonomies: it is neither "active information seeking" in Wilson's sense (users are not searching for existing information) nor purely instrumental in Longo's sense (users seek a diagnostic verdict, not a decision-input). It is precisely this gap that motivates the DHIS construct.

4.3 Information Supply–Demand Mismatch (RQ3)

The Q–R dyadic matrix (Figure 4) uncovers a complex matching landscape. First, at the level of overall response: of the 907 questions in Track B, 364 (40.1%) received no reply whatsoever, with this proportion being highest in the "report interpretation" category. In other words, even users who took the step of uploading a diagnostic report often faced silence.

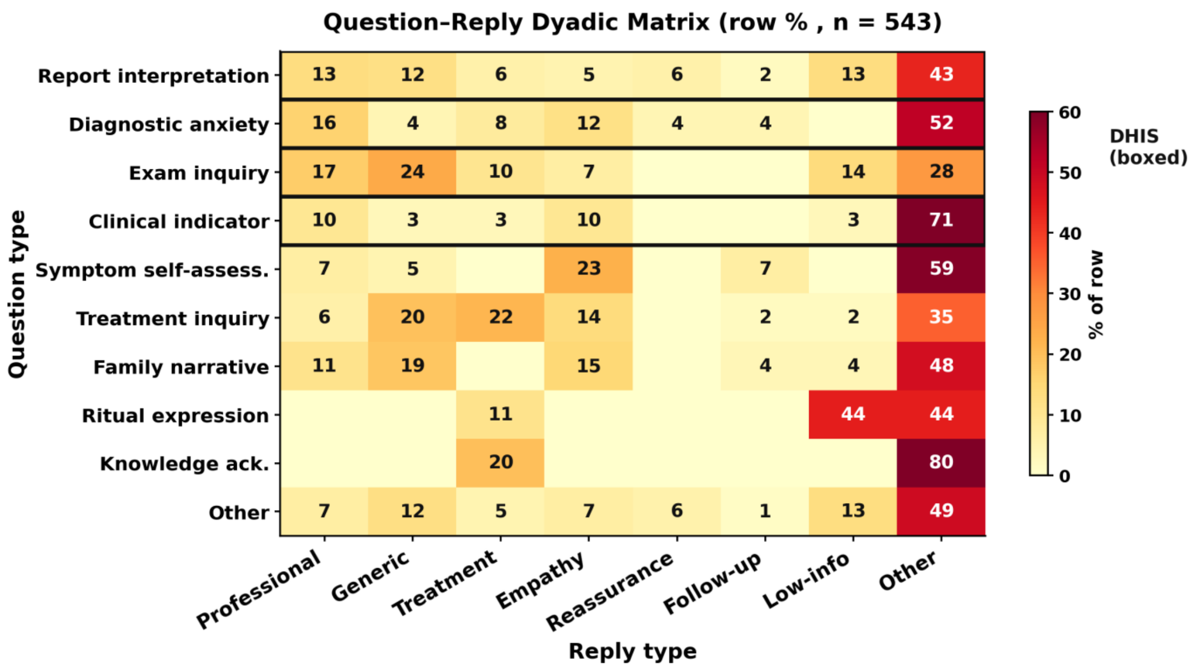


Figure 4. Question–reply dyadic matrix (row-normalized %). DHIS-type questions are boxed.

Focusing on the 543 paired dyads—in particular, the 191 DHIS-type dyads (report interpretation, diagnostic anxiety, exam inquiry, and clinical-indicator inquiry)—we find that only 13.6% (n = 26) received professional medical advice (e.g., recommend contrast CT, further biopsy, AFP testing), 11.5% received generic advice, 10.9% received low-information replies, 9.9% received follow-up questions, 6.8% received peer empathy, and 3.7% received explicit reassurance. The proportion of professional replies (13.6%) is roughly on par with the aggregated low-quality reply rate (15.2%).

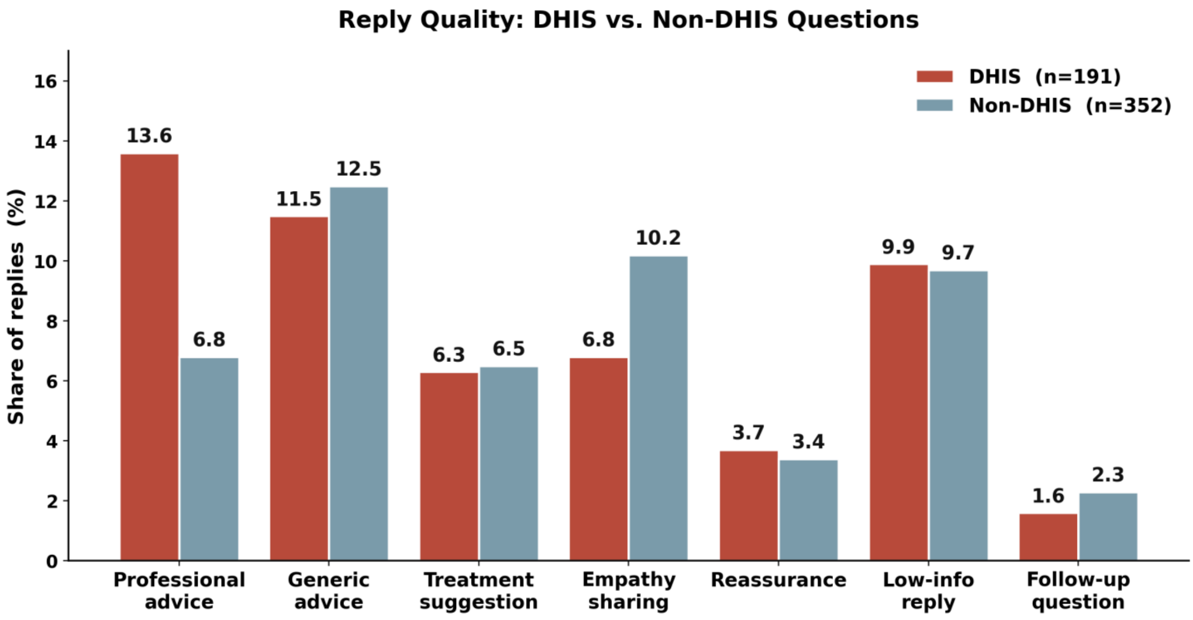


Figure 5. Reply quality distribution: DHIS vs. non-DHIS questions.

Comparing reply-quality distributions between DHIS and non-DHIS questions (Figure 5) reveals an important nuance: DHIS questions are twice as likely to receive professional advice (13.6%) as non-DHIS ones (6.8%), suggesting that the supply side is not entirely blind to the severity of diagnostic requests and can, to some degree, recognise "serious report uploads" and respond accordingly. Yet, in absolute terms, the level remains low: nearly nine in ten diagnostic questions do not receive what could be called a professional answer. This combination of "partial discrimination, but insufficient absolute provision" constitutes the core tension of platform medicalisation.

Three typical mismatch patterns emerge. First, affective substitution mismatch: a user uploads imaging of a suspicious liver lesion and asks "could this be cancer?" only to receive "don't worry, it's fine"—professional concern substituted by emotional soothing, potentially delaying needed follow-up. Second, boomerang mismatch: in response to a specific exam query, the responder asks back "what did your doctor say?"—tossing diagnostic responsibility back to the asker. Third, vague-advice mismatch: a detailed lesion description receives "go to a hospital and check"—neither harmful nor genuinely informative. These three share a common structure: a polite facade that fails to meet the user's diagnostic need—"well-intentioned low-quality replies" as a distinctive feature of the platform's health-information ecology.

5. Discussion

5.1 Platform Medicalisation: Structural Causes and Risk Map

Our findings provide robust empirical support for the platform-medicalisation hypothesis. 31.5% of pre-diagnosis comments exhibit DHIS characteristics, and almost six in ten users open with honorifics such as "doctor", "chief physician" or "hello", pre-framing the blogger or other responders as quasi-medical authorities. This usage reflects three structural drivers.

First, the access bottleneck of formal medical channels. Crowded, short outpatient encounters at tertiary hospitals systematically leave "report interpretation" under-served; physicians tend to deliver conclusions but rarely walk patients through each indicator. This information vacuum turns into lasting cognitive uncertainty after the visit, driving users onto platforms for compensatory interpretation—resonating with Eysenbach's (2008) "information asymmetry → platform compensation" model.

Second, the immediacy of health anxiety. Ambiguous reporting language for liver lesions—"hypoechoic", "hyperechoic", "nodule"—excites strong user anxiety. Terms such as "could it be cancer?", "is it serious?" and "scared" appear at high frequency in Track B, showing that the anxiety triggered by a report demands an immediate outlet. Formal care usually requires appointments, queues, and days or weeks of waiting, whereas Xiaohongshu offers an immediate response window—even if the response is of low quality.

Third, weak-tie trust and platform persona. Xiaohongshu's "blogger economy" builds a peculiar perceived credibility: even when bloggers are not actually physicians, their accumulated "medical-popularisation" persona (white-coat avatar, title cues such as "Dr." or "Chief"), together with platform social signals, establish sufficient trust. The gap between such pseudo-medical authority and verified qualification is the core risk of platform medicalisation.

Together, these drivers produce a risk map: (1) mis-judgement risk—treating platform replies as credible diagnoses and delaying care; (2) affective-misguidance risk—reassurance ("it's nothing") lulling at-risk individuals into complacency; (3) collective-anxiety amplification—ritualised comment-section prayers (tuituitui) expressing empathy yet potentially cascading into contagious group anxiety; and (4) medical-authority erosion—long-term platform reliance systematically weakening the authority of formal medical discourse.

5.2 Theoretical Positioning of DHIS

The DHIS construct fills a notable gap in the existing HISB taxonomy. Longo's (2005) instrumental-versus-affective binary fails to capture "asking others for a verdict" as a distinctive instrumental act; Lambert and Loiselle's (2007) stage model notes a preference for instrumental information at the diagnosis stage but does not distinguish between "retrieving existing information" and "soliciting a diagnostic verdict"—two fundamentally different instrumental acts. DHIS enables researchers to isolate behaviours such as "upload + interpret" from general health information retrieval and examine their distinctive supply–demand dynamics.

DHIS also has comparative-research value. Existing work has noted that Chinese users differ from Western users in aspects such as strong reliance on physician authority, high sensitivity to quasi-authority signals (white coat, hospital backdrop), and avoidance/platform-usage patterns for sensitive conditions (Lin et al., 2016; Cao et al., 2016). The DHIS construct can ground these macro-level differences in concrete behaviour categories, enabling comparative analysis of how frequently users in different cultures "seek others' verdicts" and how such verdicts are provided.

5.3 Mechanisms of Supply–Demand Mismatch: An Incentive-Structure Lens

The mismatch we document is not simply a matter of insufficient expertise among responders; it is rooted in the incentive structure of UGC platforms. The core incentive of UGC platforms is user dwell-time and engagement, and the goals of bloggers and top commenters are sustained follower engagement and account ranking. In this structure, high-stakes, high-responsibility professional answers (e.g., "your imaging highly suggests HCC; go to a tertiary hospital immediately for contrast MRI") require both expertise and a willingness to shoulder the responsibility of potential mis-judgement. Low-stakes, low-threshold responses ("go for a recheck", "don't worry"), by contrast, satisfy the interaction metric while carrying essentially zero judgement-error cost. This asymmetric risk–return structure systematically rewards low-information, responsibility-evading replies.

This also explains why "peer empathy" (9.2%) occupies a substantial share of the supply side. When responders are themselves seekers (e.g., also worried about their own lesions), they may lack the expertise to render professional judgement but can nonetheless complete the interaction by expressing "me too" empathy—fulfilling a social exchange while bearing no clinical responsibility. This dual identity of "seeker-as-responder" is a structural feature of the platform's health-information ecology.

5.4 In Dialogue with Existing Research: A Full-Trajectory Picture

Our findings complement earlier work on caregiver discourse in liver-cancer contexts. Previous studies have emphasised the proxy-disclosure and information-seeking behaviour of family caregivers after diagnosis, revealing the dual burden of information and emotion borne by family members under non-disclosure norms. The present study pushes the analytic frame pre-diagnosis, portraying the latent patient as the primary diagnostic seeker. Together, the two lines yield three complementary axes: proxy-versus-primary, post- versus pre-diagnosis, and emotion-versus-diagnosis. This full-chain perspective moves health-communication research beyond the patient–family dyad into a multi-actor network of latent patients, platforms, bloggers and top commenters—embedding each node in an intensely platformed socio-technical system.

5.5 Implications for Platform Governance

Three layers of intervention are suggested. At the platform layer, health content should be tagged in a tiered manner: general popularisation content may retain open commenting, while diagnostic-consultation content should automatically surface a disclaimer ("this platform is not a medical institution; please seek professional advice") and a care-navigation widget (e.g., appointment links to tertiary hospitals) triggered by evident report uploads. At the content-creator layer, health bloggers above a follower threshold should be subject to credential verification and targeted health-information-literacy training that clarifies the discursive boundaries of diagnostic replies (e.g., when it is appropriate to say only "seek immediate medical care" rather than offer specific judgements). At the user layer, digital health literacy should be incorporated into public-health education, with a particular focus on the cognitive distinction between "platform response" and "medical diagnosis".

6. Conclusion and Future Directions

Integrating 2,761 texts collected from two keyword-specific Xiaohongshu datasets ("liver cancer" and "liver lesion"), and applying a BERTopic-style pipeline combined with rule-based coding, this study has offered a systematic account of user health information-seeking behaviour across the full liver-cancer trajectory on a major Chinese social-media platform. Three main findings emerge. First, diagnostic seeking accounts for 31.5% of pre-diagnosis content but only 3.4% of post-diagnosis content—revealing a pronounced stage-wise differentiation. Second, emotional seeking exhibits an inverse distribution (post-diagnosis 20.7% vs. pre-diagnosis 11.5%), with the two forms of seeking undergoing a "mirror-image switch" across stages that arguably constitutes a characteristic trajectory of Chinese-context HISB. Third, the information supply side exhibits substantial mismatch: 40.1% of diagnostic questions receive no response, and of those answered, only 13.6% receive professional advice, with three typical mismatch types—affective substitution, boomerang, and vague advice—accounting for the bulk of the remainder.

The theoretical contributions are twofold. First, we propose the DHIS construct, extending the HISB taxonomy for the social-media context. Second, we propose the platform-medicalisation concept, offering a new analytic lens for the structural phenomenon by which social-media platforms increasingly take on informal medical-diagnostic roles. Methodologically, we demonstrate a mixed-methods pipeline—three-track parallel topic modelling, rule-based coding, and question–reply dyadic analysis—reusable for short-text, two-sided social-media data.

Several limitations apply. The data come from a single platform, so generalisation to Weibo, Douyin, Zhihu and others must proceed with caution. The use of rule-based coding, even with BERTopic-style clustering and inter-coder agreement checks, retains a residual coder-prior bias. The reply-side sample (543 texts) is moderate, and mismatch conclusions should be cross-validated with other UGC data. The design is cross-sectional; it cannot trace how a single user's behaviour evolves along the disease trajectory, a task for which longitudinal designs are better suited.

Three directions warrant future work: (1) extending DHIS to other chronic-disease contexts (cardiovascular, diabetes, mental health) for cross-condition comparison; (2) experimental designs that trace the downstream care-seeking effects of different reply types (professional advice, reassurance, vague advice); and (3) evaluation of platform-governance interventions (tiered labelling, credential verification) using quasi-experimental designs. Progress on these fronts would provide an actionable evidence base for public-health governance in the digital age.

Funding

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

References

Cao, W., Zhang, X., Xu, K., & Wang, Y. (2016). Modeling online health information-seeking behavior in China: The roles of source characteristics, reward assessment, and internet self-efficacy. *Health Communication, 31*(9), 1105–1114. https://doi.org/10.1080/10410236.2015.1045236

Chambers, D., Cantrell, A. J., Johnson, M., Preston, L., Baxter, S. K., Booth, A., & Turner, J. (2019). Digital and online symptom checkers and health assessment/triage services for urgent health problems: Systematic review. *BMJ Open, 9*(8), e027743. https://doi.org/10.1136/bmjopen-2018-027743

Chang, L., Li, W., Xin, X., & Wang, J. (2024). Strategies for assessing health information credibility among older social media users in China: A qualitative study. *Health Communication, 39*(12), 2767–2778. https://doi.org/10.1080/10410236.2023.2288372

Chen, C., Xu, Q., Wang, T., Yuan, Y., Hu, Z., Li, X., Zheng, X., & Lei, H. (2025). Comparative analysis of liver cancer burden trends in China and worldwide between 1990 and 2021. *Frontiers in Public Health, 13*, 1513210. https://doi.org/10.3389/fpubh.2025.1513210

Chen, J., & Wang, Y. (2021). Social media use for health purposes: Systematic review. *Journal of Medical Internet Research, 23*(5), e17917. https://doi.org/10.2196/17917

Conrad, P. (2007). *The medicalization of society: On the transformation of human conditions into treatable disorders*. Baltimore, MD: Johns Hopkins University Press.

Diao, X., Guo, C., Jin, Y., Li, B., Gao, X., Du, X., Chen, Z., Jo, M., Zeng, Y., Ding, C., Liu, W., Guo, J., Li, S., & Qiu, H. (2025). Cancer situation in China: An analysis based on the global epidemiological data released in 2024. *Cancer Communications, 45*(2), 178–197. https://doi.org/10.1002/cac2.12627

Eysenbach, G. (2008). Medicine 2.0: Social networking, collaboration, participation, apomediation, and openness. *Journal of Medical Internet Research, 10*(3), e22. https://doi.org/10.2196/jmir.1030

Eysenbach, G., Powell, J., Kuss, O., & Sa, E.-R. (2002). Empirical studies assessing the quality of health information for consumers on the World Wide Web: A systematic review. *JAMA, 287*(20), 2691–2700. https://doi.org/10.1001/jama.287.20.2691

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*. https://doi.org/10.48550/arXiv.2203.05794

Huang, Y., & Miao, W. (2024). Domesticating algorithms through data reflectivity and user reflexivity: The metaphor of Yanghao on Xiaohongshu (RED). *Convergence: The International Journal of Research into New Media Technologies*. Advance online publication. https://doi.org/10.1177/13548565241301146

Lambert, S. D., & Loiselle, C. G. (2007). Health information–seeking behavior. *Qualitative Health Research, 17*(8), 1006–1019. https://doi.org/10.1177/1049732307305199

Lederman, R., Fan, H., Smith, S., & Chang, S. (2014). Who can you trust? Credibility assessment in online health forums. *Health Policy and Technology, 3*(1), 13–25. https://doi.org/10.1016/j.hlpt.2013.11.003

Li, C., Khan, M. M., & Chen, Z. (2023). Public trust of physicians in China improved since the COVID-19 pandemic began. *Social Science & Medicine, 320*, 115704. https://doi.org/10.1016/j.socscimed.2023.115704

Li, Q., Ding, C., Cao, M., Yang, F., Yan, X., He, S., Cao, M., Zhang, S., Teng, Y., Tan, N., Wang, J., Xia, C., & Chen, W. (2024). Global epidemiology of liver cancer 2022: An emphasis on geographic disparities. *Chinese Medical Journal, 137*(19), 2334–2347. https://doi.org/10.1097/CM9.0000000000003264

Lin, W.-Y., Zhang, X., Song, H., & Omori, K. (2016). Health information seeking in the Web 2.0 age: Trust in social media, uncertainty reduction, and self-disclosure. *Computers in Human Behavior, 56*, 289–294. https://doi.org/10.1016/j.chb.2015.11.055

Longo, D. R. (2005). Understanding health information, communication, and information seeking of patients and consumers: A comprehensive and integrated model. *Health Expectations, 8*(3), 189–194. https://doi.org/10.1111/j.1369-7625.2005.00339.x

McGlynn, K. A., Petrick, J. L., & El-Serag, H. B. (2021). Epidemiology of hepatocellular carcinoma. *Hepatology, 73*(Suppl 1), 4–13. https://doi.org/10.1002/hep.31288

Sbaffi, L., & Rowley, J. (2017). Trust and credibility in web-based health information: A review and agenda for future research. *Journal of Medical Internet Research, 19*(6), e218. https://doi.org/10.2196/jmir.7579

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems, 12*(4), 5–33. https://doi.org/10.1080/07421222.1996.11518099

Wang, Q., Poudel-Adhikari, S., Wu, Y., Zhao, M., Kolstad, J., Zhao, L., Li, X., Xu, D. R., & Lu, C. (2022). Consultation length, process quality and diagnosis quality of primary care in rural China: A cross-sectional standardized-patient study. *Patient Education and Counseling, 105*(4), 902–908. https://doi.org/10.1016/j.pec.2021.07.014

Wang, X., Shi, J., & Kong, H. (2021). Online health information seeking: A review and meta-analysis. *Health Communication, 36*(10), 1163–1175. https://doi.org/10.1080/10410236.2020.1748829

Wang, J., Yang, J., & Zhang, Y. (2023). Acquiring and sharing health information on WeChat among older adults in China: A qualitative study. *BMC Geriatrics, 23*(1), 405. https://doi.org/10.1186/s12877-023-04106-7

Wilson, T. D. (1999). Models in information behaviour research. *Journal of Documentation, 55*(3), 249–270. https://doi.org/10.1108/EUM0000000007145

Wu, T., Deng, Z., Zhang, D., Buchanan, P. R., Zha, D., & Wang, R. (2018). Seeking and using intention of health information from doctors in social media: The effect of doctor–consumer interaction. *International Journal of Medical Informatics, 115*, 106–113. https://doi.org/10.1016/j.ijmedinf.2018.04.009

Xiong, Z., Zhang, L., Li, Z., Xu, W., Zhang, Y., & Ye, T. (2021). Frequency of online health information seeking and types of information sought among the general Chinese population: Cross-sectional study. *Journal of Medical Internet Research, 23*(12), e30855. https://doi.org/10.2196/30855

Zhang, X., Chen, B., Li, G., & Dong, Y. (2022). Exploring the health information seeking behavior of social media users under the background of COVID-19 pandemic: An empirical study based on social cognitive theory. *Frontiers in Psychology, 13*, 1045914. https://doi.org/10.3389/fpsyg.2022.1045914

Zhang, X., Wen, D., Liang, J., & Lei, J. (2017). How the public uses social media WeChat to obtain health information in China: A survey study. *BMC Medical Informatics and Decision Making, 17*(Suppl 2), 66. https://doi.org/10.1186/s12911-017-0470-0

Zheng, R., Qu, C., Zhang, S., Zeng, H., Sun, K., Gu, X., Xia, C., Yang, Z., Li, H., Wei, W., Chen, W., & He, J. (2018). Liver cancer incidence and mortality in China: Temporal trends and projections to 2030. *Chinese Journal of Cancer Research, 30*(6), 571–579. https://doi.org/10.21147/j.issn.1000-9604.2018.06.01