Marketing Science Institute Working Paper Series 2026

Report No. 26-117

When Professionals Become Influencers: The Impact of Live Streaming on Service Demand

Xu Zhang, Yuqing Gu, Yi Zhang, Puneet Manchanda

When Professionals Become Influencers: The Impact of Live Streaming on Service Demand

Xu Zhang
Assistant Professor of Marketing, London Business School
xzhang@london.edu

Yuqing Gu
PhD Candidate, University of Michigan
guyuqing@umich.edu

Yi Zhang
PhD Candidate, University of International Business and Economics
zhang_yi@uibe.edu.cn

Puneet Manchanda
Isadore and Leon Winkelman Professor of Marketing, University of Michigan
pmanchan@umich.edu

March 2026

Acknowledgments: The authors thank participants at the 2026 UT Dallas FORMS conference, 2026 London Area Marketing Assembly, the 2025 Conference on Artificial Intelligence, Machine Learning, and Business Analytics, the 4th Workshop on UK Digital Economics, the 1st Creator Economy Conference, and the 45th Marketing Science Conference for their helpful comments and feedback.

Financial Disclosure Statement: The authors acknowledge the financial support of the Research and Materials Development Fund of London Business School (RAMD_Zhang_X_24/25_8895 and RAMD_Zhang_X_25/26_2963).

Ethics Approval Statement: The study was approved by the Research Ethics Committee at London Business School (Approval No. REC1080).

When Professionals Become Influencers: The Impact of Live Streaming on Service Demand

Abstract

A growing number of professionals, such as physicians, use live streaming to attract clients and promote their services. Yet whether live streaming generates demand in professional service settings, and which features predict its effectiveness, remains unclear. We examine these questions using data from one of China's largest online healthcare platforms, combining consultation records for more than 7,000 physicians with detailed information on their live streaming sessions. Using a generalized synthetic control approach, we show that live streaming adoption causally increases service demand. The effect persists for several months and is heterogeneous, with larger gains for more established physicians, including chief and associate chief physicians and those with higher baseline consultation volumes. To analyze live streaming effectiveness, we conduct a multimodal analysis of live streaming videos, extracting visual, auditory, and textual features using machine learning and large language models. Audience interaction, content, linguistic features, and emotional expression emerge as the strongest predictors of effectiveness. Interactive sessions, clear and concrete language, and a neutral emotional display are associated with stronger demand responses, whereas greater emotional variability is associated with weaker responses. Extending the analysis to investment fund managers reveals similar patterns, suggesting that live streaming is an effective demand-generation tool across professional service domains.

Keywords: Live streaming, Professional Services, Healthcare, Service Marketing, Synthetic Control, Artificial Intelligence

INTRODUCTION

Influencer marketing is among the fastest growing forms of digital promotion, with global spending rising from USD 6 billion in 2017 to an estimated USD 40 billion in 2025.[1] Across platforms such as TikTok, Instagram, Twitch, and Taobao, influencers drive sales, shape consumer preferences, and build brand engagement (Beichert et al. 2024; Gu, Zhang, and Kannan 2024; Huang and Morozov 2025; Yang, Zhang, and Zhang 2025). The mechanisms underlying this effectiveness, including information value, engagement, emotional expression, and audience interaction, are increasingly well documented (Beichert et al. 2024; Cascio Rizzo et al. 2024; Lu et al. 2021; Yang, Zhang, and Zhang 2025). Recently, however, this phenomenon has expanded beyond consumer products to professional services. In these settings, professionals, defined as individuals certified or licensed by recognized third parties such as professional or regulatory bodies, are beginning to act as influencers, leveraging digital platforms to reach clients directly, often via live streaming. Examples include physicians, lawyers, nutritionists, and financial advisors. This shift raises important new questions: does the channel of influencer marketing function in the same way when professionals, whose credibility and expertise are critical, become influencers themselves? Or does this setting introduce new predictors for effectiveness distinct from those observed in consumer-product contexts?

The emergence of professional influencers has prompted debate. Advocates argue that live streaming democratizes access to knowledge, fosters transparency, and creates new ways for professionals to connect with clients. Critics, however, worry about conflicts of interest, credibility risks, and the commercialization of professional expertise. In healthcare, for instance, live streaming can provide valuable education to patients, a practice encouraged by some government agencies,[2] but it may also be perceived as self-promotion, raising concerns that visibility-seeking could undermine professionalism and service quality and calling for

[1]Source: Influencer Advertising – Worldwide 2025, Statista.

[2]See "Science Popularization, Live Streaming Classes, and Online Consultations: Short-Video Platforms Collaborate with the Medical Sector to Support Epidemic Prevention and Control" by China National Radio.

stronger regulation.[3] These tensions underscore the need for empirical evidence on both the effectiveness of professional live streaming and the drivers that predict its effectiveness.

In this paper, we address two questions. First, does professional live streaming increase demand for professional services? Second, if so, what factors predict its effectiveness? We answer these questions in healthcare, focusing on physicians on Haodf.com, one of China's largest online consultation platforms. As of 2024, close to 30,000 live streaming sessions have been conducted on the platform, attracting millions of viewers. We compile a unique dataset that combines monthly consultation records for 7,226 physicians, of whom 2,271 are streamers and the remainder are non-streamers. In addition, we collect detailed live streaming video data. This setting allows us to examine both the causal effect of adopting live streaming on service demand and the streaming features that predict its effectiveness.

To estimate the causal effect of live streaming adoption on physicians' consultation demand, we employ a generalized synthetic control approach. This method uses a factor-based model to learn latent common trends from non-streaming physicians and streaming physicians' pre-adoption periods, and then predicts each streaming physician's post-adoption consultation demand absence of adoption, providing a credible counterfactual. The estimates reveal that live streaming leads to a significant increase in demand, with the effect persisting for several months before tapering off. We further examine four potential pathways for this effect: platform visibility (ranking), perceived service quality (ratings), pricing, and production of patient-education articles. The estimated demand effect is similar when we additionally control for these observed time-varying measures. These results are robust across multiple checks. Together, these findings provide causal evidence that live streaming serves as an effective demand-generation tool in professional services.

We further examine heterogeneity in the effect of live streaming adoption. Along pre-adoption characteristics that are observable to the platform and can inform targeting, the estimated demand gains are more pronounced among more established physicians: effects are

[3] See "Regulation on Medical Science Popularization Activities Conducted by Self-Media Outlets" by Xinhua Net.

stronger for chief/associate chief physicians and for physicians with higher baseline consultation volume. We do not observe systematic differences by hospital tier. Across specialties, the point estimates are positive for the majority of categories, although statistical significance is concentrated in a subset of specialties with larger treated samples.

Motivated by the substantial heterogeneity in the effectiveness of live streaming on service demand, we examine why some physicians' live streaming sessions are more effective than others. To do so, we conduct a multimodal analysis of live streaming videos. Specifically, we extract visual, auditory, and textual information from each physician's initial live streaming session and construct features grounded in theories from marketing, psychology, organizational behavior, public health, and information systems, adapted to the context of professional live streaming. Using a combination of machine learning methods, deep learning architectures, and large language models, we derive 38 attributes across the three modalities, spanning eight feature categories: audience interaction, session content, professional appearance, emotional expression, linguistic features, acoustic-prosodic features, visual composition, and promotional strategies.

Linking these features to individual treatment effects, we find that effectiveness is primarily associated with audience interaction, session content, linguistic features, and emotional expression. Interactive sessions, concrete and simple language, and neutral emotional expression are associated with stronger live streaming demand effects, while greater emotional variability is negatively related to effectiveness. These findings partially contrast with results from pre-recorded video and commercial live streaming settings and highlight mechanisms distinctive to professional live streaming, where trust, expertise, and real-time engagement play a central role.

To test the generalizability of our findings beyond healthcare, we extend our analysis to fund managers in China, a professional service domain that differs fundamentally from healthcare. Using a similar generalized synthetic control approach and multimodal content analysis, we find that when fund managers mention funds during live streams, those funds experience significant increases in demand, as measured by trading volume. The consistency of these effects across both physicians and fund managers, despite their distinct professional contexts, provides strong evidence that live streaming serves as an effective demand-generation tool across diverse credence good domains. Moreover, we find that key live streaming features, particularly audience interaction, emotional expression, and linguistic features, operate similarly across both contexts.

Our study makes three contributions. First, we quantify the causal effect of professional live streaming on service demand, extending the influencer marketing literature to a new and important domain where the effect is a priori ambiguous. Second, we identify the live streaming attributes that predict effectiveness, highlighting factors that differ from those emphasized in consumer influencer contexts. Third, we provide actionable implications for professionals and platforms: professionals can use these insights to attract more clients, while platforms can design features and training to help professionals create effective streams. Together, these contributions advance understanding of digital demand generation in professional services and offer guidance for the evolving role of professionals as influencers.

RELATED LITERATURE

This study builds on three streams of literature. First, we contribute to research on influencer marketing, which has largely examined effectiveness in settings where influencers promote third-party products. Prior work has shown that influencer marketing can drive sales and consumption across a variety of platforms and formats, from short-form videos (e.g., TikTok, Instagram, Weibo) (Beichert et al. 2024; Bharadwaj et al. 2022; Leung et al. 2022; Yang, Zhang, and Zhang 2025) to live streaming environments (e.g., Twitch, TikTok, Taobao) (Gu, Zhang, and Kannan 2024; Huang and Morozov 2025; Li, Haviv, and Lovett 2025). Comprehensive reviews by Libai et al. (2025) and Pan et al. (2025) highlight a consistent theme: context shapes both the overall effectiveness of influencer marketing and the mechanisms that drive it. We extend this literature by examining a novel and increasingly important context in which professionals act as influencers. Unlike commercial influencers, professionals highlight credibility and expertise, with the aim of building awareness, trust, and long-term relationships rather than driving immediate purchases. While elements of entertainment and persuasion remain relevant, the balance of the mechanisms is likely to shift. These differences motivate our focus on factors such as audience interaction, perceived professionalism, emotional expression, and linguistic features, and allow us to assess whether established theories of influencer marketing generalize to professional live streaming.

Second, our work contributes to research on how professionals attract clients. Prior studies have identified a range of tools that facilitate client acquisition, including online reviews (Kaye, Luca, and Vats 2024; Xu, Armony, and Ghose 2021), platform endorsements (Bairathi, Zhang, and Lambrecht 2025; Zhan, Zhang, and Fu 2026), client testimonials (Shimp, Wood, and Smarandescu 2007), and advertisements (Clow, Tripp, and Kenny 1996). By contrast, professionals acting as influencers and engaging in live streaming represent a novel and still underexplored channel. Although live streaming has the potential to raise awareness, foster relationships, and generate future clients, empirical evidence on its effectiveness in professional services remains limited. We address this gap by examining the live streaming effectiveness of professionals in two distinct domains, healthcare and finance, and identifying the factors that predict effectiveness.

Lastly, our work relates to the growing body of literature on video analysis. Video analysis has become increasingly relevant for marketing research (Rajaram and Manchanda 2020; Tian, Dew, and Iyengar 2024; Yang, Zhang, and Zhang 2025) given both the rising consumption of video content by consumers and the growing use of video by firms. To examine the predictors of influencer marketing effectiveness, we adopt a multimodal framework to analyze live streaming videos, incorporating visual, auditory, and textual modalities (Snoek and Worring 2005). After extracting information from these three modalities in the raw video data, we move beyond direct reliance on representation learning approaches (e.g., embeddings and autoencoders), which can obscure interpretation. Instead, we employ a theory-driven approach to guide variable construction, complemented by machine learning methods, deep learning architectures, and large language models (LLMs) where appropriate for each modality and variable. For example, text features capture sentiment and linguistic complexity leveraging domain-specific dictionaries and LLMs; audio features reflect emotional prosody and tone using deep learning models; and visual features measure production quality and visual appeal using cloud-based computer vision services and vision LLMs. This design ensures that each constructed variable has a clear theoretical meaning and can be interpreted using SHapley Additive exPlanations (SHAP) (Lundberg and Lee 2017). This framework enables us to identify the key predictors of live streaming effectiveness in professional settings and to assess whether existing theories generalize to this context.

INSTITUTIONAL SETTING

The online medical consultation market has experienced significant growth over the past decade and is projected to surpass USD 10 billion and reach 127 million users by 2025.[4] Despite this impressive valuation, the market's penetration rate remains low, at only 1.63% in 2025, highlighting substantial potential for growth. Our study focuses on the online medical consultation market in China, which is the second-largest market globally in consultation revenue and the largest in number of users utilizing online medical consultation services.

In China, most medical consultations are conducted in person at clinics or hospitals which are part of the traditional healthcare system. Since the early 2010s, technological advancements have facilitated the emergence of online platforms that connect physicians with patients for online consultations. These consultations are primarily used by patients to manage minor cases promptly or to seek second opinions for more serious conditions (e.g., oncology or cardiology concerns). Typically, these consultations occur through mobile apps developed by the platforms, through various formats such as chat-based consultations, app-enabled phone calls, and video consultations.

[4] Source: Online Doctor Consultations: Market Data & Analysis 2025 by Statista.

Our data are sourced from one of the largest healthcare platforms in China, Haodf.com. Established in 2006, the platform had approximately 285,000 registered physicians by 2024, representing over 10% of all licensed hospital-employed physicians nationwide. These physicians span 28 specialties, including general internal medicine, cardiology, pediatrics, and surgery. Haodf has served a cumulative total of 89 million patients, facilitating 49 million consultations in 2024 alone. To join the platform, physicians must have a full-time position in an offline clinic or hospital and hold the necessary professional licenses. The online consultations offered through Haodf are typically provided by physicians outside their regular working hours and beyond their full-time duties at clinics or hospitals.

Patients using the platform must first register and verify their identity before making a consultation request. On the platform, patients can view information about physicians, including their specialty, affiliated hospitals, educational background, availability,[5] and consultation fees. Unlike in-person consultations at hospitals where prices are regulated by the hospital and government, the fees for online consultations on the platform are set by individual physicians and are, on average, about three times higher than offline hospital prices. This additional revenue stream from online consultations serves as a key incentive for physicians to participate in the platform. The platform collects a commission from the physicians for each consultation via a fixed commission rate, although the rate is not observed by the researchers.

In addition to providing access to online consultations, Haodf also serves as a health knowledge hub in China, akin to a "Wikipedia" for healthcare. Physicians can write educational articles to share with the community, and the platform provides information on symptoms, diagnoses and treatments based on previous consultations. This knowledge-sharing feature attracts traffic from both patients seeking consultations and the general public looking for health education.

With the significant rise of live streaming e-commerce in China, which facilitated over

[5]Physicians can indicate if they are available for a consultation or not at any given point in time. However, if they indicate they are available, no details about the number and time of available slots are provided.

USD 5.4 trillion in transaction value in 2024, Haodf introduced a live streaming feature that allows physicians to engage with users.[6] Launched in July 2020, this feature enables physicians to conduct live streaming sessions directly within the platform's mobile app. These sessions are free for users, and physicians are encouraged by the platform to participate. Haodf provides information on upcoming live streams on its main platform, but does not provide any additional promotion for specific live streams.[7] Physicians use live streaming sessions to share medical knowledge, provide free consultations, or both.[8] Sessions typically focus on a specific disease or treatment topic and are interactive, allowing patients to ask questions in real time.

Live streaming offers clear benefits to patients by allowing them to ask medical questions and learn about health-related topics at no cost. For physicians, these sessions provide an opportunity to build trust with patients and gain insights into patient needs.[9] Reflecting the growing adoption of this channel, approximately 30,000 live streaming sessions had been conducted on the platform as of July 2024, attracting millions of patient participants.

Beyond these direct benefits, industry reports and other anecdotal evidence suggest that live streaming also serves as a strategic tool for physicians. By educating patients and demonstrating expertise, physicians can strengthen patient relationships and increase demand for their paid consultations (between a physician and a specific patient), thereby generating supplementary income. To check this, we carried out in-depth interviews with ten physicians on the platform around the pros and cons of live streaming on this platform (see details in Online Appendix Section J). Their responses confirmed that the main benefits of live streaming include improving patient welfare, expanding access to care, and earning supplementary

[6]Source: "China's Live Commerce Data Report 2024" by Statista.

[7]After a live streaming session ends, a recorded version is posted on the platform the next day. These recordings attract little attention, typically receiving fewer than 10 views on the first day and almost none within a week, compared with hundreds of views for a typical live stream.

[8]The live streaming sessions on Haodf are regulated by the platform to ensure quality. Physicians are prohibited from mentioning specific drug names but can offer general advice on medical care and treatments that patients may need in the future. Haodf does not sell medications; if a prescription is issued, patients fill it through external pharmacies or other offline outlets.

[9]"Why do physicians participate in online live streaming?" *NetEase*, accessed January 19, 2026.

income.

DATA

The objective of this study is to measure the impact of live streaming on the demand for physicians' paid online consultations. To achieve this, we collect two datasets: one containing consultation information and the other containing live streaming data. These datasets are then combined to form the final analysis sample.

For the consultation data, the data provider provides monthly information on individual physician attributes, including physician name, physician ID, affiliated hospital, professional title, specialty, cumulative number of consultations, consultation prices, patient ratings, and rankings.[10] For the live streaming data, we collect information on all live streaming sessions conducted on the platform, including the original video, physician ID, and timestamps. The sample period ranges from September 2020 to July 2024.

The two datasets are matched using physician ID and timestamps. Detailed data cleaning steps are provided in Web Appendix A. The final dataset includes 2,271 physicians who conducted at least one live streaming session during the sample period and 4,955 physicians who did not engage in any live streaming sessions during the same period.

Physician Characteristics

We first report information on physicians' professional titles, affiliated hospitals, and geographic locations for the two groups of physicians in Table 1. These characteristics rarely change during our sample period, so we report the most recent observation for each physician. Panel A summarizes physicians who conducted at least one live streaming session during the sample period (*streaming physicians*), while Panel B summarizes those who did not engage in any live streaming (*non-streaming physicians*).

Streaming physicians tend to hold more senior professional titles: 27% are Chief Physi-

[10] We do not have any patient-level characteristics.

Table 1: Physician Level Descriptive Statistics

	Obs	Mean	SD	Min	Max
Panel A: Streaming Physicians					
Chief Physician	2271	0.27	0.44	0.00	1.00
Associate Chief Physician	2271	0.38	0.49	0.00	1.00
Attending Physician	2271	0.28	0.45	0.00	1.00
Resident	2271	0.06	0.24	0.00	1.00
Other Professional Title	2271	0.01	0.08	0.00	1.00
Tertiary Hospital	2271	0.43	0.50	0.00	1.00
Non-teritiary Hospital	2271	0.57	0.50	0.00	1.00
Tier 1 City	2271	0.19	0.39	0.00	1.00
Tier 2 City	2271	0.55	0.50	0.00	1.00
Tier 3 City	2271	0.14	0.35	0.00	1.00
Other City	2271	0.12	0.32	0.00	1.00
Panel B: Non-Streaming Physicians					
Chief Physician	4955	0.19	0.39	0.00	1.00
Associate Chief Physician	4955	0.30	0.46	0.00	1.00
Attending Physician	4955	0.37	0.48	0.00	1.00
Resident	4955	0.10	0.30	0.00	1.00
Other Professional Title	4955	0.04	0.20	0.00	1.00
Tertiary Hospital	4955	0.38	0.49	0.00	1.00
Non-teritiary Hospital	4955	0.62	0.49	0.00	1.00
Tier 1 City	4955	0.25	0.43	0.00	1.00
Tier 2 City	4955	0.34	0.47	0.00	1.00
Tier 3 City	4955	0.16	0.37	0.00	1.00
Other City	4955	0.25	0.43	0.00	1.00

cians and 38% are Associate Chief Physicians compared to 19% and 30% among non-streaming physicians, respectively. In contrast, lower-ranking titles such as Attending Physician and Resident are more common among non-streaming physicians. Streaming physicians are also slightly more likely to be affiliated with tertiary hospitals (43% vs. 38%), which are large, highly reputable institutions that offer advanced medical care.

In terms of geographic distribution, streaming physicians are more concentrated in tier 2 cities (55%) compared to non-streaming physicians (34%), while non-streaming physicians are more prevalent in tier 1 and other cities.[11] Overall, we observe clear differences in characteristics between streaming and non-streaming physicians. We address these differences in the subsequent empirical strategy section.

Table 2: Physician-Month Level Descriptive Statistics

	Obs	Mean	SD	Min	Max
Panel A: Streaming Physicians					
Pre-Live Streaming					
Number of Consultations	11884	16.54	44.04	0.00	888.00
Consultation Price	11884	67.04	97.17	1.00	2000.00
Number of Educational Articles	11884	15.39	62.25	0.00	2940.00
Ranking Index	11884	3.48	0.47	2.10	5.00
Patients' Rating on Effectiveness	5416	99.16	3.54	50.00	100.00
Patients' Rating on Attitude	5416	99.36	3.35	50.00	100.00
Post-Live Streaming					
Number of Consultations	11343	17.87	46.75	0.00	843.00
Consultation Price	11343	69.25	95.33	1.00	2000.00
Number of Educational Articles	11343	14.77	92.44	0.00	4182.00
Ranking Index	11343	3.52	0.46	2.40	5.00
Patients' Rating on Effectiveness	4790	99.09	4.05	40.00	100.00
Patients' Rating on Attitude	4790	99.29	3.90	41.50	100.00
Panel B: Non-Streaming Physicians					
Number of Consultations	151918	14.72	45.43	0.00	988.00
Consultation Price	151918	57.12	157.27	0.50	9999.00
Number of Educational Articles	151918	11.34	73.22	0.00	3606.00
Ranking Index	151918	3.44	0.41	1.80	5.00
Patients' Rating on Effectiveness	81483	98.96	4.77	30.00	100.00
Patients' Rating on Attitude	81483	99.15	4.58	35.50	100.00

In Table 2, we report time-varying physician attributes measured at the physician-month level. Panel A presents summary statistics for streaming physicians during the six-month

[11] Tier 1 cities include Beijing, Shanghai, Guangzhou, and Shenzhen. Tier 2 cities refer to the "new first-tier cities" as defined by China Business Network, including Changsha, Chengdu, Chongqing, Dongguan, Foshan, Hangzhou, Hefei, Nanjing, Qingdao, Shenyang, Suzhou, Tianjin, Wuhan, Xi'an, and Zhengzhou. Tier 3 cities are provincial capital cities that are not part of the new first-tier list.

window before (pre-live streaming) and after (post-live streaming) their first live streaming sessions. We observe that, on average, the number of consultations per month increases from 16.54 before live streaming to 17.87 after. Consultation prices also rise, from an average of 67.04 RMB to 69.25 RMB, suggesting that physicians are likely to adjust pricing after gaining visibility through live streaming. The number of educational articles posted on physicians' personal pages decreases slightly from 15.39 to 14.77. The platform uses a non-personalized ranking index (ranging from 1 to 5) to determine the position of physicians in search results, where a higher value indicates a better ranking. The average ranking index increases slightly from 3.48 to 3.52 after live streaming. Patients' average ratings on consultation effectiveness and physicians' attitude remain at similar levels before and after live streaming, from 99.16 to 99.09 and from 99.36 to 99.29, respectively. These ratings are aggregated monthly as cumulative averages. Because ratings are voluntary, this measure is missing for physicians who receive no ratings in a given month, resulting in a smaller sample size.

Panel B of Table 2 reports the same metrics for non-streaming physicians over the full sample period. Compared to streaming physicians, non-streaming physicians have a slightly lower number of monthly consultations (14.72) and lower prices (57.12 RMB). They also post fewer educational articles on average (11.34) and have slightly lower ranking index (3.44). Patient ratings are comparable across the two groups, with both effectiveness and attitude ratings close to 99.

Live Streaming Video Data

We collect detailed information on physicians' live streaming sessions. Over the sample period, we assembled 23 TB of video data. For each session, we observe the streaming date, the associated physician ID, and the original video recording. We then use a multimodal analysis framework, informed by prior theory and enabled by machine learning and AI techniques, to extract features from three modalities: text, audio, and visual content. Details of this framework are presented later in the paper.

As noted in the Introduction, in professional settings such as healthcare, the impact of live streaming on demand is ambiguous. Live streaming may increase physician visibility, build trust through real-time interactions, and enhance perceived expertise. These factors could, in turn, stimulate consultation demand. On the other hand, the time required to prepare and conduct live streams may divert attention from patient care, potentially affecting service quality. Moreover, live streaming may create perceptions of over-commercialization or self-promotion, which can undermine a physician's credibility. Given these potentially contrasting outcomes, empirical analysis is needed to assess live streaming's impact on service demand in professional contexts.

Empirical Strategy

To estimate the effect of live streaming adoption on physicians' consultation demand, a natural starting point is to compare the pre- and post-treatment trends between streaming and non-streaming physicians using a difference-in-differences (DiD) strategy. However, the challenge is that the decision to adopt live streaming is endogenously made by individual physicians. As a result, the parallel trends assumption underlying DiD may not hold if there are unobserved confounders that affect the two groups differently over time.[12] To address this concern, we build on the synthetic control approach of Abadie and Gardeazabal (2003); Abadie, Diamond, and Hainmueller (2010) and implement the generalized synthetic control (GSC) method proposed by Xu (2017), which is well suited to settings with multiple treated units and unobserved interactive fixed effects.

Formally, we use the following specification:

$$y_{it} = \delta_{it} * Streaming_{it} + \lambda_i' f_t + \alpha_i + \xi_t + \epsilon_{it} \tag{1}$$

[12] Online Appendix Figure W1 presents model-free trends in service quantity for treated and control physicians. While the patterns suggest a positive impact of live streaming adoption on service demand, the parallel trends assumption does not fully hold in the pre-treatment period.

where y_{it} represents the service quantity by physician i in month t. The term $Streaming_{it}$ indicates the live streaming adoption status of physician i in month t. δ_{it} is the treatment effect of interest, which measures the individual impact of live streaming on physician i's service quantity in month t.

What sets the generalized synthetic control approach apart from standard DiD is the interactive fixed effects term $\lambda_i' f_t$. Here, $f_t = [f_{1t}, \ldots, f_{rt}]'$ is an $r \times 1$ vector representing unobserved factors common across units in month t, estimated using the control group data. $\lambda_i = [\lambda_{i1}, \ldots, \lambda_{ir}]'$ is an $r \times 1$ vector of unknown factor loadings specific to physician i, estimated using the pre-treatment data. The number of factors r is determined by cross-validation. The detailed estimation steps follow Section 3 of Xu (2017).[13]

Intuitively, the inclusion of the interactive fixed effects term allows us to account for unobserved, time-varying confounders that influence all physicians but with heterogeneous intensities. In this framework, f_t can be viewed as capturing common latent factors—such as seasonal fluctuations in patient demand or broader market conditions—that vary over time but are not directly observed. The corresponding factor loadings λ_i reflect the degree to which each physician is affected by these latent factors. By combining information from the control group and the pre-treatment observations of the treated group to estimate the latent factor structure, the GSC approach enables the construction of credible counterfactual outcomes even when the parallel trends assumption does not hold.

We further include additive two-way fixed effects, α_i for physician fixed effects and δ_t for year-month fixed effects, to control for time-invariant heterogeneity across physicians and for temporal shocks common to all physicians, respectively.[14] Finally, ϵ_{it} denotes the idiosyncratic error term, assumed to be independent of all explanatory variables.

The GSC framework allows for rich unobserved time-varying confounding through com-

[13] When we estimate the impact of live streaming on different outcome variables (e.g., consultation demand, consultation prices, patient review ratings, platform ranking scores) f_t and λ_i are estimated individually for each outcome variable.

[14] The additive physician and time fixed effects can be viewed as special cases of the unobserved factor structure by setting $f_{1t} = 1$ and $\lambda_{i2} = 1$, and rewriting $\lambda_{i1} = \alpha_i$ and $f_{2t} = \xi_t$. We include these terms explicitly to ensure that the two-way fixed effects are always incorporated in the specification.

mon latent factors with heterogeneous loadings and delivers a close pretreatment fit between treated physicians and their synthetic controls. As with other evaluations of endogenous adoption decisions using observational data, the resulting estimates should be interpreted as causal under the maintained assumption that, conditional on the controls and latent factor structure captured by the model, adoption timing is not driven by physician-specific time-varying shocks that also directly affect consultation demand (e.g., changes in capacity constraint/availability). We assess the sensitivity of our findings to such concerns in the robustness check section.

Based on this approach, let $T_{0,i}$ denote the last non-streaming month for treated physician i. Then, for the set of N live streaming physicians L, the average treatment effect on the treated in post-treatment month t, $A\hat{T}T_t$, is calculated as the average difference between the observed outcome for a treated physician, $Y_{it}(1)$, and its synthetic-control counterfactual, $\hat{Y}_{it}(0)$:

$$A\hat{T}T_t = \frac{1}{N} \sum_{i \in L, t > T_{0,i}} [Y_{it}(1) - \hat{Y}_{it}(0)] = \frac{1}{N} \sum_{i \in L, t > T_{0,i}} \hat{\delta}_{it} \tag{2}$$

where the standard errors for $A\hat{T}T_t$ are computed using parametric bootstrapping.

Unlike the parallel trends assumption required in the DiD approach, the GSC method assumes that unobserved time-varying confounders can be represented by a factor model that can be consistently estimated from untreated data. As shown in Online Appendix Figure W2, after accounting for the factor model, the treated group and the synthetic control group exhibit a close match during the pre-treatment period. The same figure also supports the no-anticipation assumption because it shows no systematic behavioral changes among treated physicians prior to live streaming adoption. Finally, GSC relies on the Stable Unit Treatment Value Assumption (SUTVA), which in this context implies that live streaming by treated physicians does not spill over to affect the outcomes of non-streaming physicians. While SUTVA cannot be tested directly, we conduct robustness checks later in this section to assess the sensitivity of our results to potential violations.

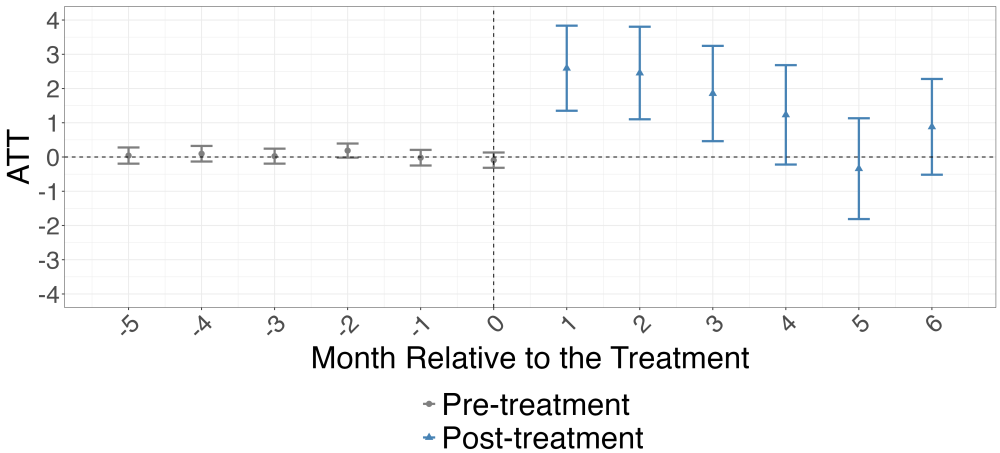


Figure 1: Impact of Live Streaming on Number of Services

Notes: This figure shows the estimated ATT on the total number of services and their 95% confidence intervals for each month before and after live streaming adoption month. The estimation includes physician fixed effects and calendar year-by-month fixed effects. The standard errors are based on parametric bootstraps of 1,000 times.

Main Results

In this section, we detail our findings on how live streaming affects physicians' online consultation demand. Figure 1 reports the estimated average treatment effects of live streaming adoption on monthly service volume. The x-axis shows months relative to the adoption of live streaming ($t = 1$), and the y-axis shows the estimated treatment effects over a six-month window before and after adoption with bootstrapped 95% confidence intervals. Pretreatment estimates near zero suggest a good match between synthetic control physicians and treated physicians, indicating that the GSC method effectively accounts for both observed and unobserved attributes.

The results suggest that the adoption of live streaming leads to an immediate and statistically significant increase in consultations with effects persisting for about three months. In the month immediately following adoption, physicians conduct 2.593 more consultations (s.e. = 0.634), i.e., a 16% increase; in the following month, the increase is 2.453 (s.e. = 0.689), about a 15% increase; by the third month, the increase is 1.854 (s.e. = 0.710), about an 11%

increase. From the fourth month onward, the difference becomes statistically insignificant.[15]

Robustness Checks

We conduct six additional exercises to assess the robustness of our findings around the impact of live streaming adoption on service demand.

First, we examine whether our results are sensitive to alternative estimation strategies. Instead of the GSC approach used in the main analysis, we re-estimate the effect using the following four methods: (1) two-way fixed effects model, (2) Callaway and Sant'Anna Difference-in-Differences (CS-DID) estimator (Callaway and Sant'Anna 2021), (3) Sun and Abraham Difference-in-Differences (SA-DID) estimator (Sun and Abraham 2021), and (4) Borusyak, Jaravel, and Spiess Difference-in-Difference (BJS-DID) estimator (Borusyak, Jaravel, and Spiess 2024). The results, reported in Online Appendix Table W1, are consistent with our main estimates in terms of both magnitude and statistical significance.

Second, we address potential violations of the Stable Unit Treatment Value Assumption (SUTVA). If spillovers from streaming physicians to non-streaming physicians exist, they are most likely to occur among physicians in the same specialty with similar platform rankings (Zhan, Zhang, and Fu 2026). To mitigate this concern, we exclude from the control group any physicians who share the same specialty and are ranked close to streaming physicians. The results, presented in Online Appendix Figure W4, indicate that the positive impact of live streaming remains robust in both magnitude and statistical significance. Another plausible spillover channel is hospital affiliation. To test robustness along this dimension, we exclude control physicians from the same hospital as streaming physicians. As shown in Online Appendix Figure W5, the results again confirm the robustness of our main findings.

Third, we test the sensitivity of our findings to the estimation window. The main spec-

[15] Live streaming may also affect physicians' offline consultation demand. Although we do not observe offline consultation volume for all physicians in our sample, we use a proxy for a subsample of physicians and find that the effect of live streaming on offline consultations is not statistically distinguishable from zero. This may be because most patients use the platform to access physicians outside their local market (because live streaming reaches an audience beyond the physician's local market) or because our proxy is noisy. A more detailed discussion is provided in Online Appendix C.

ification uses a six-month period before and after live streaming adoption. We adjust this window to three, nine, and twelve months. The results, reported in Online Appendix Figures W6, Figure W7, and Figure W8, remain broadly consistent with the main findings.

Fourth, we assess whether our estimated treatment effect is driven by pre-existing demand trends rather than by live streaming adoption itself. In particular, physicians who later adopt live streaming may already be experiencing increasing demand prior to adoption. We address this concern in two ways. We first conduct a timing placebo test in which, for adopting physicians, we artificially move the adoption date one to six months earlier and re-estimate the model. We do not find comparable effects under these placebo treatment dates, suggesting that our main findings are unlikely to be driven by differential pre-trends, gradual underlying trends, or anticipatory behavioral responses prior to actual adoption. Second, we randomly assign treatment status to 2,271 physicians from the control group and re-estimate the model. This placebo exercise also yields no comparable effects, reducing the concern that our results are mechanically generated by the empirical specification or by chance correlations in treatment assignment. Details are reported in Appendix D, Figures W9 and W10.

Fifth, we examine whether our findings could be explained by unobserved changes in physicians' capacity or time availability rather than by live streaming adoption itself. Any time-invariant differences in capacity between adopters and non-adopters are absorbed by physician fixed effects. The more relevant concern is a time-varying increase in capacity that both triggers adoption and enables physicians to handle more consultations. We view this explanation as unlikely for several reasons. First, consultations on the platform are initiated by patients, so physicians cannot mechanically generate more consultations simply by increasing supply. Second, the platform-wide consultation request acceptance rate is consistently high at 93–95%, and most rejected requests reflect mismatch in specialty rather than limited physician capacity, suggesting that capacity is generally not the binding constraint. Third, if physicians have additional available time, allocating it to paid consultations would often be more immediately remunerative than conducting free live streams, especially if they already face unmet consultation demand. Finally, in our physician interviews, none cited increased capacity or additional free time as a reason for adopting live streaming; instead, they emphasized patient reach and visibility. Taken together, these considerations make it unlikely that changes in physician capacity, rather than live streaming adoption, explain our results.

Sixth, we examine the sensitivity of our findings to alternative data-cleaning choices, including the missing-at-random check, the treatment of physicians with missing first-month consultation data, the treatment of physicians with missing video features, the inclusion of all non-streaming physicians without pre-selection, and the exclusion of physicians who conduct live streams on other platforms. The results, reported in Online Appendix Section E, confirm the robustness of our main findings.

Alternative Pathways

The previous analysis reports the overall effect of live streaming on service demand. Given the dynamic platform setting, we next examine if other pathways, along with live streaming adoption, could increase demand. Live streaming adoption can influence both platform-side exposure and physician-side practice. In particular, live streaming may (i) improve a physician's visibility on the platform (ranking), (ii) shape patients' perceived quality (ratings) of physicians, (iii) affect pricing decisions, and (iv) encourage the creation of additional informational content (patient-education articles). Examining whether these factors move with adoption helps clarify how live streaming translates into higher service demand.

We take two complementary steps to examine these pathways. First, we re-estimate the post-adoption ATT for service demand while additionally controlling for observed time-varying measures of ranking, ratings, price, and patient-education article production. Because these variables may themselves respond to adoption, they should not be interpreted as standard exogenous controls. Nevertheless, examining the sensitivity of the estimated ATT

Table 3: Test for Alternative Pathways

	(1) Baseline	(2) Control Ranking	(3) Control Ratings	(4) Control Price	(5) Control Articles	(6) Control All
ATT (Month 1)	**2.593**	**2.717**	**2.901**	**2.766**	**2.756**	**2.689**
	(0.634)	(0.628)	(0.660)	(0.627)	(0.680)	(0.653)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
ATT (Month 2)	**2.453**	**2.590**	**2.859**	**2.689**	**2.677**	**2.556**
	(0.689)	(0.715)	(0.731)	(0.764)	(0.751)	(0.676)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
ATT (Month 3)	**1.854**	**1.699**	**2.162**	**2.020**	**2.006**	**1.658**
	(0.710)	(0.682)	(0.722)	(0.768)	(0.745)	(0.711)
	[0.009]	[0.013]	[0.003]	[0.009]	[0.007]	[0.020]
ATT (Month 4)	1.231	1.031	**1.503**	1.337	1.321	0.983
	(0.741)	(0.713)	(0.752)	(0.770)	(0.752)	(0.743)
	[0.096]	[0.148]	[0.046]	[0.082]	[0.079]	[0.186]
ATT (Month 5)	-0.340	-0.453	0.029	-0.170	-0.187	-0.499
	(0.751)	(0.714)	(0.826)	(0.766)	(0.743)	(0.776)
	[0.650]	[0.526]	[0.972]	[0.825]	[0.801]	[0.520]
ATT (Month 6)	0.881	0.830	1.378	1.167	1.148	0.782
	(0.713)	(0.640)	(0.755)	(0.728)	(0.754)	(0.708)
	[0.217]	[0.195]	[0.068]	[0.109]	[0.128]	[0.270]

Notes: This table presents results from tests of alternative pathways. Column (1) shows the baseline specification, estimating the impact of live streaming on consultation demand with physician and time fixed effects. Columns (2)–(5) each augment this baseline with one additional control: ranking in column (2), patients' ratings in column (3), price in column (4), and the number of educational articles written by the physician in column (5). The ATT remains stable across all these specifications. Because ratings are missing for some physicians, columns (3) and (6) include a missing-value indicator (ratings set to zero when missing) to keep the sample size constant. Column (6) adds the full set of controls—ranking, patients' ratings, price, and number of educational articles—and the ATT continues to hold. Standard errors in parentheses are based on parametric bootstraps of 1,000 times. P-values are reported in square brackets below the standard errors. Coefficients significant at the 5% level are in bold.

to their inclusion is informative about whether the main result is materially affected by contemporaneous movements in these observed measures. Second, we estimate the effect of live streaming adoption on each of these candidate pathways using the same GSC framework.

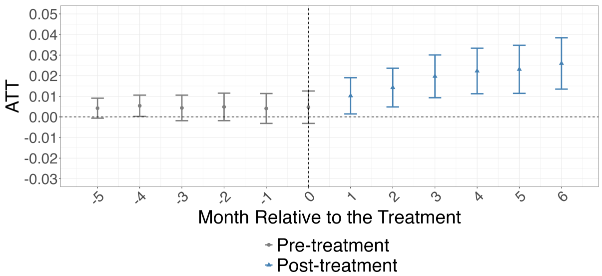
Table 3 reports the first set of results. Column (1) reproduces the baseline estimate without additional controls. Columns (2) through (5) add ranking, ratings, price, and article counts one at a time, and column (6) includes all four simultaneously. The estimated demand effect remains similar across specifications, suggesting that the main result is not highly sensitive to conditioning on these observed time-varying measures.

We next examine whether adoption is associated with changes in these candidate pathways. Figure 2 reports the corresponding GSC estimates for ranking, ratings, price, and article counts. Among the four outcomes, we find a statistically significant positive effect only for physicians' ranking, where a higher value indicates a better ranking in search results.[16] This pattern suggests that ranking is the most salient observed pathway through which live streaming may additionally affect demand (Jeziorski, Leng, and Seiler 2025). Because ranking is a post-adoption variable and may be endogenous, controlling for it is not part of our main identification strategy and should be interpreted only as a sensitivity exercise rather than a clean estimate of the direct effect. By contrast, the baseline specification captures the overall effect of adoption, including any effect operating through ranking. Reassuringly, the ATT estimates with and without ranking are very similar. Bootstrap tests show that the differences are not statistically significant for Months 1, 2, and 3, nor for the average ATT over the first six months (see Appendix D.5 and Table W2). Overall, we do not find evidence that live streaming adoption systematically changes ratings, price, or article production, and although adoption improves ranking, accounting for this observed pathway does not materially alter the estimated treatment effect.

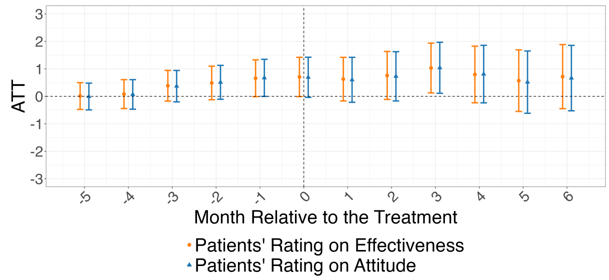
HETEROGENEITY

The preceding analysis estimates the average effect of live streaming adoption on service demand across all streamers. We examine heterogeneity along characteristics observed prior to adoption as they can be used by the platform to prioritize outreach and training toward physician profiles that are more likely to benefit from live streaming. Specifically, we estimate subgroup-specific effects along four dimensions: physician seniority (chief and associate chief versus non-chief), hospital tier (tertiary versus non-tertiary), baseline consultation volume (above versus below the median), and specialty (18 categories). We report these results in Figure 3.

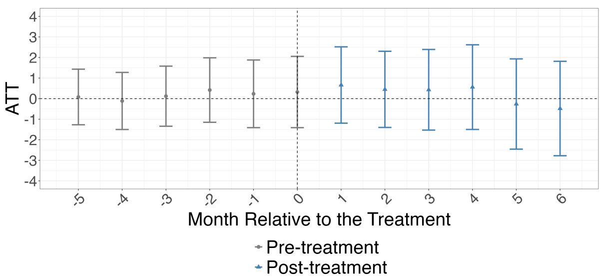
[16] Note that live streaming is not a direct input into the platform's ranking algorithm. However, live streaming may increase a physician's visibility on the platform, for example through greater page traffic and more consultations, which may in turn improve ranking over time.



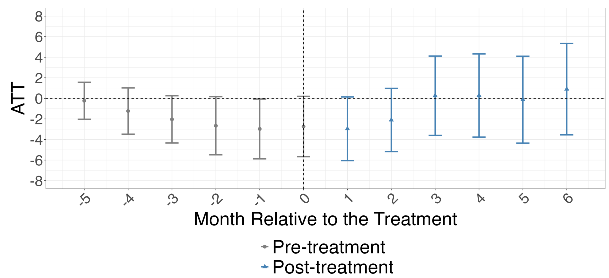
(a) Ranking



(b) Ratings



(c) Price



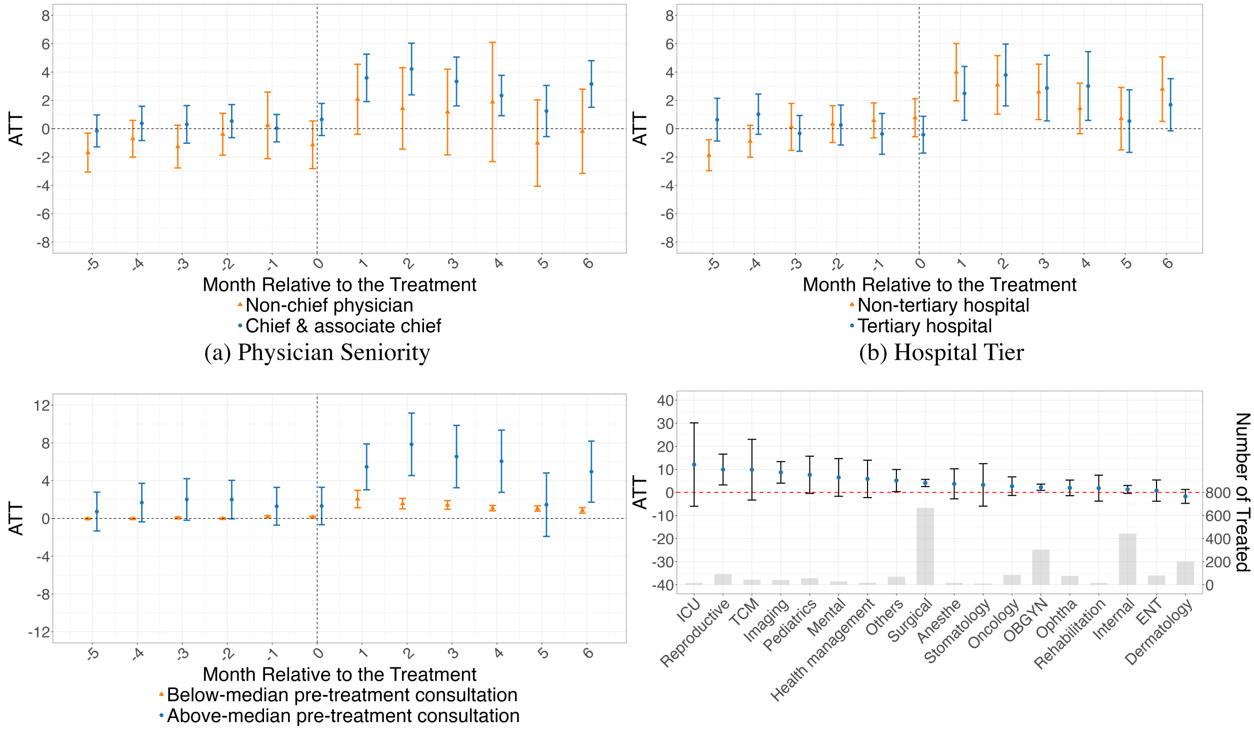
(d) Article Counts

Figure 2: Effect of Live Streaming Adoption on Alternative Outcome Variables

Notes: This figure presents the estimated average treatment effect on the treated (ATT) of live streaming adoption on four alternative outcome variables using the generalized synthetic control method. Panel (a) shows the effect on physicians' ranking index (ranging from 0 to 5 and higher value indicates better ranking), Panel (b) on patient rating, Panel (c) on consultation price, and Panel (d) on the number of patient educational articles. The shaded areas represent 95% confidence intervals based on parametric bootstraps of 1,000 iterations.

Seniority. Panel (a) presents the results by physician seniority. The demand effects are concentrated among senior physicians. Chief and associate chief physicians exhibit a clear, statistically significant, and persistent post-adoption increase in service demand, whereas non-chief physicians display weaker and less precisely estimated effects. This pattern suggests that live streaming may be more effective when physicians have stronger pre-existing reputation or credibility, which can help convert live streaming engagement into additional consultations.

Hospital tier. Panel (b) shows the results by hospital tier. We observe positive post-adoption effects for physicians in both tertiary and non-tertiary hospitals. The estimated



(a) Physician Seniority

(b) Hospital Tier

(c) Baseline Consultation Volume

(d) Specialty

Figure 3: Heterogeneity by Pre-Adoption Characteristics

Notes: Panels (a)–(c) show the estimated ATT (with 95% confidence intervals) on the number of services for each month before and after live streaming adoption. Panel (a) compares chief/associate chief physicians ($N = 1{,}671$) with non-chief physicians ($N = 600$). Panel (b) compares physicians in tertiary hospitals ($N = 1{,}171$) with those in non-tertiary hospitals ($N = 1{,}110$). Panel (c) compares physicians with above-median ($N = 1{,}136$) and below-median ($N = 1{,}135$) pre-treatment consultation volume (split at the median). The standard errors are based on parametric bootstraps of 1,000 iterations. Panel (d) shows the average estimated ATT for the first three months after the first live streaming session (with 95% confidence intervals) on the number of services by physician specialty.

effects are broadly similar in direction and magnitude across the two groups, suggesting that the demand benefits of live streaming do not differ substantially by hospital tier.

Baseline consultation volume. Panel (c) presents the results by pre-treatment baseline consultation volume. Heterogeneity by baseline demand is pronounced. Physicians with high pre-treatment consultation volume experience substantially larger post-adoption absolute gains, with effects that are both stronger and more persistent than those for physicians with low pre-treatment volume. In contrast, the low-volume group exhibits smaller increases.

This pattern is consistent with complementarities between live streaming and existing demand: physicians with an established patient base or higher baseline visibility may translate streaming exposure into consultations more effectively.

Specialty. Finally, we examine heterogeneity across specialties, as shown in Panel (d). The dots show the average estimated ATT over the first three months after live streaming adoption for 18 specialties, with the scale shown on the left axis. The gray bars at the bottom show the number of streaming physicians in each specialty, with the scale shown on the right axis. Point estimates are positive for most specialties, although precision varies with the number of treated physicians. Some specialties exhibit larger point estimates, while others have wider confidence intervals because of smaller treated sample sizes. Overall, the results suggest broadly positive effects across specialties.

Taken together, the heterogeneity results indicate that the demand benefits of live streaming are strongest among physicians who are already relatively established on the platform (higher seniority and higher baseline volume). Because these dimensions are observable prior to adoption, they provide actionable guidance for platform outreach and training efforts.

WHAT PREDICTS THE EFFECTIVENESS OF LIVE STREAMING?

While the average effect of live streaming adoption on service demand is positive, our heterogeneity analyses indicate that estimated effects vary systematically across physicians. In particular, the gains are larger for more established physicians (chief/associate chief and those with higher baseline consultation volume). These patterns raise an important practical question for the platform: Which observable aspects of live streaming activity are systematically associated with stronger demand responses and may thus help identify or support more effective streaming practices?

Motivated by the video-centric nature of live streaming, we provide a comprehensive characterization of the video attributes that may matter for live streaming effectiveness.

Our goal is to map a broad set of video attributes to downstream demand responses and to identify which dimensions are most strongly associated with performance. In contrast, most prior studies of live streaming focus on a single attribute dimension or only one modality in isolation (Chen et al. 2025; Lin, Yao, and Chen 2021; Luo et al. 2025; Ma et al. 2024; Tang, Hao, and Li 2023; Xu et al. 2023). This broader approach is important because live streams are inherently multimodal and high-dimensional: multiple attributes co-occur and may interact, so focusing on only one dimension or one modality may overlook other relevant sources of variation. Moreover, from a practical perspective, creators are unlikely to adjust one dimension of a live stream without simultaneously affecting others. Taken together, these considerations make a comprehensive analysis of video attributes important for understanding their predictive relationship with performance.

The professional service context also differs in ways that may shape which attributes matter. Professional influencers, such as physicians, establish trust primarily through formal credentials, demonstrated expertise, and domain (medical) knowledge. In contrast, typical influencers often build trust through perceived relatability, using aspirational or entertaining information, which leads to differences in the type of content presented. In addition, professional influencers aim to promote their own services and increase awareness, whereas many typical influencers monetize by promoting third-party brands and frequently use time-limited price discounts to encourage immediate purchases (Liu 2023). This creates differences in how information is communicated and how audiences are convinced. Finally, most prior studies have examined pre-recorded influencer videos (Rajaram and Manchanda 2020; Yang, Zhang, and Zhang 2025; Zhang, Qiu, and Ye 2025) rather than live streaming, which involves real-time interaction and may operate through different mechanisms. Despite the rapid growth of professional live streaming, systematic evidence on which video attributes are associated with effectiveness in professional settings remains scarce. We fill this gap by assembling and measuring a rich set of attributes and documenting how they are associated with demand responses following adoption.

To do so, we identify eight categories of features motivated by prior literature in marketing, psychology, organizational behavior, public health, and information systems, and adapt them to the context of professional live streaming. We then introduce a multimodal machine learning framework to extract these features from the video, audio, and textual components of the streams. A key advantage of this approach is its flexibility: rather than imposing a parsimonious specification ex ante, it allows us to incorporate a broad and expandable set of feature categories. Although we focus on eight categories in the current analysis, the framework can readily accommodate additional categories and features. Finally, we present evidence on which extracted features are most strongly associated with demand responses following adoption.

Theoretically-Informed Live Streaming Feature Categories

Building on established theories and empirical findings, we define eight categories of features that capture different dimensions of live streaming videos in the telehealth setting. Each category is grounded in relevant prior research yet tailored to the unique context of professional live streaming.

Audience Interaction. The first category captures audience interaction during the live streaming session, a key mechanism through which live streaming can affect outcomes. Prior work shows that higher audience engagement increases value creation in entertainment live streaming (Lu et al. 2021) and that audience availability varies systematically over the course of the day (Qu et al. 2025). Accordingly, we operationalize audience interaction along two dimensions. We capture audience availability using the session's start time, measured in decimal hours (*session time*). We capture realized physician–patient interaction during the session using three measures: the total length of physicians' responses to patient questions in words (*answer length*), the number of patient questions answered (*Q&A count*), and the share of physicians' speech devoted to answering patient questions (*Q&A session share*).

Session Content. The second category focuses on the substantive content of the live stream. Prior evidence from television advertising shows that content attributes influence ad effectiveness (Guitart and Stremersch 2021; Jiang and Kim 2024). Related evidence from influencer marketing and entertainment live streaming indicates that content tied to product engagement and explicit audience recognition can increase sales and revenue (Lu et al. 2021; Yang, Zhang, and Zhang 2025). In our professional telehealth context, we use Large Language Models to classify each Q&A exchange during the live session into one of the following five categories: (i) disease diagnosis advice (*disease Q&As*), (ii) disease treatment advice (*treatment Q&As*), (iii) drug usage advice (*drug Q&As*), (iv) preventive care advice (*prevention Q&As*), and (v) other types of medical advice not covered by the above categories (*other Q&As*).

Professional Appearance. The third category captures the perceived professionalism conveyed by the physician's appearance, a factor often overlooked in influencer marketing and video promotion research. When professionals act as influencers, their appearance can shape audience trust and engagement. Prior work shows that professional appearance is valued across fields such as law (Abel 1979), education (Pamuji and Limei 2023), service (Yohana, Akbar, and Suparjo 2020), the public sector (Öberg and Bringselius 2015), and celebrity culture (Feng et al. 2025). In healthcare, medical uniforms and business suits have been linked to greater patient trust in offline visits (Petrilli et al. 2018; Rehman et al. 2005). We measure physicians' appearance through three dimensions. First, for professional attire, we assess whether the physician wears medical clothing (e.g., a white coat or surgical gown; *medical uniform*) and whether the physician wears a suit (*formal clothing*). Second, for personal protective equipment (PPE), we note that our sample period covers the Covid-19 pandemic,[17] during which physicians may be perceived as more careful and professional when wearing PPE. We therefore measure the proportion of time the physician wears a mask (*mask session share*) and the proportion of time they wear a medical hat (*hat session share*). Third,

[17] In the regressions, we include year-month fixed effects to account for Covid-related time trends.

for facial appearance, we include the proportion of time they wear glasses (*glass session share*) and a measure of the physician's facial attractiveness (*facial attractiveness*).

Emotional Expression. The fourth category draws on research in consumer behavior, organizational behavior, and psychology, which demonstrates that emotions can be contagious (Barsade 2002; Hatfield, Cacioppo, and Rapson 1992), particularly in real-time interactive environments such as live streaming. In professional live streaming contexts, however, the role of emotional expression remains ambiguous. Studies in influencer marketing suggest that greater emotional expressiveness, may benefit influencers (Cascio Rizzo et al. 2024; Lin, Yao, and Chen 2021). In contrast, research on offline healthcare consultations finds that patients may prefer physicians to maintain a neutral demeanor during interactions (Hall, Roter, and Rand 1981; Oe and Murata 2019). We measure emotional expression across three modalities. First, for audio features, we calculate the proportion of sentences spoken by the physician in a neutral tone (*neutral tone session share*), and the variance in emotional tone across all sentences of the physician's speech (*audio emotion variance*). Second, for facial features, we compute the proportion of time the physician displays a neutral expression (*neutral face session share*), and the variance in facial emotions (*facial emotion variance*). Third, for text-based features, we use widely adopted sentiment analysis methods to measure the average sentiment of the physician's speech (*text sentiment*).

Linguistic Features. The fifth category captures the linguistic characteristics of physician communication in live streaming sessions. Prior research shows that communication style significantly shapes patient satisfaction: patients prefer less medical jargon (Allen et al. 2023; Thomas et al. 2014), more concrete language (Wang, Zhang, and Meng 2025), and greater linguistic alignment between physician and patient. In addition, readability has been linked to reduced communication frictions (Crossley et al. 2020). Accordingly, we measure four linguistic features: (i) the complexity of physician responses (*answer complexity*), (ii) the concreteness of language (*answer concreteness*), (iii) the number of medical jargon words (*medical jargon count*), and (iv) linguistic similarity between physician responses and patient questions (*Q&A similarity*).

Acoustic-Prosodic Features. The sixth category captures audio presentation attributes of the live stream. In healthcare contexts, Liu et al. (2020) show that patients prefer physicians with higher voice pitch, while Liu, Si, and Gao (2022) report preferences for faster speech rates. From the audio track, we extract four sets of features: (i) the physician's average speech rate (*speech rate*), (ii) the mean and variance of voice loudness (*average voice volume*, *voice volume variance*), (iii) the mean and variance of voice pitch (*average voice pitch*, *voice pitch variance*), and (iv) the perceived timbre of the physician's speech measured by average central position of the speech's spectrum (*average spectral centroid*).

Visual-Composition Features. The seventh category captures the visual presentation of the live stream (Brucks, Rifkin, and Johnson 2025). Specifically, we measure (i) average distance between the physician's face and the center of the picture (*face distance*); (ii) the pitch, roll, and yaw of the physician's face, capturing whether the physician faces the camera directly or at an angle (*face pitch angle*, *face yaw angle*, *face roll angle*); whether the physician appears to be looking into the camera, measured by the average eye gaze angle (*average eye gaze theta*) (the smaller this value is, the physician is looking at the camera) and the proportion of video frames in which the physician's facial information cannot be detected (*invalid frame ratio*).

Promotional Strategies The final category captures physicians' promotional behaviors during live streaming. In typical influencer settings, where influencers promote third-party products, promotional activities such as product mentions (Yang, Zhang, and Zhang 2025) or coupon offerings (Liu 2023) are essential for generating revenue. At the same time, however, prior studies show that promoting sponsored products can harm influencers' reputations because audiences often prefer organic content without commercial intention (Cheng and

Zhang 2025). In professional live streaming, by contrast, physicians are simultaneously educating audiences about medical knowledge and raising awareness of their own services rather than promoting third-party products. As a result, the impact of promotional behavior may differ in this context. We construct two measures of promotional strategies: (i) the number of discount vouchers the physician distributes for free consultations (*voucher count*), and (ii) the number of times the physician recommends that patients get a one-on-one consultation (*service recommendation*).

A Multimodal Machine Learning Framework for Feature Extraction

To construct the eight categories of features, we implement a multimodal machine learning framework (Hu and Ma 2025), illustrated in Figure 4. The process begins with separating the video files into their constituent modalities, including audio, text, and visual content.

Using the Python package *MoviePy*, we extract the audio track from each video to serve as the basis for acoustic feature analysis, (see Web Appendix G for technical details). The audio is then transcribed into text using the deep learning model *SenseVoice* (An et al. 2024), which generates word-level transcripts with punctuation. Because our setting involves extensive medical terminology and potential background noise, we refine the transcriptions using the large language model *DeepSeek-V3* to improve recognition accuracy for domain-specific terms (see Web Appendix F for the full prompt) . The resulting transcripts serve as the basis for textual analysis. Lastly, for the visual modality, we extract one frame per minute from each streaming session. Given that the average session lasts about one hour, this yields approximately 60 images per live stream for visual-feature analysis (see Web Appendix H for technical details). Once the modality-specific data are obtained, we perform modality-specific analyses for text, audio, and visual features.

Tables 4 and 5 report descriptive statistics for the video features of live streaming sessions conducted by the 2,271 treated physicians in their first month after adoption.[18] We focus on

[18] Among these physicians, 2,042 conducted only one live streaming session in their first month. For the remaining 229 physicians who conducted multiple sessions, we average feature values across all sessions conducted in the first

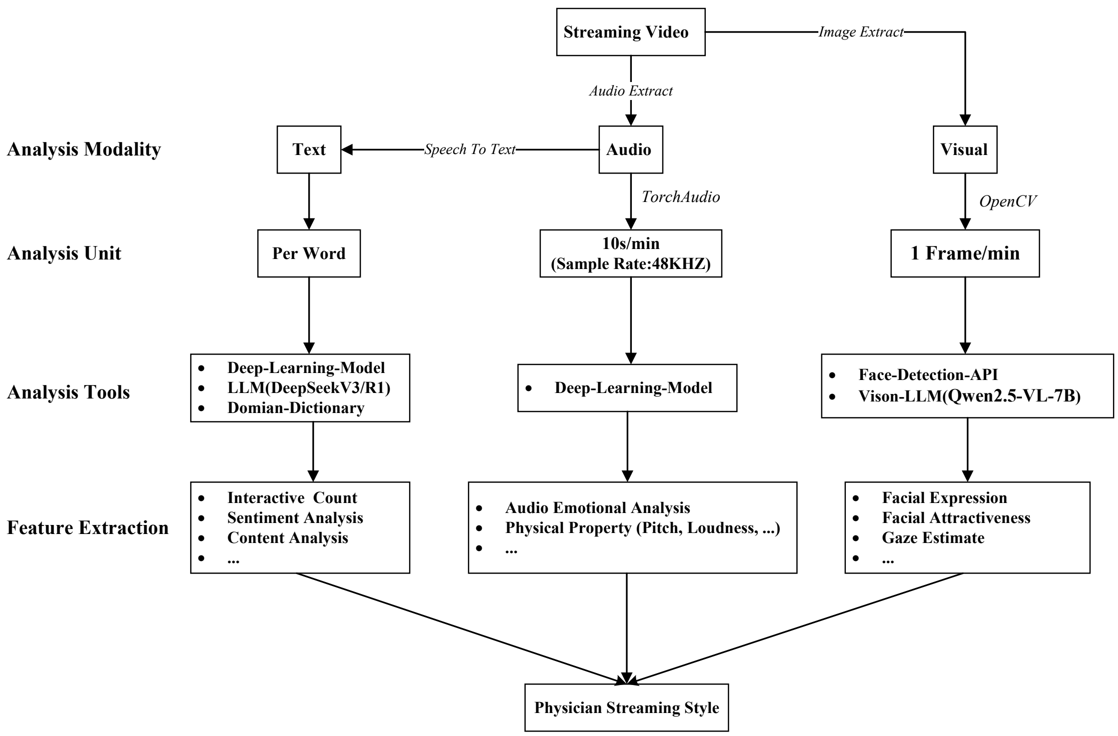


Figure 4: Video Data Processing Framework

this initial window to capture physicians' early streaming practices before cumulative exposure and evolving behavior potentially shape later outcomes. On average, a live streaming session includes 30.59 rounds of questions and answers with 10.03 focused on disease diagnosis, 10.46 on disease treatment, 4.02 on drug usage, 3.75 on preventive care, and 2.32 on other topics. Live streaming sessions begin as early as 8:00 a.m. and as late as 10:30 p.m. with an average start time of 6:00 p.m. Physicians devote, on average, 32% of session time to answering patient questions. In terms of appearance, 47% of physicians wear medical attire and 19% wear professional clothing during live streaming sessions. Only 4% wear a mask or hat, while 67% wear glasses.

For emotional expression, 97% times physicians speak in a neutral tone with some variation when physicians sometimes speak happily, angrily, or sadly. Similarly, 96% of the time physicians have a neutral face with some variation on a happy or angry or sad face. The mean textual sentiment score is 0.15 on a scale from -0.71 (negative) to 0.88 (positive),

month.

Table 4: Video Features Descriptive Statistics - Part 1

	Obs	Mean	SD	Min	Max
Panel A: Audience Interaction					
Session time	2271	17.70	2.85	8.00	22.48
Answer length	2271	126.50	46.69	12.00	627.50
Q&A count	2271	30.59	23.04	1.00	301.00
Q&A session share	2271	0.32	0.14	0.00	1.00
Panel B: Session Content					
Disease Q&As	2271	10.03	9.05	0.00	106.14
Treatment Q&As	2271	10.46	9.31	0.00	93.29
Drug Q&As	2271	4.02	5.37	0.00	58.00
Prevention Q&As	2271	3.75	4.80	0.00	66.50
Other Q&As	2271	2.32	3.38	0.00	57.00
Panel C: Professional Appearance					
Medical uniform	2271	0.47	0.49	0.00	1.00
Formal clothing	2271	0.19	0.39	0.00	1.00
Mask session share	2271	0.04	0.19	0.00	1.00
Hat session share	2271	0.04	0.19	0.00	1.00
Glass session share	2271	0.67	0.47	0.00	1.00
Facial attractiveness	2271	49.07	8.89	22.74	78.19
Panel D: Emotional Expression					
Neutral tone session share	2271	0.97	0.08	0.00	1.00
Audio emotion variance	2271	0.03	0.05	0.00	0.52
Neutral face session share	2271	0.96	0.09	0.00	1.00
Face emotion variance	2271	0.03	0.05	0.00	0.41
Text sentiment	2271	0.15	0.23	-0.71	0.88
Panel E: Linguistic Features					
Average complexity	2271	11.65	1.82	5.25	33.77
Average concreteness	2271	2.91	0.12	2.28	3.67
Medical jargon count	2271	8.26	3.84	0.00	55.00
Q&A similarity	2271	0.19	0.04	0.00	0.57

Table 5: Video Features Descriptive Statistics - Part2

	Obs	Mean	SD	Min	Max
Panel F: Acoustic-prosodic Features					
Speech rate	2271	242.56	42.18	28.80	355.87
Average voice volume	2271	-26.40	5.31	-51.69	-13.03
Voice volume variance	2271	4.79	3.03	0.12	20.14
Average voice pitch	2271	889.37	412.28	190.43	2712.32
Voice pitch variance	2271	446.69	193.10	18.46	1165.71
Average spectral centroid	2271	2343.73	462.46	809.78	5374.76
Panel G: Visual-composition Features					
Face distance	2271	105.09	57.16	6.84	337.96
Face pitch angle	2271	-3.75	6.11	-41.41	21.90
Face yaw angle	2271	5.02	9.24	-46.42	44.21
Face roll angle	2271	0.09	3.53	-16.95	22.08
Average eye gaze theta	2271	13.52	6.44	1.24	49.58
Invalid frame ratio	2271	0.23	0.31	0.00	0.98
Panel H: Promotional Strategies					
Voucher count	2271	0.32	1.09	0.00	13.00
Service recommendation	2271	3.12	4.17	0.00	34.00

indicating an overall slightly positive tone.

For the linguistic features, at the per-answer level, the mean linguistic complexity is 11.65, with higher values indicating greater textual complexity. The mean concreteness of language is 2.91, where higher scores reflect more concrete expressions. On average, each reply contains eight professional medical terms. The mean question–answer similarity is 0.19.

For the acoustic prosodic features, the average speech rate of the physicians during live streaming is 243 characters per minute, which falls within the typical range for Mandarin. The average loudness is -26 LKFS (Loudness, K-weighted, relative to full scale). The mean fundamental frequency (pitch) is 890 Hz; higher pitch values are generally associated with a brighter perceived timbre. The mean spectral centroid is 2344 (units as defined by the analysis pipeline). This metric also indexes timbral brightness; unlike pitch, it captures the distribution of signal energy across frequencies. Higher values indicate a greater concentration of energy in the higher-frequency range and, consequently, a brighter timbre.

Finally, for the visual features, the mean distance between the physician's face and the screen center is 105 pixels; larger distances indicate a position closer to the frame edges. Head-pose angles are as follows: pitch $-4°$ (slight downward tilt), yaw $5°$ (slight rightward rotation), and roll $0°$ (no tilt), indicating an overall upright posture. The average angle between the physician's gaze and the camera is $14°$, indicating a slight deviation from direct eye contact. The physician is absent from the livestream frame for 23% of the time, primarily because the face is occluded while presenting slides to explain medical content.

For promotional strategies, physicians mention coupons an average of 0.32 times per live streaming session, with substantial variation: some mention them as many as 13 times, while others do not mention them at all. Physicians also recommend their consultation services an average of 3.12 times per session, again with considerable variation. Some make such recommendations more than 30 times in a session, while others do not engage in any self-promotion.

Video Features and Live Streaming Effectiveness

Using the GSC method, we estimate the effect of live streaming on service demand at the physician level. To examine the association of these individual physician effects (ITE) with live streaming video features, we employ a flexible machine learning approach. Specifically, we use the video attributes to predict the ITE (δ_{i1} from Equation 1). This allows us to identify the attributes that are most predictive of live streaming "success." Following Datta, Ailawadi, and Van Heerde (2017) and Yang, Zhang, and Zhang (2025), we use the estimated ITE rather than raw consultation counts because it summarizes demand changes attributable to live streaming relative to the control group, net of observed covariates and the unobserved components captured by the interactive fixed effects in the GSC estimator.[19] Accordingly, our goal is to explain cross-sectional variation in these estimated responses rather than draw inferences about any single physician-level effect.

To predict the ITE, we construct an ensemble model that linearly combines one linear algorithm (Elastic Net) and two non-linear algorithms (Random Forest and eXtreme Gradient Boosting). By integrating both linear and non-linear specifications, our ensemble approach is capable of capturing additive effects as well as complex, non-linear interactions among video features (Mueller and Spinnewijn 2025). To mitigate the risk of overfitting and ensure the generalizability of our findings, we adopt a hold-out validation strategy where 80% of the observations are randomly selected for model training and the remaining 20% are reserved for out-of-sample prediction and performance evaluation. The resulting out-of-sample R^2 is 0.056. Although this value may appear modest in absolute terms, it should be assessed given the context. Live streaming sessions are substantially longer and more complex than short-form videos, making their outcomes inherently harder to predict using observable video features alone. In addition, our first-stage specification already absorbs important sources of variation, including physician fixed effects and time effects, leaving less residual variation to be explained.

After fitting the model, we use SHapley Additive exPlanations (SHAP) (Lundberg and Lee 2017) to interpret the importance of video features. Grounded in cooperative game theory, SHAP values fairly decompose a model's predictions into the marginal contributions of individual features, thereby quantifying each feature's positive or negative effect on live stream performance. This provides interpretability at both the instance and global levels and is widely used for explainable machine learning and feature-importance assessment (Chakraborty et al. 2025; Zhou et al. 2021).

We report group-level feature importance in Figure 5. For each observation, we compute

[19] Because the ITEs are estimated rather than directly observed, we conduct a robustness check by re-estimating the model using weights based on the bootstrapped variance of the individual ITE estimates. As shown in Online Appendix Figure W11, the results remain consistent. We do not use this as our main specification because the variance is itself estimated and may therefore introduce additional noise.

the mean absolute SHAP value of features within the same group to mitigate differences in group size and use this measure as the group-level feature importance.[20] We then aggregate across observations by taking the mean of these group-level values to summarize each group's global importance. The results indicate that audience interaction, session content, emotional expression, and linguistic features are most strongly associated with live streaming performance. Among these, audience interaction has the highest predictive importance (mean absolute SHAP = 0.33), substantially exceeding the other groups, followed by session content (0.27), emotional expression (0.18), and linguistic features (0.17). Visual composition (0.11), acoustic-prosodic features (0.11), and promotional strategies (0.10) exhibit comparable, moderate contributions, while professional appearance (0.04) has the lowest importance.

As a robustness check, we repeat the prediction analysis using alternative ensemble specifications: (1) replacing the linear component ElasticNet with Lasso and Ridge regressions, and (2) substituting the nonlinear component eXtreme Gradient Boosting with multilayer perceptron. The corresponding SHAP results, reported in Online Appendix W13 and W14, are qualitatively and quantitatively similar to those from the baseline ensemble model results: the same broad feature categories remain the most important predictors of the ITE, with the same relative ranking across models. This pattern suggests that our conclusions regarding feature importance are robust to the choice of prediction algorithm.

Overall, the results suggest that live streaming effectiveness in professional services is primarily associated with interaction, session content, and communication style (including emotional and linguistic features) rather than purely audiovisual presentation. This pattern is consistent with the distinctive role of professionals: patients value live streaming not only as entertainment or promotion, but as an opportunity to obtain credible information and evaluate expertise and trustworthiness in real time. In this context, interaction is likely to matter because it facilitates knowledge exchange in the short run (answering diagnostic and

[20]Using the mean absolute values of the top three most important features from each group yields similar results; see Online Appendix Table W12.

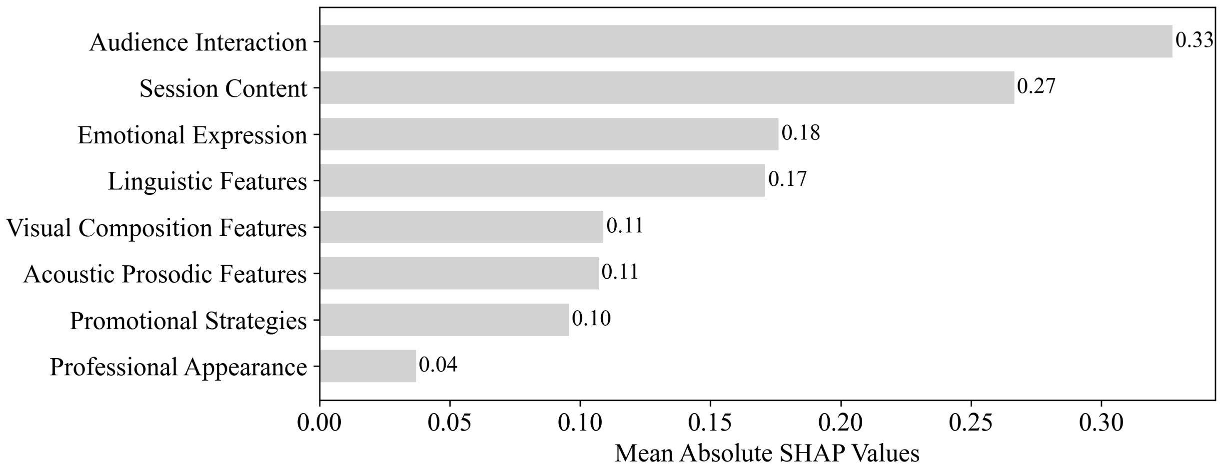


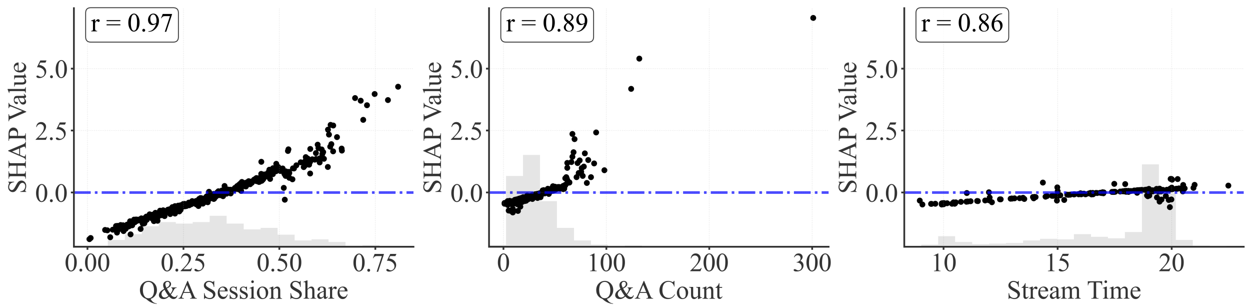
Figure 5: Importance of Feature Groups

treatment questions) and helps build perceived competence and reputation in the longer run.

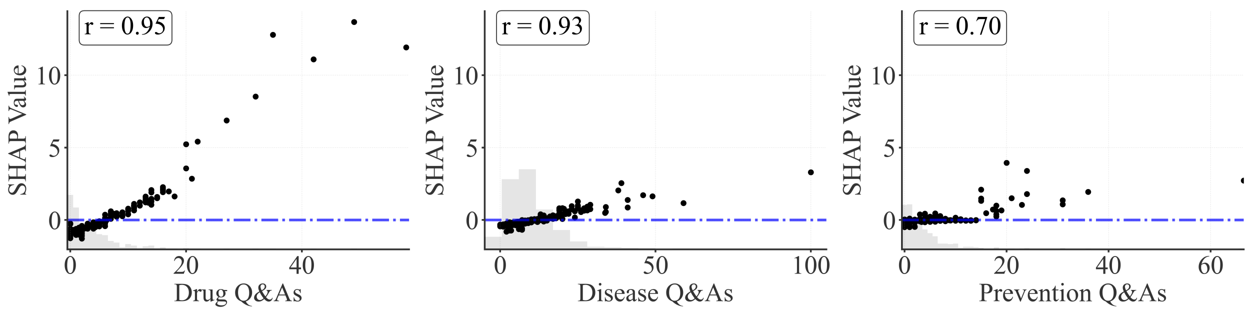
Focusing on the four most influential categories for predicting effectiveness, Figure 6 presents the three most important features within each category, ranked by mean absolute SHAP value. For each feature, the upper-left corner of the corresponding plot reports the correlation between SHAP values and feature values, indicating the direction of the effect.

For audience interaction (Panel (a)), a higher Q&A share and more Q&A exchanges are positively associated with live streaming performance. In contrast to commercial live streams where interaction often centers on entertainment or promotional content, interaction here largely reflects clinical exchanges: patients raise health-related questions and physicians provide professional guidance, address medical concerns, and build trust. Sessions conducted around 20:00 (8:00 p.m.) also tend to perform best. For within session content (Panel (b)), medically substantive responses are most strongly associated with performance, with discussions of drug usage and disease diagnosis contributing more than preventive care and other topics. Together, Panels (a) and (b) suggest that effectiveness is more linked to clinically focused engagement rather than peripheral or frivolous interaction.

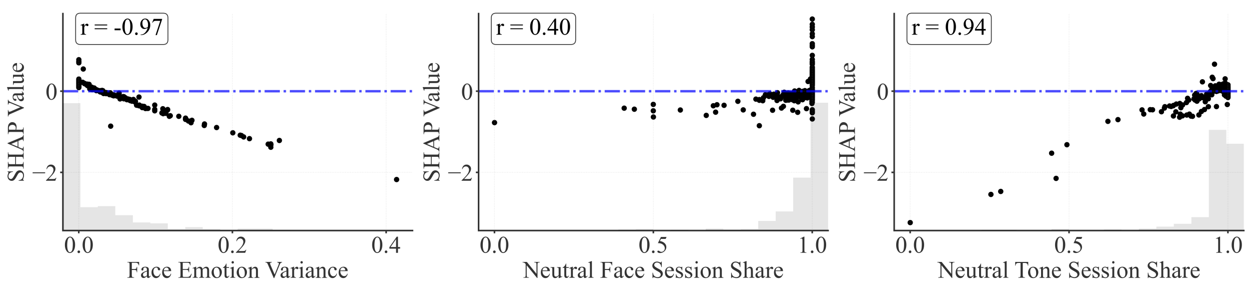
Regarding emotional expression in Panel (c), greater variability in facial expressions is negatively associated with performance. In contrast, neutral emotional expression in both



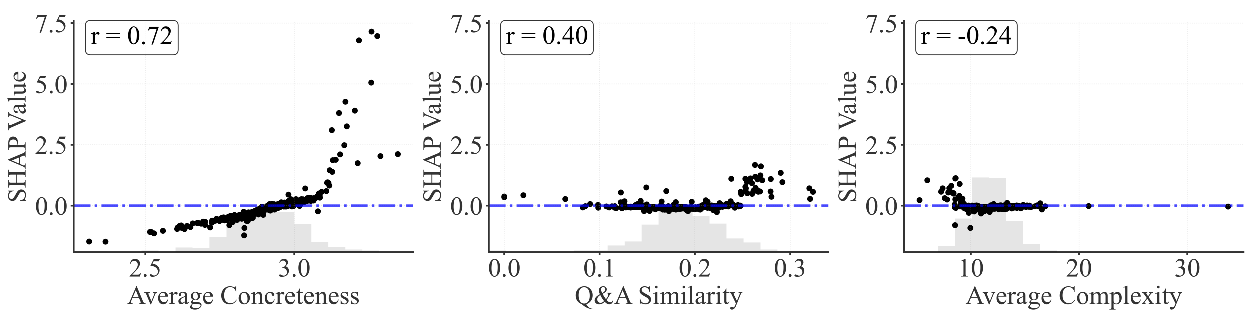
(a) Audience Interaction



(b) Session Content



(c) Emotional Expression



(d) Linguistic Features

Figure 6: Individual Feature Importance within Feature Group

Notes: Scatter points represent individual sample observations in the test set, while the shaded background histograms illustrate the probability density of feature distributions. The X-axis denotes feature values and the Y-axis represents the corresponding SHAP values.

facial appearance and vocal tone is positively associated with performance. This pattern suggests that, in professional live streaming contexts, greater emotional variability may have unintended negative effects, in sharp contrast to prior findings from commercial influencer settings (Cascio Rizzo et al. 2024; Lin, Yao, and Chen 2021). Finally, for linguistic features, live streaming sessions characterized by greater concreteness, stronger responsiveness to audience questions, and higher comprehensibility, as reflected in lower average linguistic complexity, tend to achieve better performance.

BEYOND HEALTHCARE: EVIDENCE FROM FINANCE

While our primary analysis focuses on physicians, a natural question arises: Does the effectiveness of live streaming extend to other professional service domains? To address this question and strengthen the external validity of our findings, we examine the live streaming behavior and outcomes of investment fund managers, a distinctly different type of professional service provider operating in the financial sector.

Investment fund managers represent a useful context for testing generalizability for several reasons. First, like physicians, investment fund managers provide credence goods where clients face significant information asymmetry and must rely heavily on the certified professional's expertise (Darby and Karni 1973). Second, investment fund managers operate in a fundamentally different domain (financial services rather than healthcare) characterized by distinct service attributes, client decision-making processes, and regulatory environments. Third, investment fund managers in China have increasingly adopted live streaming as a promotional channel, creating a parallel setting in which to test the generalizability of findings.[21] Together, these features make the fund-manager setting an informative complement to our physician evidence, even though the treatment and outcome are necessarily measured differently in finance.

[21] Fund live streaming in China exceeded 54,000 sessions in 2023 (21.43% year-over-year growth) with approximately 3.36 billion cumulative views. Source: "Securities Times (Feb 29, 2024)."

Data and Empirical Strategy

We construct a fund-day panel from the *Tiantian Fund* platform spanning August 2020 to August 2025. We focus on exchange-traded funds (ETFs) because, unlike other fund types, ETFs disclose daily trading volume, which enables high-frequency identification of investor response. By contrast, comparable metrics for other fund types are typically available only at the quarterly level. Given that our dependent variable is fund-level daily trading volume, the unit of analysis is the fund-day; we interpret trading volume as a high-frequency proxy for investor demand.[22] The sample includes 1,003 ETFs, of which 239 are mentioned in fund managers' live streams. ETF inception dates range from March 2006 to June 2025, and the average fund age is 3.7 years. Average daily trading volume is 60.0 million shares.

Similar to our physician analysis, we estimate the causal effect of live streaming on demand using generalized synthetic control (GSC). GSC uses information from the full panel of treated and untreated funds to infer what would have happened to treated funds in the absence of live stream mentions. By modeling latent common factors with fund-specific loadings, the approach accounts for unobserved shocks that vary over time and affect funds differently. Because mentions are not randomly assigned and may reflect fund managers' strategic choices, we interpret this finance extension as complementary evidence and emphasize pre-treatment fit as an important diagnostic for the design.

Main Results: Impact of Live Streaming on Demand

We report the estimated effects of ETF mentions in fund managers' live streams on ETF trading volume in Online Appendix Figure W20. Consistent with the findings from the physician setting, ETF mentions in live streams increase subsequent trading activity. Esti-

[22] Although daily net fund flow is a possible alternative, it reflects both purchases and redemptions, and redemptions may occur for reasons unrelated to live stream content. In our context, where live streams are intended primarily to stimulate investment, trading volume provides a more direct and higher-frequency measure of investor response. We use share volume rather than dollar volume because intraday ETF prices fluctuate continuously, so dollar volume varies mechanically with execution prices even when the number of shares traded is unchanged. Share volume is therefore less sensitive to contemporaneous price movements and provides a cleaner proxy for quantity-based trading activity.

mated effects are 22 million shares on day 1 ($p = 0.064$, an 18% increase), 25 million shares on day 2 ($p = 0.036$, a 20% increase), and 23 million shares on day 3 ($p = 0.051$, an 19% increase), suggesting that the effect emerges shortly after the mention. By day 4, the estimated effect is 21 million shares ($p = 0.042$, a 18% increase), and it remains positive and statistically significant over the subsequent two weeks.[23]

Overall, the estimates indicate a sustained increase in trading volume following live stream mentions. The similar patterns observed across the healthcare and finance settings suggest that live streaming is followed by increased behavioral responses in both contexts. This cross-context consistency suggests that our findings are not specific to healthcare and may apply more broadly to other professional service domains.

Video Features and Live Streaming Effectiveness for Fund Managers

To examine the predictors of effectiveness in investment fund managers' live streams, we apply the same framework used in the physician setting to the set of live streaming sessions in which ETFs are mentioned. Specifically, we use multimodal methods to extract video features from three modalities: text, audio, and visual content.

We adapt the feature categories from the physician setting to reflect the institutional characteristics of the fund manager context. For linguistic features, we use a finance-specific lexicon from the Tsinghua Open Chinese Lexicon (THUOCL) dictionary (Han et al. 2016) to capture domain-relevant terminology. Signals of professional appearance also differ across settings. Unlike physicians, investment fund managers do not typically rely on attire such as white coats or masks to convey professionalism. Instead, we measure whether fund managers wear suits and glasses, as well as their facial attractiveness. In addition, 39% of fund live streams contain no Q&A interaction, so we do not further classify Q&A content in this setting. Finally, because industry regulations prohibit fund managers from directly promoting

[23]The marginal statistical significance on days 1 and 3 may reflect short delays in investor response, as viewers process the information, seek additional opinions, and place orders with a lag. Because some live streaming sessions mention multiple ETFs, we re-estimate the analysis using only sessions in which a single ETF is mentioned. The results for 182 ETFs, reported in Online Appendix Figure W21, show a broadly similar pattern.

funds through incentives or similar tactics during live streams, we exclude promotional-strategy features from the analysis.

We use the same ensemble model as in the physician setting to predict the average individual treatment effect (ITE) over the first four days following a fund live stream based on stream features. We use the four-day average to reduce noise in the daily estimates and to reflect the short lag with which investor responses become consistently positive in our earlier analysis. The model is trained on 80% of the sample and evaluated on the remaining 20% holdout sample. We assess feature importance using SHAP values. Online Appendix Figure W22 reports group-level feature importance for fund live streams. The results indicate that linguistic features, audience interaction, and emotional expression are the three most important predictors of the estimated treatment effect of fund managers' live streams, broadly consistent with the patterns observed in the physician setting.

To aid interpretation, Online Appendix Figure W23 plots the relationships between individual features within the three most predictive categories and their corresponding SHAP values. The variables shown parallel those examined in the physician live streaming analysis. For linguistic features, simpler and more concrete communication is associated with higher SHAP values, as reflected in lower linguistic complexity and greater textual concreteness. We also find that the use of finance-related terms is positively associated with effectiveness, suggesting that viewers value substantive content. For audience interaction, both the share of the live stream devoted to Q&A and the number of Q&A exchanges exhibit positive SHAP values, indicating positive contributions to live streaming effectiveness. Live streams conducted before 15:00 are also associated with higher SHAP values, plausibly because ETFs can be traded only during market hours, leading to higher viewer engagement. For emotional expression, more neutral facial expressions and vocal tone are positively associated with live stream effectiveness, whereas greater emotional variability, as captured by higher facial emotion variance, is negatively associated with effectiveness.

Discussion

The fund manager analysis yields several important insights. First, the significant positive effect of live streaming on service demand demonstrates that our physician findings generalize to a markedly different professional service context. The consistency in the direction of effects (despite fundamental differences in service characteristics, client needs, and regulatory environments) provides evidence that live streaming serves as an effective demand generation tool across two distinct credence service domains.

Second, the consistency of three key dimensions across both contexts highlights potentially generalizable associations underlying live streaming effectiveness. First, audience interaction is strongly positively associated with demand generation in both the physician and fund manager settings, suggesting that real-time engagement is a critical driver across professional domains. Second, a simple and concrete linguistic style is positively associated with effectiveness. In professional live streaming, where viewers often lack domain expertise, plain and specific language may reduce comprehension costs and cognitive load, thereby enhancing behavioral responses. Third, greater emotional expressiveness is negatively associated with effectiveness in both contexts. Physicians who maintain neutral facial expressions show associations with better outcomes, and fund managers who maintain neutral facial expressions and vocal tone also show association with stronger effects. Together, these findings suggest that professional live streaming operates through mechanisms that differ from those in pre-recorded settings, where audience interaction is limited, and from non-professional influencer or entertainment live streaming contexts, where emotional expressiveness typically enhances effectiveness (Cascio Rizzo et al. 2024; Lin, Yao, and Chen 2021). The consistency of these patterns across healthcare and financial services further suggests that platforms and professionals should start by testing these associations, with a focus on audience interaction and the maintenance of a neutral professional demeanor.

CONCLUSION

The rapid rise of live streaming has created new opportunities for professionals to reach and engage clients. Yet little is known about whether live streaming is effective in professional service contexts or which factors predict its success. Using online healthcare as our primary setting, and extending the analysis to financial services, this paper provides causal evidence that professional live streaming increases service demand and takes the first step in terms of identifying the video attributes associated with its effectiveness.

Our findings yield four key insights. First, adopting live streaming leads to a causal and economically meaningful increase in service demand. For physicians, this effect persists for several months and gradually tapers off over time. Second, the effectiveness of live streaming is associated with specific session attributes. Audience interaction and clear, concrete communication are positively associated with demand generation, while heightened emotional expressiveness, which is often advantageous in consumer influencer settings, is negatively associated with effectiveness in professional contexts. This contrast highlights that mechanisms emphasized in consumer-product influencer marketing may be less relevant in settings where credibility, expertise, and trust are central. Third, the effects are heterogeneous across physicians; for example, the demand gains are larger for more established physicians in terms of seniority and baseline volume. Fourth, extending the analysis to fund managers shows that live streaming also increases service demand in financial services, and that the key associations of effectiveness, namely audience interaction, linguistic clarity, and neutral emotional expression, are similar across both contexts. Taken together, these results indicate that live streaming can serve as an effective demand-generation tool across two distinct credence service domains.

This study advances understanding of how live streaming functions as a demand-generation tool in professional service contexts. First, it adds to professional service marketing by identifying live streaming as a novel client acquisition channel and by highlighting the features that are associated with making professional live streams more effective. Second, it shows that when professionals, rather than commercial influencers, act as content creators, both the effectiveness of live streaming and the underlying associations differ in important ways, particularly with respect to emotional expression. Third, the study demonstrates the value of combining causal inference with multimodal analysis of video content, illustrating how rich visual, auditory, and textual data can be systematically linked to heterogeneous treatment effects in digital settings.

Furthermore, our findings offer implications for practice and policy. For platforms, targeting more senior physicians and providing incentives for them to adopt and continue live streaming may lead to better outcomes. Platforms should also carry out testing to see if tools and training programs that facilitate real-time interaction and support clear communication help professionals create more effective live streams. For professionals, the results suggest the value of experimenting with engaging audiences through meaningful interaction and maintaining a neutral and professional demeanor. For regulators, the findings highlight that live streaming can change demand patterns in sensitive domains such as healthcare and financial decision-making. Regulators therefore need to examine the benefits and costs of whether it needs to be monitored and/or regulated.

In conclusion, our study opens several avenues for future research. First, while we document that live streaming leads to higher consultation demand, it is unclear that this leads to better physician-patient matching and, in turn, improved health outcomes. In other words, we cannot pinpoint the welfare effects from this new channel. Second, the "success" of live streaming could lead to adverse behaviors and outcomes, such as misinformation, over-commercialization, and long-term effects on trust. Third, extending this framework beyond the two professional domains examined here, preferably in a non-Chinese setting, would further enrich our understanding of how live streaming reshapes professional service delivery. Finally, future studies could test the causal role of the specific live streaming attributes identified here by moving beyond predictive models to randomized experiments.

REFERENCES

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller (2010), "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program," *Journal of the American Statistical Association*, 105 (490), 493–505.

Abadie, Alberto and Javier Gardeazabal (2003), "The economic costs of conflict: A case study of the Basque Country," *American Economic Review*, 93 (1), 113–132.

Abel, Richard L (1979), "The rise of professionalism," *Brit. JL & Soc'y*, 6, 82.

Allen, Katherine A, Victoria Charpentier, Marissa A Hendrickson, Molly Kessler, Rachael Gotlieb, Jordan Marmet, Emily Hause, Corinne Praska, Scott Lunos, and Michael B Pitt (2023), "Jargon be gone–patient preference in doctor communication," *Journal of Patient Experience*, 10.

An, Keyu, Qian Chen, Chong Deng, Zhihao Du, Changfeng Gao, Zhifu Gao, Yue Gu, Ting He, Hangrui Hu, Kai Hu, Shengpeng Ji, Yabin Li, Zerui Li, Heng Lu, Xiang Lv, Bin Ma, Ziyang Ma, Chongjia Ni, Changhe Song, Jiaqi Shi, Xian Shi, Hao Wang, Wen Wang, Yuxuan Wang, Zhangyu Xiao, Zhijie Yan, Yexin Yang, Bin Zhang, Qingling Zhang, Shi-Ya Zhang, Nan Zhao, and Siqi Zheng (2024), "FunAudioLLM: Voice Understanding and Generation Foundation Models for Natural Interaction Between Humans and LLMs," *ArXiv*, abs/2407.04051 https://api.semanticscholar.org/CorpusID:271039436.

Bairathi, Mimansa, Xu Zhang, and Anja Lambrecht (2025), "The value of platform endorsement," *Marketing Science*, 44 (1), 84–101.

Barsade, Sigal G (2002), "The ripple effect: Emotional contagion and its influence on group behavior," *Administrative Science Quarterly*, 47 (4), 644–675.

Beichert, Maximilian, Andreas Bayerl, Jacob Goldenberg, and Andreas Lanz (2024), "Revenue generation through influencer marketing," *Journal of Marketing*, 88 (4), 40–63.

Bharadwaj, Neeraj, Michel Ballings, Prasad A Naik, Miller Moore, and Mustafa Murat Arat (2022), "A new livestream retail analytics framework to assess the sales impact of emotional displays," *Journal of Marketing*, 86 (1), 27–47.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess (2024), "Revisiting event-study designs: robust and efficient estimation," *Review of Economic Studies*, 91 (6), 3253–3285.

Brucks, Melanie S, Jacqueline R Rifkin, and Jeff S Johnson (2025), "Video-call glitches trigger uncanniness and harm consequential life outcomes," *Nature*, 650 (8101), 409–416.

Callaway, Brantly and Pedro HC Sant'Anna (2021), "Difference-in-differences with multiple time periods," *Journal of Econometrics*, 225 (2), 200–230.

Cascio Rizzo, Giovanni Luca, Francisco Villarroel Ordenes, Rumen Pozharliev, Matteo De Angelis, and Michele Costabile (2024), "How high-arousal language shapes micro-versus macro-influencers' impact," *Journal of Marketing*, 88 (4), 107–128.

Chakraborty, Ishita, Khai Chiong, Howard Dover, and K Sudhir (2025), "Can AI and AI-Hybrids detect persuasion skills? Salesforce hiring with conversational video interviews," *Marketing Science*, 44 (1), 30–53.

Chen, Cheng, Yuheng Hu, Yingda Lu, and Yili Hong (2025), "Two Streams to Success: Unpacking the Informational and Relational Roles of Live Streaming in Online Shopping," *Production and Operations Management*, forthcoming.

Cheng, Mengjie and Shunyuan Zhang (2025), "Reputation burning: Analyzing the impact of brand sponsorship on social influencers," *Management Science*, 71 (7), 5910–5932.

Clow, Kenneth E, Carolyn Tripp, and James T Kenny (1996), "The importance of service quality determinants in advertising a professional service: an exploratory study," *Journal of Services Marketing*, 10 (2), 57–72.

Crossley, Scott A, Renu Balyan, Jennifer Liu, Andrew J Karter, Danielle McNamara, and Dean Schillinger (2020), "Predicting the readability of physicians' secure messages to improve health communication using novel linguistic features: Findings from the ECLIPPSE study," *Journal of Communication in Healthcare*, 13 (4), 344–356.

Darby, Michael R and Edi Karni (1973), "Free competition and the optimal amount of fraud," *The Journal of Law and Economics*, 16 (1), 67–88.

Datta, Hannes, Kusum L Ailawadi, and Harald J Van Heerde (2017), "How well does consumer-based brand equity align with sales-based brand equity and marketing-mix response?," *Journal of Marketing*, 81 (3), 1–20.

Feng, Xiaohang Flora, Shunyuan Zhang, Xiao Liu, Kannan Srinivasan, and Cait Poynor Lamberton (2025), "An AI method to score celebrity visual potential from human faces," *Journal of Marketing Research*, 62.

Gu, Xian, Xiaoxi Zhang, and PK Kannan (2024), "Influencer mix strategies in livestream commerce: Impact on product sales," *Journal of Marketing*, 88 (4), 64–83.

Guitart, Ivan A and Stefan Stremersch (2021), "The impact of informational and emotional television ad content on online search and sales," *Journal of Marketing Research*, 58 (2), 299–320.

Hall, Judith A, Debra L Roter, and Cynthia S Rand (1981), "Communication of affect between patient and physician," *Journal of Health and Social Behavior*, pages 18–30.

Han, Shiyi, Yuhui Zhang, Yunshan Ma, Cunchao Tu, Zhipeng Guo, Zhiyuan Liu, and Maosong Sun (2016), "THUOCL: Tsinghua open Chinese lexicon," *Tsinghua University*.

Hatfield, Elaine, John T Cacioppo, and Richard L Rapson (1992), "Primitive emotional contagion," *Emotion and Social Behavior*, 14, 151–177.

Hu, Allen and Song Ma (2025), "Persuading investors: A video-based study," *The Journal of Finance*, 80 (5), 2639–2688.

Huang, Yufeng and Ilya Morozov (2025), "The promotional effects of live streams by Twitch influencers," *Marketing Science*, forthcoming.

Jeziorski, Przemyslaw, Dingzhe Leng, and Stephan Seiler (2025), "Causal Inference with Endogenous Price Response," *CESifo Working Paper*.

Jiang, Zhenling and Donggwan Kim (2024), "The Effects of TV Advertising and Ad Content on Consumer Financial Decisions: Evidence from Mortgage Refinancing," *The Wharton School Research Paper*.

Kaye, Aaron P, Michael Luca, and Sonal Vats (2024), "Differential Impacts of Online Ratings in the Market for Medical Services," https://apkaye.github.io/aaronpkaye.website/Preliminary_Draft_ZocDoc.pdf.

Leung, Fine F, Flora F Gu, Yiwei Li, Jonathan Z Zhang, and Robert W Palmatier (2022), "Influencer marketing effectiveness," *Journal of Marketing*, 86 (6), 93–115.

Li, Nan, Avery Haviv, and Mitchell J Lovett (2025), "Opposing Influences of YouTube Influencers: Purchase and Usage Effects in the Video Game Industry," *Marketing Science*, forthcoming.

Libai, Barak, Ana Babić Rosario, Maximilian Beichert, Bas Donkers, Michael Haenlein, Reto Hofstetter, PK Kannan, Ralf van der Lans, Andreas Lanz, H Alice Li, Dina Mayzlin, Eitan Muller, Daniel Shapira, Jeremy Yang, and Lingling Zhang (2025), "Influencer marketing unlocked:

Understanding the value chains driving the creator economy," *Journal of the Academy of Marketing Science*, 53 (1), 4–28.

Lin, Yan, Dai Yao, and Xingyu Chen (2021), "Happiness begets money: Emotion and engagement in live streaming," *Journal of Marketing Research*, 58 (3), 417–438.

Liu, Shan, Guangsen Si, and Baojun Gao (2022), "Which voice are you satisfied with? Understanding the physician–patient voice interactions on online health platforms," *Decision Support Systems*, 157, 113754.

Liu, Shan, Muyu Zhang, Baojun Gao, and Guoyin Jiang (2020), "Physician voice characteristics and patient satisfaction in online health consultation," *Information & Management*, 57 (5), 103233.

Liu, Xiao (2023), "Dynamic coupon targeting using batch deep reinforcement learning: An application to livestream shopping," *Marketing Science*, 42 (4), 637–658.

Lu, Shijie, Dai Yao, Xingyu Chen, and Rajdeep Grewal (2021), "Do larger audiences generate greater revenues under pay what you want? Evidence from a live streaming platform," *Marketing Science*, 40 (5), 964–984.

Lundberg, Scott M and Su-In Lee (2017), "A unified approach to interpreting model predictions," *Advances in Neural Information Processing Systems*, 30.

Luo, Lijuan, Ling Liu, Yujie Zheng, and Yuwei Wang (2025), "The power of voice: Investigating the effects of streamer voice characteristics on sales performance in live streaming E-commerce," *Journal of Retailing and Consumer Services*, 84, 104260.

Ma, Xiangyang, Hongzhong Chen, Xiaoping Lang, Tieshan Li, Niwei Wu, and Bao Duong (2024), "Research on the impact of streamers' linguistic emotional valence on live streaming performance in live streaming shopping environments," *Journal of Retailing and Consumer Services*, 81, 104040.

Mueller, Andreas I. and Johannes Spinnewijn (2025), "The Nature of Long-Term Unemployment: Predictability, Heterogeneity, and Selection," *Journal of Political Economy*, 133 (12), 3846–3902.

Öberg, Shirin Ahlbäck and Louise Bringselius (2015), "Professionalism and organizational performance in the wake of new managerialism," *European Political Science Review*, 7 (4), 499–523.

Oe, Akemi and Yoshitoshi Murata (2019), "Identification of Appropriate Facial Expressions for Medical Doctors," *IARIA, International Journal on Advances in Life Sciences*, 11 (1), 44–55.

Pamuji, Slamet and Sun Limei (2023), "The Managerial Competence Of The Madrasa Head In Improving Teacher Professionalism And Performance At Mi Al-Maarif Bojongsari, Cilacap District," *Pengabdian: Jurnal Abdimas*, 1 (2), 66–74.

Pan, Meizhi, Markus Blut, Arezou Ghiassaleh, and Zach WY Lee (2025), "Influencer marketing effectiveness: A meta-analytic review," *Journal of the Academy of Marketing Science*, 53 (1), 52–78.

Petrilli, Christopher M, Sanjay Saint, Joseph J Jennings, Andrew Caruso, Latoya Kuhn, Ashley Snyder, and Vineet Chopra (2018), "Understanding patient preference for physician attire: a cross-sectional observational study of 10 academic medical centres in the USA," *BMJ open*, 8 (5), e021239.

Qu, Changle, Sunhao Dai, Ke Guo, Liqin Zhao, Yanan Niu, Xiao Zhang, and Jun Xu (2025), "KuaiLive: A Real-time Interactive Dataset for Live Streaming Recommendation," *arXiv preprint arXiv:2508.05633*.

Rajaram, Prashant and Puneet Manchanda (2020), "Unboxing engagement in YouTube influencer videos: An attention-based approach," *arXiv preprint arXiv:2012.12311*.

Rehman, Shakaib U, Paul J Nietert, Dennis W Cope, and Anne Osborne Kilpatrick (2005), "What to wear today? Effect of doctor's attire on the trust and confidence of patients," *The American Journal of Medicine*, 118 (11), 1279–1286.

Shimp, Terence A, Stacy L Wood, and Laura Smarandescu (2007), "Self-generated advertisements: Testimonials and the perils of consumer exaggeration," *Journal of Advertising Research*, 47 (4), 453–461.

Snoek, Cees GM and Marcel Worring (2005), "Multimodal video indexing: A review of the state-of-the-art," *Multimedia Tools and Applications*, 25 (1), 5–35.

Sun, Liyang and Sarah Abraham (2021), "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of Econometrics*, 225 (2), 175–199.

Tang, Xiaoli, Zefeng Hao, and Xiaolin Li (2023), "The influence of streamers' physical attractiveness on consumer response behavior: based on eye-tracking experiments," *Frontiers in Psychology*, 14, 1297369.

Thomas, Marlyn, Meena Hariharan, Suvashisa Rana, Sunayana Swain, and Asher Andrew (2014), "Medical jargons as hindrance in doctor–patient communication," *Psychological Studies*, 59, 394–400.

Tian, Zijun, Ryan Dew, and Raghuram Iyengar (2024), "Mega or micro? Influencer selection using follower elasticity," *Journal of Marketing Research*, 61 (3), 472–495.

Wang, Siqi, Xiaofei Zhang, and Fanbo Meng (2025), "Convergent or not? The effect of linguistic convergence on the effectiveness of online physician-patient communication," *Industrial Management & Data Systems*, forthcoming.

Xu, Wei, Xiaohan Zhang, Runyu Chen, and Zhan Yang (2023), "How do you say it matters? A multimodal analytics framework for product return prediction in live streaming e-commerce," *Decision Support Systems*, 172, 113984.

Xu, Yiqing (2017), "Generalized synthetic control method: Causal inference with interactive fixed effects models," *Political Analysis*, 25 (1), 57–76.

Xu, Yuqian, Mor Armony, and Anindya Ghose (2021), "The interplay between online reviews and physician demand: An empirical investigation," *Management Science*, 67 (12), 7344–7361.

Yang, Jeremy, Juanjuan Zhang, and Yuhan Zhang (2025), "Engagement that sells: Influencer video advertising on TikTok," *Marketing Science*, 44 (2), 247–267.

Yohana, Corry, Ma'ruf Akbar, and Suparjo (2020), "The effect of integrity, professionalism, and innovation on service performance," *Journal of Business and Behavioural Entrepreneurship*, 4 (1), 26–42.

Zhan, Jiajia, Xu Zhang, and Hongqiao Fu (2026), "Information disclosure via platform endorsement in online healthcare," *Journal of Marketing Research*, forthcoming.

Zhang, Zhipeng, Keda Qiu, and Yan Ye (2025), "Influence of audiovisual features of short video advertising on consumer engagement behaviors: Evidence from TikTok," *Journal of Business Research*, 201, 115662.

Zhou, Mi, George H Chen, Pedro Ferreira, and Michael D Smith (2021), "Consumer behavior in the online classroom: Using video analytics and machine learning to understand the consumption of video courseware," *Journal of Marketing Research*, 58 (6), 1079–1100.

Web Appendix A: Data Cleaning Process

We begin by describing the data cleaning procedure for the first dataset, which contains consultation-related information. Using an automated web scraper, we collect monthly data on physicians' attributes, including name, platform ID, affiliated hospital, professional title, specialty, cumulative number of consultations, consultation prices, patient ratings, and rankings, as measured at the end of each calendar month.[1]

The raw dataset includes 285,694 physicians. We apply the following steps to remove inactive accounts and outliers:

1. We exclude 150,741 physicians who had zero consultations throughout the entire sample period. These physicians are listed on the platform but do not provide online services. This leaves a sample of 134,953 physicians.

2. We remove 327 physicians whose monthly consultation volume exceeds 1,000 because such cases likely indicate that multiple practitioners are sharing a single account. This step yields a sample of 134,626 physicians.

3. We exclude 9,223 physicians with missing information on one or more key variables, including consultation number, consultation price, ranking index, or patient ratings, during the observation period.[2] These missing values are likely due to technical issues during the scraping process, such as unstable internet connections.[3] The final cleaned sample consists of 125,403 physicians.

We next describe the cleaning process for the live streaming dataset. This dataset includes information on all live streaming sessions conducted on the platform, including the video content, associated physician ID, and session timestamp. A total of 6,499 physicians conducted at least one live streaming session during the observation window. Among these, 4,784 physicians could be matched to the cleaned consultation-related dataset. For these physicians, we apply the following additional cleaning steps:

1. Our consultation data collection began in September 2020, while the live streaming feature was introduced in July 2020. Because our main analysis requires at least six months of pre-treatment observations, we exclude 1,608 streaming physicians who have fewer than six months of data prior to their first live streaming session. This results in 3,176 physicians. As discussed in Online Appendix D, we perform robustness checks with shorter pre-treatment windows and find our results to be consistent.

[1] Given that there are over 200,000 physicians on the platform, it typically takes several days to complete data collection for all physicians in a given month. To estimate values as of the end of each month, we perform prorated calculations when necessary. For example, if a physician's cumulative number of consultations is recorded on two non-month-end dates, we prorate the change across the number of days between observations to approximate the value at month-end. The number of consultations in a given month is then calculated as the difference in cumulative consultations between two consecutive months.

[2] For the cumulative number of educational articles, we infer missing values when the counts are the same in the previous and following months. After this procedure, we obtain a complete record for this variable.

[3] We examine correlations between the missing-data indicator and several variables, including number of services, service price, ranking index, and patients' ratings on effectiveness and attitude, and find no significant associations. This suggests that missing data is unlikely to bias our estimates.

2. In some cases, the interval between two scraped observations is much longer than one month, which undermines the precision of the prorated consultation estimates. This is particularly problematic when such gaps occur around the time of live streaming adoption. We exclude 661 such physicians, leaving a sample of 2,515.[4]

3. We exclude physicians whose first live streaming session contained no patient questions (preventing construction of textual attributes) or whose face was too distant from the camera to detect facial expressions. After these exclusions, the treatment group comprises 2,272 streaming physicians.[5]

4. We exclude one physician specializing in pathology because this is the only streaming physician from this specialty, making it impossible to include specialty fixed effects. After applying these steps, our final treatment group comprises 2,271 streaming physicians.

Finally, we discuss the sample cleaning process for our control group. Abadie, Diamond, and Hainmueller (2010) find that synthetic control works best with a small donor pool of untreated units that are close to the treated units in the space of the predictors because this helps reduce over-fitting and interpolation biases. Thus, we follow their suggestion by trimming the control group so that the trend of the outcome variable for control units is similar to that of the treatment units during the pretreatment period in the raw data. Specifically, we select physicians who have been consistently active on the platform in our observed data, to mirror the pattern observed in the treated group. To be specific, we choose physicians with an average monthly consultation volume greater than 5, and who are active in over 75% of the months during our sample period. This gives us a final sample of 4,955 control physicians. In Figure W18 of Online Appendix E, we use all non-streaming physicians as the control group and the results remain consistent with our main findings.

[4]Results remain consistent when these physicians are included. See Figure W15 and Figure W16 of Section D.

[5]Results remain robust when these physicians are included. See Figure W17 in Online Appendix D.

Web Appendix B: Pre-treatment Trends

Figure W1 presents a model-free comparison of service quantity for treated and control physicians over a window from six months before to six months after the first live streaming month. For treated physicians, consultation counts are aligned relative to each physician's live streaming adoption month. For the control group, we construct a synthetic control for each treated physician by averaging consultations across all control physicians and aligning the time axis to the treated physician's adoption month. We then average across these synthetic controls to obtain the control series.

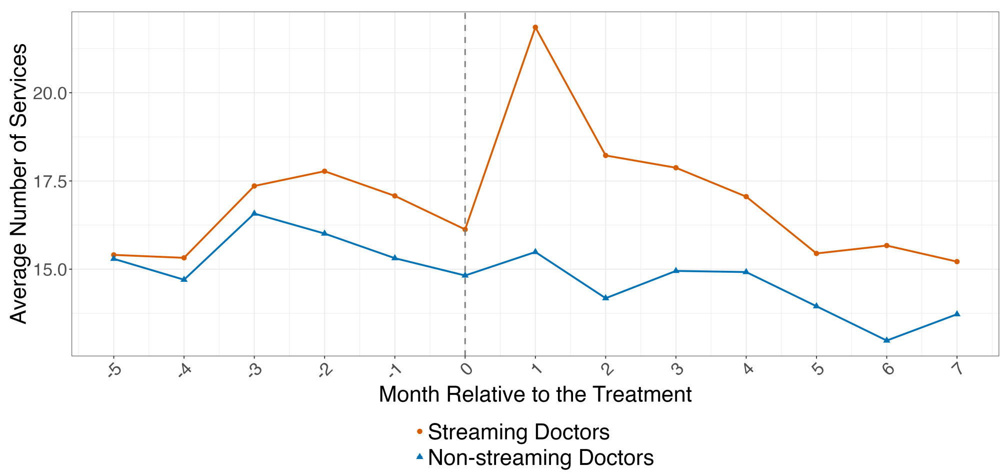


Figure W1: Model-Free Comparison of Treated and Control Physicians

Figure W2 presents the pre-treatment trends for treated physicians and their synthetic controls. The figure shows the average number of services from six months before to six months after the first live streaming month. As shown, the two groups closely track each other in the pre-treatment period, with smooth trends and no systematic differences, providing suggestive support for the no-anticipation assumption.

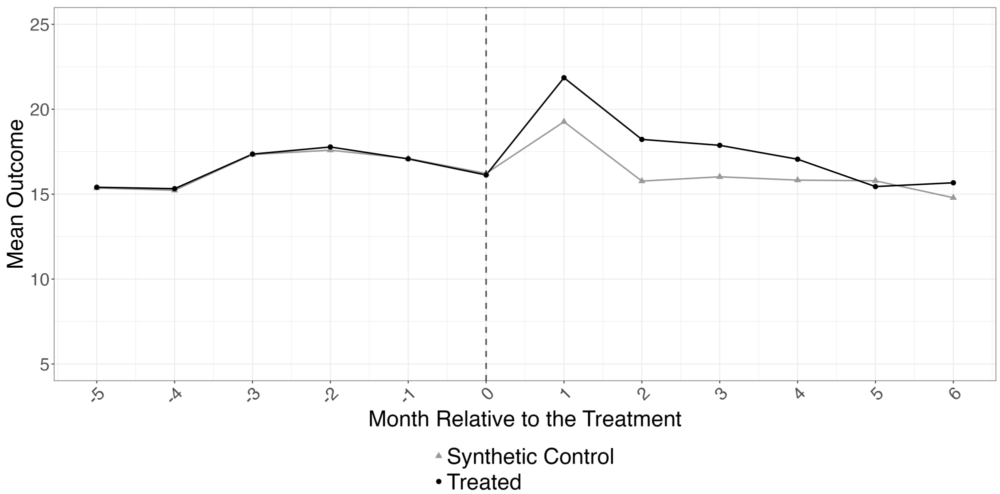
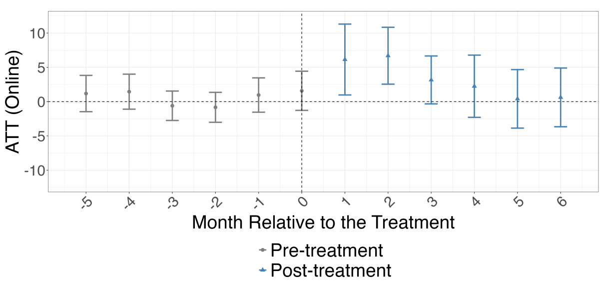


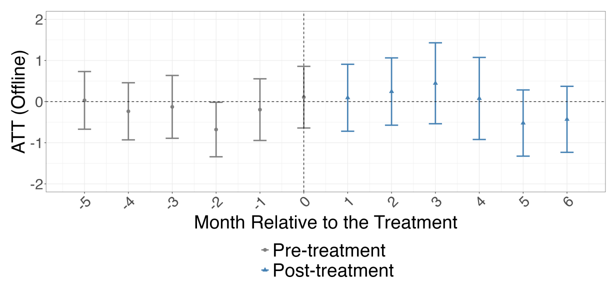
Figure W2: Comparison between Treated Physicians and Synthetic Control Physicians

Web Appendix C: Impact of Live Streaming on Offline Demand

To examine the impact of live streaming on offline consultation demand, we focus on a subsample of 2,372 physicians, 356 of whom adopted live streaming on Haodf. If a patient visits a physician offline at the hospital, the platform provides an option to report the visit and leave feedback. Because not all offline patients report their visits on the platform, and the reporting rate is unobserved, this measure serves only as a proxy for offline service demand. We apply the same GSC approach to this subsample to estimate the impact of live streaming adoption on offline service demand. Figure W3 presents the results. Panel (a) shows that live streaming increases streaming physicians' online consultation demand, consistent with our findings from the main sample, and that the effect lasts for approximately two to three months. Panel (b), in contrast, shows no statistically significant effect on streaming physicians' offline consultation demand. This null effect may reflect the fact that most patients use the platform to access physicians outside their local market, or that live streaming reaches an audience beyond the physician's local market, or that the proxy is noisy.



(a) Online Consultations



(b) Offline Consultations

Figure W3: Impact of Live Streaming on Online and Offline Consultations

Notes: This figure plots the estimated ATT (with 95% confidence intervals) of live streaming on online consultations (Panel a) and offline consultations (Panel b) for each month relative to the live streaming adoption month, using the generalized synthetic control method on a subsample. The dashed vertical line indicates the month of live streaming adoption.

Web Appendix D: Robustness

Alternative Estimation Strategy

We assess the robustness of our findings using four alternative estimation strategies.

Two-way fixed effects model. As a first alternative to the GSC method, we re-estimate the impact of live streaming on service demand using a two-way fixed effects model, controlling for physician and year–month fixed effects. The results, reported in column (1) of Table W1, show an estimated impact of 2.330, which is close to our GSC estimate of 2.593.

Callaway and Sant'Anna DID estimator. A limitation of the two-way fixed effects estimator is that it relies on both "clean" comparisons (between treated and not-yet-treated units) and "forbidden" comparisons (between units already treated at different times). The latter can bias results when treatment effects are heterogeneous across adoption cohorts (Goodman-Bacon 2021). To address this issue, we implement the Callaway and Sant'Anna (2021) DID estimator, which restricts comparisons to treated versus not-yet-treated units, thereby avoiding forbidden comparisons. Column (2) of Table W1 reports the results, which are consistent with our main estimates.

Sun and Abraham DID estimator. As a third robustness check, we adopt the estimator proposed by Sun and Abraham (2021), which also addresses the forbidden comparisons problem in two-way fixed effects models. Unlike the Callaway and Sant'Anna approach, the Sun and Abraham estimator compares earlier-treated cohorts with last-treated or never-treated units. Column (3) of Table W1 presents these results, which again align with our main findings.

Borusyak, Jaravel, and Spiess DID estimator. Finally, we implement the imputation-based estimator of Borusyak, Jaravel, and Spiess (2024), another method designed for staggered DID settings. Similar to the above approaches, it avoids relying on problematic comparisons between already-treated units. Unlike methods that require specific control group definitions, however, the imputation approach flexibly constructs counterfactuals using any untreated observations. Column (4) of Table W1 reports the results, which remain consistent with our main estimates.

Table W1: Alternative Methods

	(1)	(2)	(3)	(4)
	TWFE	CSDID	SADID	BJSDID
ATT	**2.330**	**2.101**	**2.970**	**2.894**
	(0.504)	(0.501)	(0.760)	(0.741)
	[0.000]	[0.000]	[0.000]	[0.000]

Robustness Checks Concerning the Potential Violation of SUTVA

To address potential violations of the Stable Unit Treatment Value Assumption (SUTVA), we conduct two tests. First, we exclude from the control group physicians who are in the same department and with similar ranking as treatment physicians during the pretreatment periods because these physicians are most likely to experience negative spillover effects (Zhan, Zhang, and Fu (2026)). Figure W4 shows the result is consistent with our main finding.

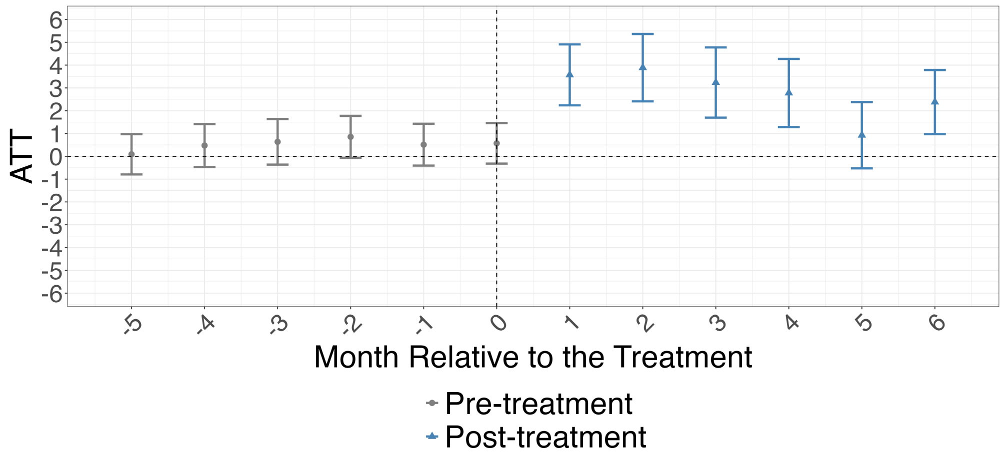


Figure W4: Exclude Physicians from the Same Specialty

Second, considering the potential spillover via hospital affiliation, we exclude control physicians from the same hospital as streaming physicians. Figure W5 shows the results, which are consistent with our main findings.

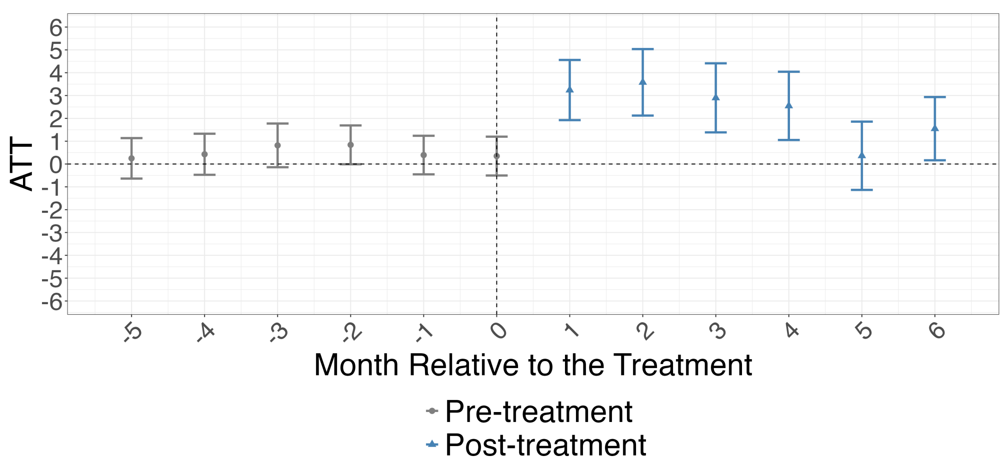


Figure W5: Exclude Physicians from the Same Hospital

Alternative Estimation Window

We test the robustness of our results to different observation windows: (1) 3 months before and after the first live stream (Figure W6), (2) 9 months before and after (Figure W7), and (3) 12 months before and after (Figure W8). For the 3-month window, the sample includes 2,750 streaming physicians, more than the main sample because it also includes physicians with 3–6 months of pre-treatment data. As expected, pretreatment balance worsens with only 3 months of history, but the estimated effects remain consistent with our main results. For the 9-month and 12-month windows, the samples include 2,125 and 1,793 streaming physicians, respectively. In both cases, the results remain stable and aligned with our main findings.

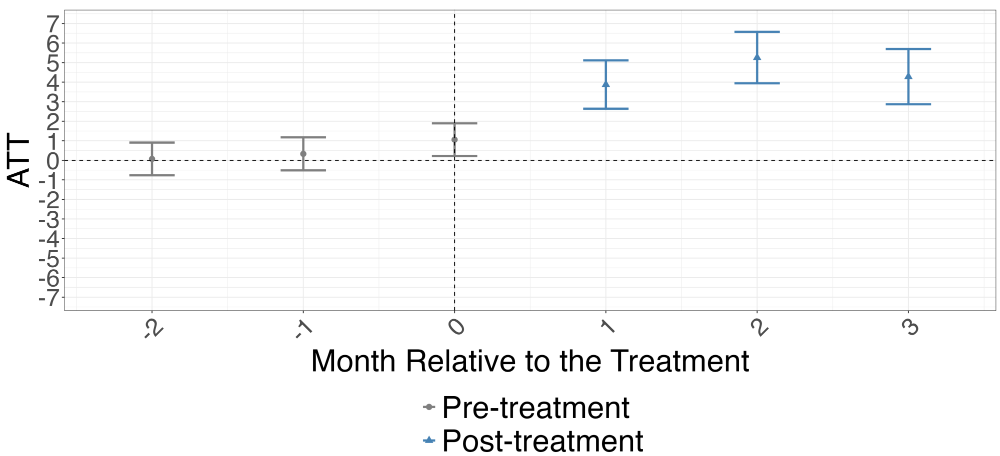


Figure W6: Impact of Live Streaming on Consultation Demand

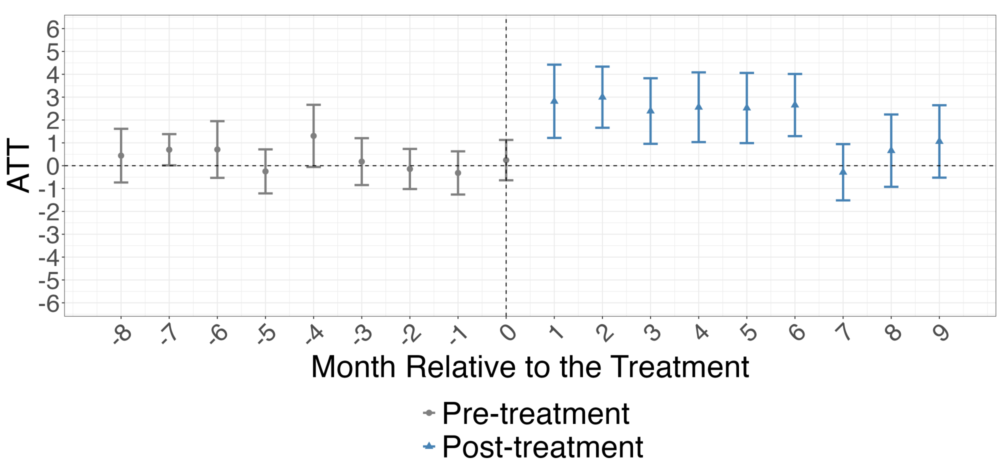


Figure W7: Impact of Live Streaming on Consultation Demand

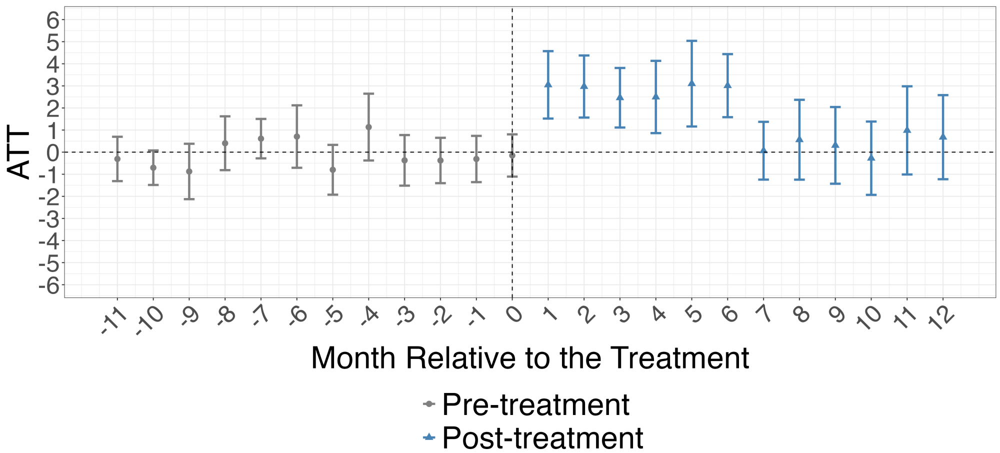


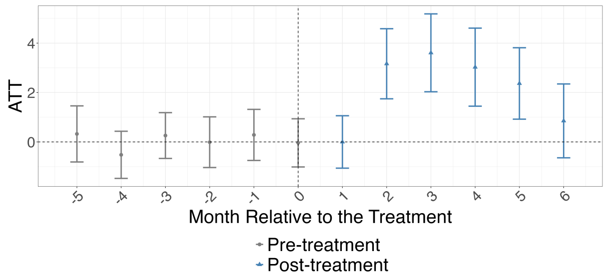
Figure W8: Impact of Live Streaming on Consultation Demand

Placebo Tests

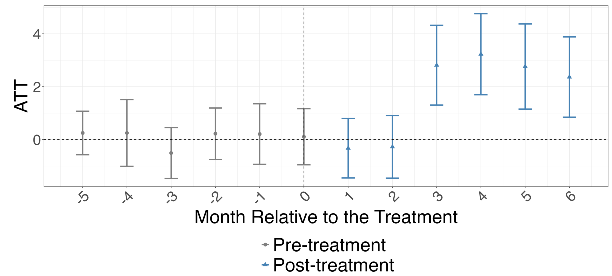
We conduct two sets of placebo tests in this section. First, we shift each treated physician's live streaming adoption date backward by 1 to 6 months before their actual adoption month and re-estimate the model. Figure W9 presents the results: the pre-treatment ATT estimates are mostly centered around zero and statistically insignificant before the actual treatment time, indicating no anticipatory effects and lending support to the parallel trends assumption.[6]

Second, we randomly assign streaming adoption times to (randomly chosen) 2,271 physicians from the control group and re-estimate the model using these placebo-treated physicians as the treatment group and the remaining control group physicians as the comparison group. Figure W10 shows that the estimated ATTs are statistically insignificant from zero in both the pre- and post-treatment periods, confirming that our main results are not driven by spurious correlations or unobserved confounders among untreated physicians.

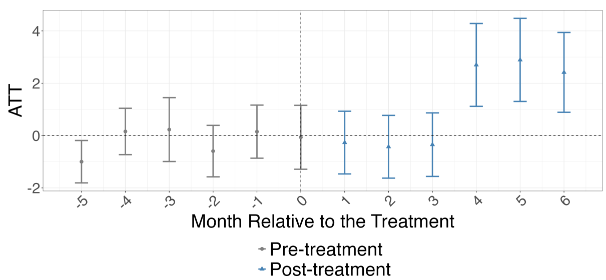
[6]Note that the pre-treatment matching degrades as the placebo shift window gets longer. This is mostly due a reduction is sample size as more and more physicians drop out of the sample.



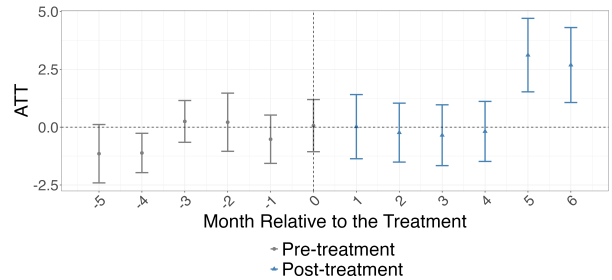
(a) 1 Month Before



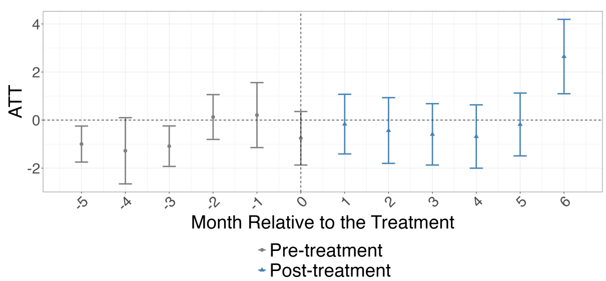
(b) 2 Months Before



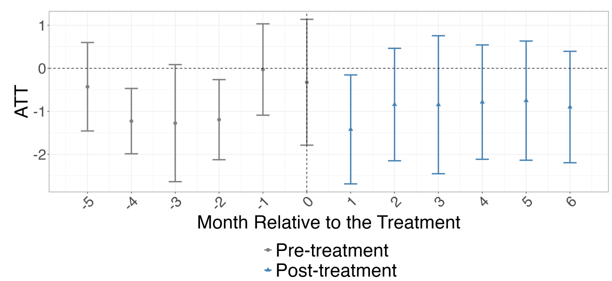
(c) 3 Months Before



(d) 4 Months Before



(e) 5 Months Before



(f) 6 Months Before

Figure W9: Placebo Tests: Shifting Treatment Timing

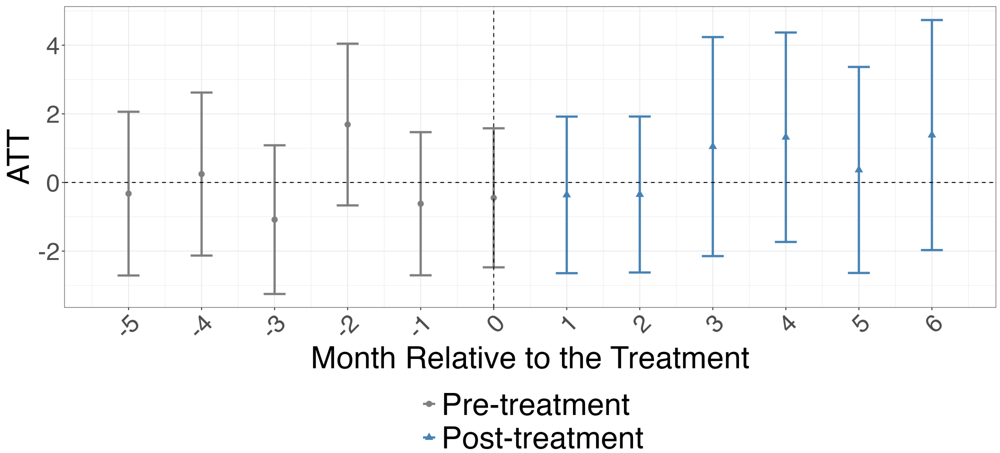


Figure W10: Placebo Tests: Randomly Assigned Treatment in Control Sample

Testing the Difference in ATT across Specifications

In this section, we test whether the ATT estimates for Months 1, 2 and 3 differ significantly between the specifications with and without ranking controls (columns (1) and (2) of Table 3). To do so, we follow the procedure suggested in Blackburn et al. (2021). Specifically, we compute the period-by-period difference in ATT between the two specifications, [ATT(with ranking) − ATT(without ranking)], and construct bootstrap standard errors for this difference using the standard deviation of the 1000 bootstrap draw differences. This accounts for the correlation between the two estimators arising from the common sample. We repeat the same procedure for the average ATT over the full post-treatment window, [ATT (Months 1-6)].

The results are reported in Table W2. The t-statistic tests the null hypothesis that the two specifications produce the same ATT. We cannot reject this null hypothesis, with p-values ranging from 0.87 to 0.95, confirming that our baseline ATT results are robust to the inclusion of ranking controls.

Table W2: Bootstrap Test for Differences in ATT Across Specifications

	Difference	SE	t statistic	p value
ATT (Month 1)	0.1236	0.8825	0.14	0.889
ATT (Month 2)	0.1365	0.9921	0.138	0.891
ATT (Month 3)	−0.1552	0.9548	−0.163	0.871
ATT (Month 1–6)	−0.0431	0.6266	−0.069	0.945

Alternative Estimation Group Feature Importance

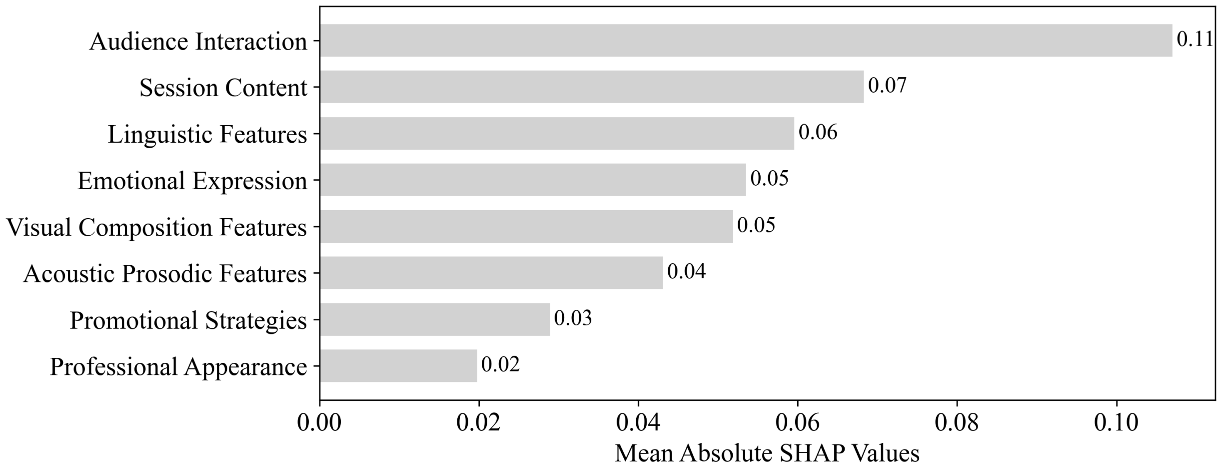


Figure W11: Robustness of Feature Group Importance with Weights

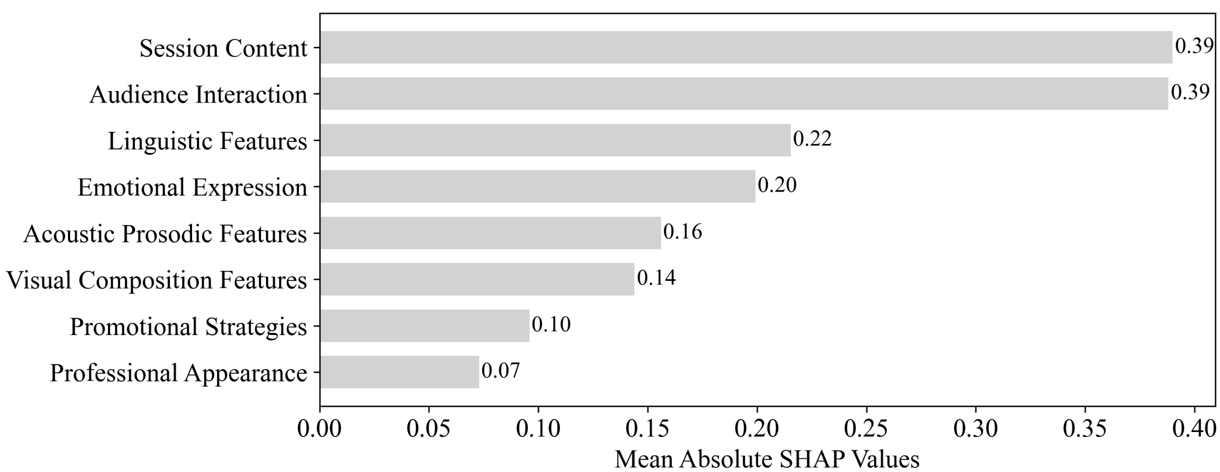


Figure W12: Robustness of Feature Group Importance: Top-Three Features per Group

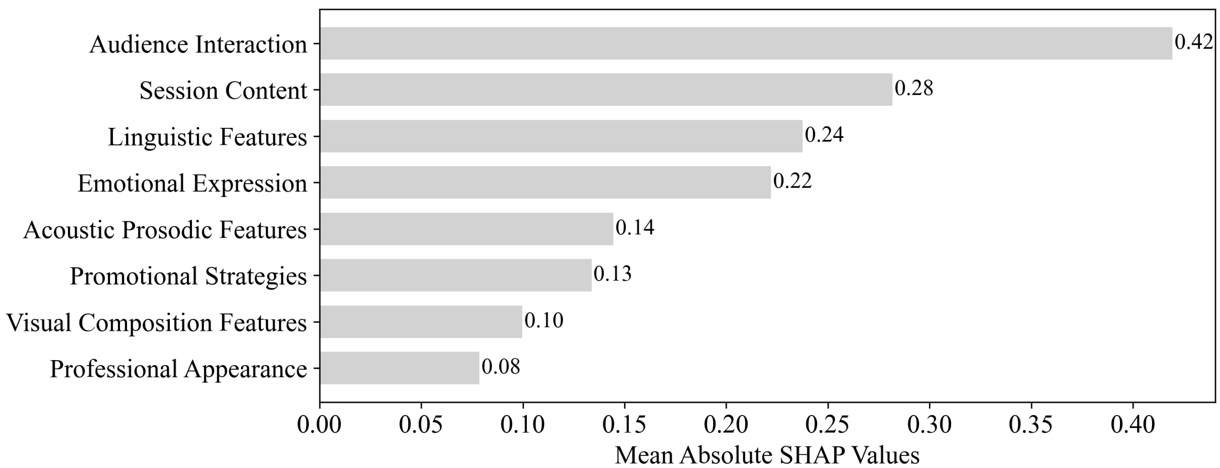


Figure W13: Robustness of Feature Group Importance: Alternative Linear Model

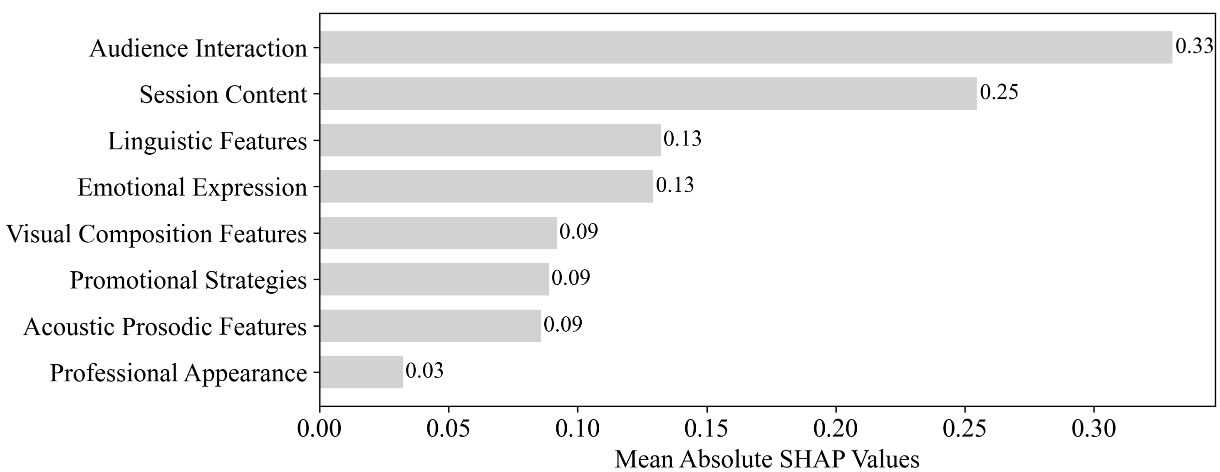


Figure W14: Robustness of Feature Group Importance: Alternative Nonlinear Model

Web Appendix E: Sensitivity Checks on Data Cleaning Process

Missing at Random (MAR) Check

To assess whether missing values may bias our findings, we test whether physician–month missingness differs systematically between treated and control groups. For each key variable (number of consultations, price, ranking index, and patient ratings on effectiveness and attitude), we construct a missing-value indicator and regress it on treatment status. Table W3 shows no meaningful differences in missingness for consultations, price, or ranking between treated and control groups, suggesting that missing values for these variables are unlikely to bias our main results. Missingness for patient ratings declines modestly after live streaming adoption, likely because higher consultation volume increases the chance that at least one patient leaves a rating in a given month. In our regressions, we account for this by including a missing indicator whenever patient ratings enter the specification.

The results, reported in Table W3, show that the estimated coefficients are either statistically insignificant or very close to zero. This indicates that the patterns of missingness are comparable across treated and control groups, suggesting that missing values are not correlated with live streaming and are therefore unlikely to bias our main results.

Table W3: Missing Values Incidence

	(1) Consultations	(2) Price	(3) Ranking	(4) Effectiveness Rating	(5) Attitude Rating
ATT	-0.00432	-0.00000	-0.00000	-0.02404	-0.02431
	(0.00459)	(0.00000)	(0.00000)	(0.00756)	(0.00752)
	[0.34596]	[0.00014]	[0.21846]	[0.00148]	[0.00123]

Notes: Physician fixed effects and year-month fixed effects are included in all specifications. Robust standard errors are reported in parentheses. P-values are reported in square brackets below the standard errors.

Including Physicians with Missing First-Month Consultations

In our main regressions, we exclude physicians whose number of consultations is missing in the first month they adopted live streaming. In this section, we conduct sensitivity checks to assess whether our results are robust to this exclusion. We consider two approaches for including these physicians. First, we include them while leaving the number of consultations for their first month post live streaming as missing. The estimation results, shown in Figure W15, are consistent with our main findings. Second, we include physicians with missing first-month consultation numbers and impute these values using pro-rated differences between the last available observation date and the next available observation date. The estimation results, plotted in Figure W16, also align with our main findings. The only deviation is a slightly positive estimate at period $t = 0$, which likely reflects pro-rated calculations capturing part of the live streaming effect from $t = 1$. We acknowledge this drawback, which motivates our decision to exclude these physicians from the main analysis. Nevertheless, even when included, the estimated results remain very close to our baseline findings.

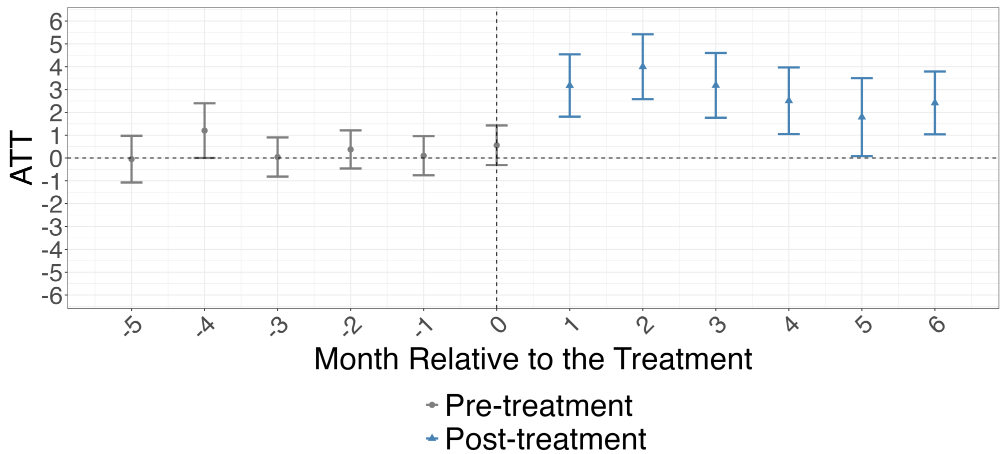


Figure W15: Impact of Live Streaming on Consultation Demand - With Missing Demand

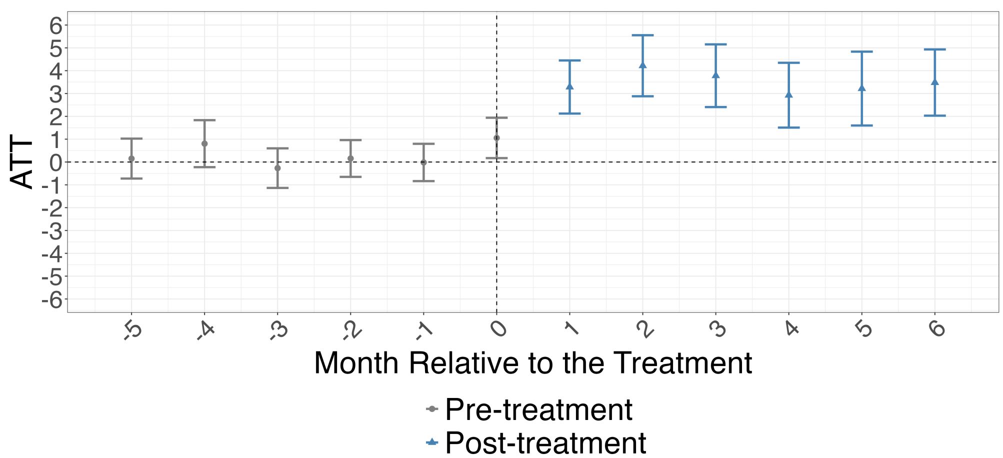


Figure W16: Impact of Live Streaming on Consultation Demand - Fill Missing Demand

Including Physicians with Missing Video Features

In our main regressions, we exclude physicians whose first live streaming video features cannot be constructed. This occurs, for example, when a physician is positioned too far from the camera and their face cannot be detected, making it impossible to generate facial emotion measures. To test whether our GSC estimation results are sensitive to this sample selection rule, we re-estimate the model including these physicians. The ATT results, presented in Figure W17, are similar to our main findings.

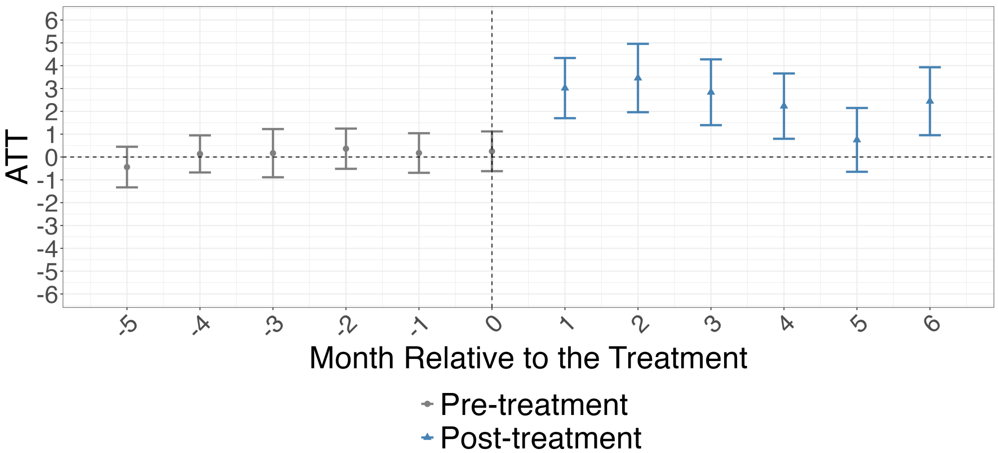


Figure W17: Impact of Live Streaming on Consultation Demand

Including All Non-Streaming Physicians in the Control Group

In this section, we test whether our results are sensitive to the choice of control sample. Specifically, we estimate a GSC model using all non-streaming physicians as the control group and present the results in Figure W18. The pre-treatment fit worsens, which is consistent with Abadie, Diamond, and Hainmueller (2010)'s argument that synthetic control methods work best with a smaller donor pool of untreated units that are similar to the treated units in the predictor space, thereby reducing risks of overfitting and interpolation bias. Nevertheless, even with all non-streaming physicians included and poorer matching quality, the estimated treatment effects remain consistent with our main findings. This further confirms that our conclusions are not sensitive to the choice of control sample.

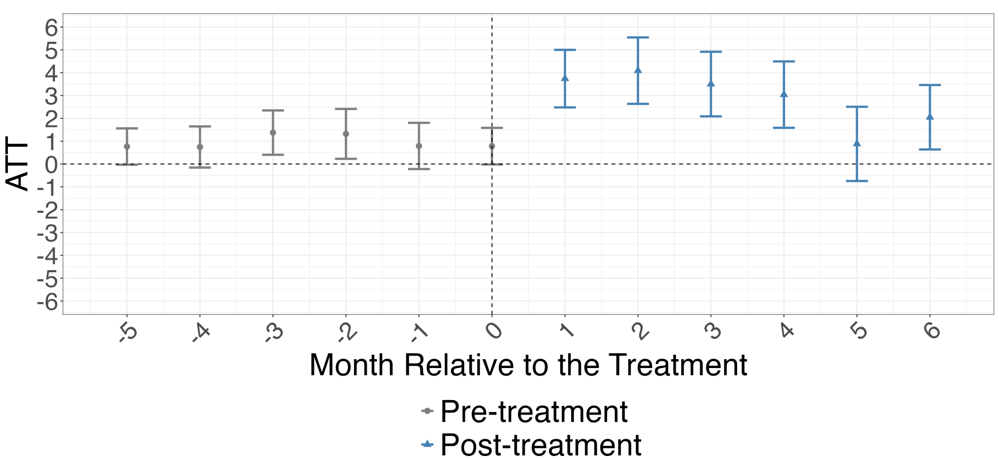


Figure W18: Impact of Live Streaming on Consultation Demand - All Non-Streaming Physicians

Exclusion of Physicians Who Conduct Live Streams on Other Platforms

It is possible that physicians who conduct live streaming sessions on Haodf are also active on other social media platforms. If their activities on other platforms are not strongly correlated with live streaming adoption on Haodf, then physician fixed effects and other controls should account for this. However, if a physician's adoption of live streaming on Haodf coincides with more intensive activity on other social media platforms, our estimates could be biased.

To address this concern, we search for physicians in our sample by name on the two other most popular social media platforms in China, Weibo and Douyin (the Chinese counterpart of TikTok). We find only three physicians who simultaneously conduct live streaming on Haodf and maintain an active presence on other social media platforms. We then repeat our main analysis excluding these three physicians and the results remain robust (see Figure W19).

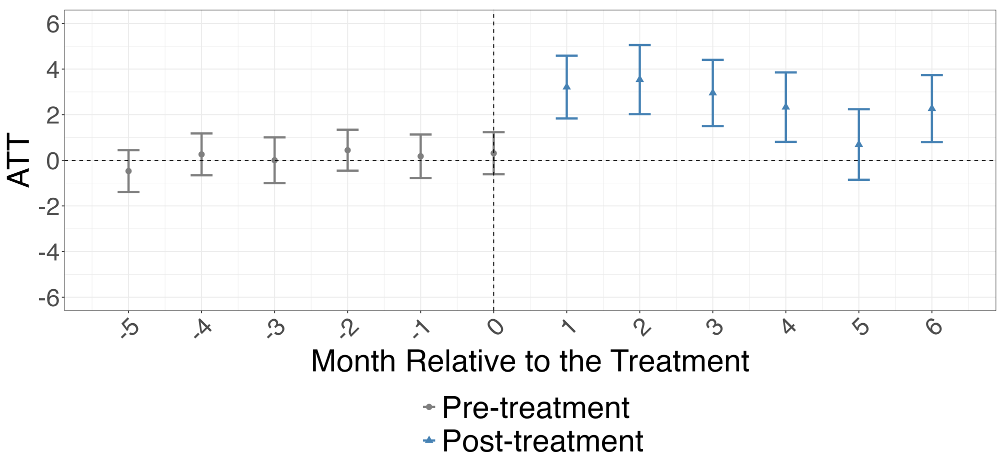


Figure W19: Impact of Live Streaming on Consultation Demand

Web Appendix F: Technical Details of Text Analysis

Post-Transcription Text Refinement

We employ the *SenseVoice* model for speech-to-text transcription. However, substantial domain-specific terminology and contextual complexity in medical live streams can introduce transcription errors and ambiguities. To mitigate these limitations, we utilize the large language model *DeepSeek-V3* to enhance the accuracy and consistency of the transcribed text. The following prompt is used to guide the correction of transcriptions generated through the speech-to-text process.

You are a helpful medical assistant tasked with carefully proofreading a transcript generated from a physician's livestream. Please thoroughly analyze the text and systematically correct any potential errors, including those related to speech recognition, miswritten medical terminology, and incorrect punctuation. Ensure that the revised text is accurate, logically coherent, and conforms to standard medical terminology. Only modify the erroneous or inaccurate parts while retaining the original structure of the text. Please output only the corrected text, without any explanatory or instructional content.

Q&A Extraction

After obtaining the livestream transcripts, we extract a series of linguistic features. First, we identify all question–answer (QA) exchanges between physicians and viewers. Owing to fragmented expressions, implicit references, and informal conversational patterns, rule-based methods exhibit limited performance in detecting QA boundaries. We therefore employ the large language model *DeepSeek-V3* to automatically extract QA pairs. The following prompt is used to guide the extraction process.

You are a helpful medical assistant tasked with analyzing a transcript of a physician's livestream. The physician provides medical popularization during the livestream and replies to questions from viewers. Carefully analyze the transcript and extract viewer questions and the corresponding physician replies. When extracting, preserve the original wording of each question and each reply exactly as they appear in the transcript — do not summarize, paraphrase, or use ellipses ("...") to shorten content. Output the extracted question–answer pairs as a JSON object that follows the format below.

Required output format:

- *If question–answer pairs are present, output a JSON object in which each key is a viewer's question (exactly as in the transcript) and its value is the physician's reply (exactly as in the transcript).*
- *If there are no viewer questions in the transcript, output an empty JSON object:* {}.
- *Output only the JSON result and no additional text or commentary.*

Example input:

"Is Ewing sarcoma highly curable? If Ewing sarcoma is localized and receives standardized surgery and radiotherapy, the cure rate is quite high — roughly seventy to eighty percent can be completely cured. However, if it is metastatic at presentation, the prognosis is poorer, with about a twenty to thirty percent chance of cure."

Example JSON output:

{"Is Ewing sarcoma highly curable?":"If Ewing sarcoma is localized and receives standardized surgery and radiotherapy, the cure rate is quite high — roughly seventy to eighty percent can be completely cured. However, if it is metastatic at presentation, the prognosis is poorer, with about a twenty to thirty percent chance of cure."}

Correction of Extracted QA Pairs

To further improve the accuracy of QA extraction and reduce the potential risk of hallucinations in large language models, we perform a secondary correction using the updated large language model DeepSeek-R1. This additional step is necessary because the initial extraction might still contain inaccuracies or mismatches between questions and answers. By leveraging DeepSeek-R1's enhanced reasoning and factual verification capabilities, we aim to ensure higher precision in the final QA dataset. The following prompt is used to guide the secondary correction process.

You are a professional medical text proofreading assistant responsible for accurately correcting Q&A pairs extracted from physician livestream transcripts.

Tasks:

1. *Filter audience Q&A pairs: Determine whether each extracted Q&A pair consists of an audience question and the corresponding physician reply. Remove non-audience questions, irrelevant content, and duplicate questions or answers.*

2. *Complete missing Q&A pairs: Review the livestream transcript to ensure all audience questions and corresponding physician replies are captured. If any are missing, re-extract and add them. Organize the extracted Q&A pairs in dictionary format, with the audience question as the key and the physician's reply as the value.*

3. *Check completeness of questions and replies: Ensure that audience questions and physician replies are fully and clearly extracted. If key content is missing (e.g., truncated, omitted, or skipped sentences), re-extract the pair.*

4. *Eliminate model hallucinations: Rigorously verify that all Q&A content matches the livestream transcript. Remove any fabricated, altered, or off-text Q&A pairs, and re-extract the correct content.*

5. *Standardize punctuation: Add appropriate punctuation (e.g., periods, question marks) to Q&A pairs lacking punctuation to ensure compliance with written expression standards.*

Output only the corrected Q&A pairs. Do not include any additional explanatory text.

Q&A Classification

We categorize the Q&A pairs into five groups based on the intent of the audience questions: disease diagnosis, disease treatment, drug usage, prevention care, and other. Classification is performed using the *Doubao-1.5-32k* large language model. The prompt used is as follows:

You are a helpful medical assistant. Please strictly classify each question into one and only one of the following five categories based on the question and the physician's response:

1. *Disease Diagnosis: Includes questions about symptom analysis, interpretation of examination reports, inference of causes, or determining whether a medical consultation is needed. Examples: "Is persistent low fever a sign of COVID-19?" "Is it serious to find a lung nodule during a physical examination?"*

2. *Disease Treatment: Involves questions about treatment options (surgical, pharmaceutical, physical therapy, etc.), evaluation of treatment efficacy, and management of side effects. Examples: "Is chemotherapy necessary for breast cancer?" "How to perform conservative treatment for lumbar disc herniation?"*

3. *Drug Usage: Covers questions about drug indications, dosages, side effects, drug interactions, and alternative medications. Examples: "Should I make up for a missed dose of antihypertensive drugs?" "What cold medicines are safe for pregnant women?"*

4. *Prevention Care: Encompasses questions about disease prevention (vaccination, lifestyle), health management (diet, exercise, mental health), postoperative follow-up, regular check-ups for chronic diseases, screening for high-risk groups, and rehabilitation care. Examples: "How to prevent diabetes?" "What dietary precautions should be taken after cancer surgery?"*

5. *Others: Q&A that do not belong to the above four categories or vague descriptions that cannot be clearly classified.*

Output requirements:

1. *Only output one of the five categories, such as Disease Diagnosis; do not add new categories or select multiple categories.*

2. *If a question involves multiple categories (e.g., medication and treatment), classify according to the core intent of the question.*

3. *Do not output any explanations or additional text; output the category only.*

Voucher Count Measure

Physicians sometimes distribute consultation coupons to viewers based on their reported conditions to enhance engagement during live streams. We employ the *Doubao-1.5-32k* large language model to extract the number of consultation coupons that physicians offered in each stream. The prompt is specified as follows:

You are a professional medical data analysis assistant responsible for accurately extracting information on consultation coupon/charity coupon giveaways from physicians' live streaming transcripts. You need carefully analyze the transcript and determine whether the physician explicitly mentions giving away consultation coupons/charity coupons.

Extraction rules:

1. *If an explicit quantity is stated (e.g., "give 3," "give one," "give 5 consultation coupons"), return that number.*

2. *If a giveaway is mentioned without a specified quantity (e.g., "text consultation coupon for you," "consultation coupon for you," "that free coupon for you," "this questionnaire for you," "already gave a charity coupon," "a charity consultation for you," "already given," "already sent"), return 1.*

3. *If the physician provides usage instructions, expected distribution, or summary statements (e.g., "will give consultation coupons," "there will be coupon giveaways," "will distribute 10–20 consultation coupons," "distributed over 30," "sent out 28"), return 0.*

4. *If consultation coupons/charity coupons are not mentioned, return 0.*

Special notes:

1. *Recognize varied expressions (e.g., "give," "give away," "distribute," "provide," etc.).*

2. *Count only quantities the physician explicitly commits to giving.*

Output requirement:

1. *Return only the final integer.*

2. *Do not include any explanations or additional text.*

Service Recommendation Measure

In addition to distributing consultation coupons, physicians may also encourage viewers during live streams to seek consultations with them. We use the Doubao-1.5-32k large language model to identify and count these consultation recommendations. The prompts are defined as follows:

You are a helpful physician's assistant. Based on the live streaming transcript, identify the exact number of times the physician explicitly recommends that patients contact the physician for a consultation or appointment through any service channel.

Eligible cases (all must be met):

1. *The physician explicitly refers to themself as the point of contact (e.g., "find me," "consult me," "register with me," "come see me," "book an appointment with me," "my clinic/my office").*

2. *There is a clear directive to take action to seek care (e.g., "find," "register," "consult," "book," "make an appointment," "inquire," "come").*

3. *The recommendation specifies a service channel, either:*

   (a) *Online via Haodf (e.g., "Haodf platform/Haodf Online/App/Website"), or*

   (b) *Offline/in person in a concrete healthcare setting (e.g., "hospital/clinic/my office/offline").*

Ineligible cases:

1. *Mentions of "online consultation" without "Haodf."*

2. *Ambiguous platform references (e.g., "our platform" only).*

3. *Patient-initiated requests (e.g., "Can I come see you?") without the physician explicitly recommending or directing the patient to do so.*

4. *Vague suggestions without a concrete directive or without referring to the physician (e.g., "You may need offline care," "Just go to the hospital and check.").*

Counting rules:

1. *Repeated recommendations within the same paragraph count as 1.*

2. *Repeated recommendations for the same clinical need within a continuous exchange count as 1.*

Output requirement: Return only an Arabic numeral; no additional explanation.

Linguistic Features Analysis

We construct four linguistic features, which we explain in detail below.

Text Sentiment: We measure text sentiment using the Linguistic Inquiry and Word Count(LIWC) dictionary (Pennebaker et al. 2015), which is widely employed in text semantic analysis (Chang, Mukherjee, and Chattopadhyay 2023; Grewal, Stephen, and Bart 2025). Based on this dictionary, we count the number of positive words (P) and negative words (N) in each text. Sentiment is computed as $Sentiment = \frac{P-N}{P+N+\varepsilon}$, where $\varepsilon = 0.001$ prevents division by zero when no affective words are present. The sentiment score ranges from -1 (negative) to 1 (positive).

Linguistic Complexity: We measure the linguistic complexity of physicians' livestreams using the Gunning–Fog Index(GFI). The calculation is as follows:

$$\text{GFI} = 0.4 \times \left(\frac{\text{Total Words}}{\text{Total Sentences}} + 100 \times \frac{\text{Complex Words}}{\text{Total Words}} \right), \tag{3}$$

The traditional GFI assesses the readability of English text: higher values indicate lower readability and greater linguistic complexity, with "complex words" defined as those containing three or more syllables. Given the absence of syllable-based counting in Chinese, this study operationalizes complex words as adverbs and conjunctions within sentences. These two categories, by encoding logical relations such as contrast, causality, and qualification, increase sentence processing difficulty.

Language Concreteness: We assess language concreteness in physicians' live streams using a specialized dictionary comprising 737,531 Simplified Chinese entries (Xie and Bi 2022), each assigned a concreteness score (for example, "take medicine" has a score of 3.64 and "cold medicine" has a score of 4.61). For each sentence, we calculate the mean concreteness of its constituent words and use this average as the sentence-level concreteness measure.

Q&A Similarity: We compute the similarity between audience questions and physician responses using cosine similarity (Wang and Chaudhry 2018). This method represents texts as term-frequency vectors and calculates the cosine of the angle between them in vector space; higher values indicate greater similarity.

Web Appendix G: Technical Details of Audio Analysis

Speech Emotion Analysis

We employ the *SenseVoice* model to perform speech based emotion recognition in physicians' live streams. The model takes the waveform of each utterance as input and outputs a corresponding emotion label. We identify four emotions (neutral, happy, angry, and sad) and compute their share within each stream. We then recode these into three categories: neutral = 0, positive (happy) = 1, and negative (angry, sad) = -1, from which we derive the mean speech emotion score and its variance.

Acoustic Prosodic Feature Analysis

For the acoustic-prosodic features, we first calculate the physician's average speech rate as total words produced during the live stream divided by the stream duration (in minutes). To reduce computational load, subsequent analyses focus on representative audio segments from the first 10 seconds of each minute (with a sample rate of 48 kHz). We then extract three core acoustic measures using *TorchAudio*, which are associated with consumer persuasiveness (Xu et al. 2025):

Loudness: Measured with torchaudio.transforms.Loudness(), in units of LKFS. This metric reflects the sound energy, with higher values indicating a greater perceived loudness.

Pitch: Measured using torchaudio.functional.detect_pitch_frequency(), in units of Hz. Higher values correspond to a brighter tonal quality as the fundamental frequency moves toward the upper range of the speaker's voice.

Average Spectral Centroid: Measured using torchaudio.functional.spectral_centroid(). This metric integrates pitch and energy distribution to reflect the brightness of the sound. Larger values indicate a higher proportion of high-frequency energy, resulting in a brighter sound.

Web Appendix H: Technical Details of Visual Analysis

Facial Feature Analysis

We utilize the *Aliyun* Vision Analysis API to identify physicians' facial attributes. This API returns multiple facial visual features, including face location, facial attractiveness, presence of eyeglasses, presence of a mask, hat-wearing status, and facial emotion (neutral, happy, angry, sad).

Head Pose Estimation

We employ the machine learning model FSA-Net(Yang et al. 2019) to estimate head pose in real time during physicians' live streams, outputting three Euler angles: pitch (degree of upward or downward tilt), yaw (degree of left or right horizontal rotation), and roll (degree of left or right lateral tilt).

Gaze Estimation

We employ the machine learning model *L2CS-Net* (Abdelrahman et al. 2023) to assess physicians' on-camera gaze engagement during live streams. Gaze engagement was quantified by the angle between the physician's gaze vector and the screen's normal vector.

Clothing Analysis

We deploy the vision-language model Qwen2.5-VL-7B [7] to extract physicians' attire information. To mitigate occlusion and motion blur in videos, we analyze every video frame. The prompt used for attire classification is:

Please determine the attire of the person in the image by selecting one of the following categories: [White Lab Coat, Surgical Scrubs, Suit/Shirt, Casual Wear]. "Casual Wear" refers to any clothing other than a white lab coat, surgical scrubs, or a suit/shirt (e.g., T-shirt). If no person is present in the image or the attire cannot be determined, return "Unable to determine." Do not include any additional explanatory text.

[7] Qwen2.5-VL

Web Appendix I: Fund Manager Results

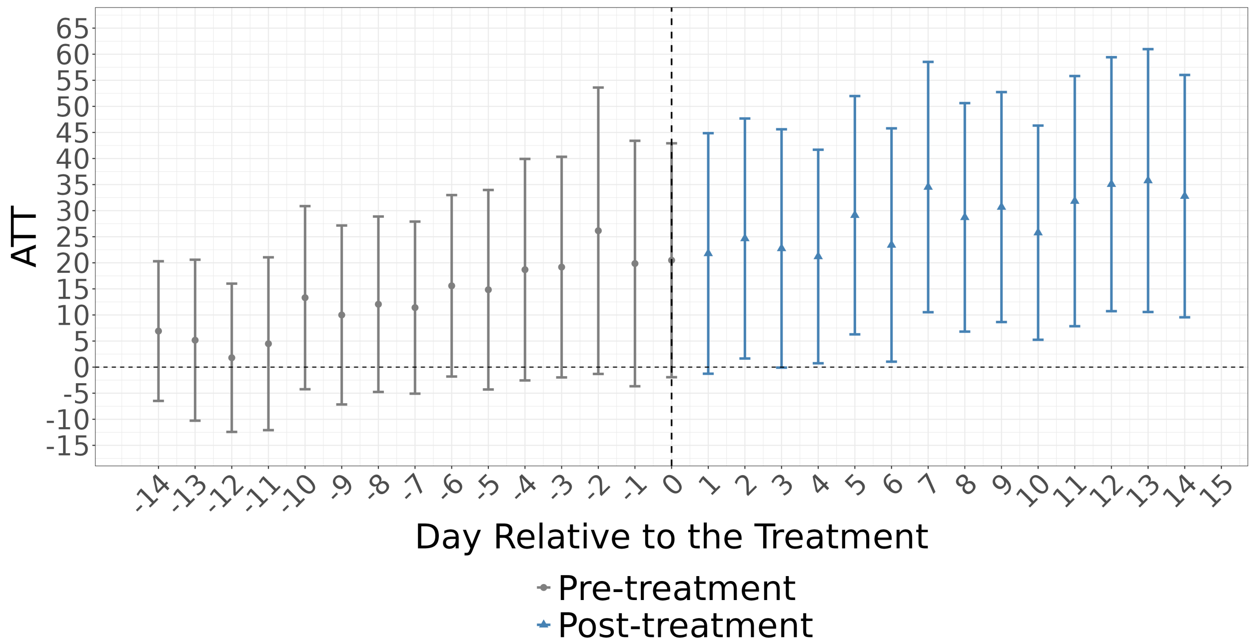


Figure W20: Impact of Live Streaming on Fund Trading Volume

Notes: This figure shows the estimated ATT on the fund trading volume and their 95% confidence intervals for each trading day before and after live streaming mention date. The estimation includes fund fixed effects and trading-day fixed effects. The standard errors are based on parametric bootstraps of 1,000 times.

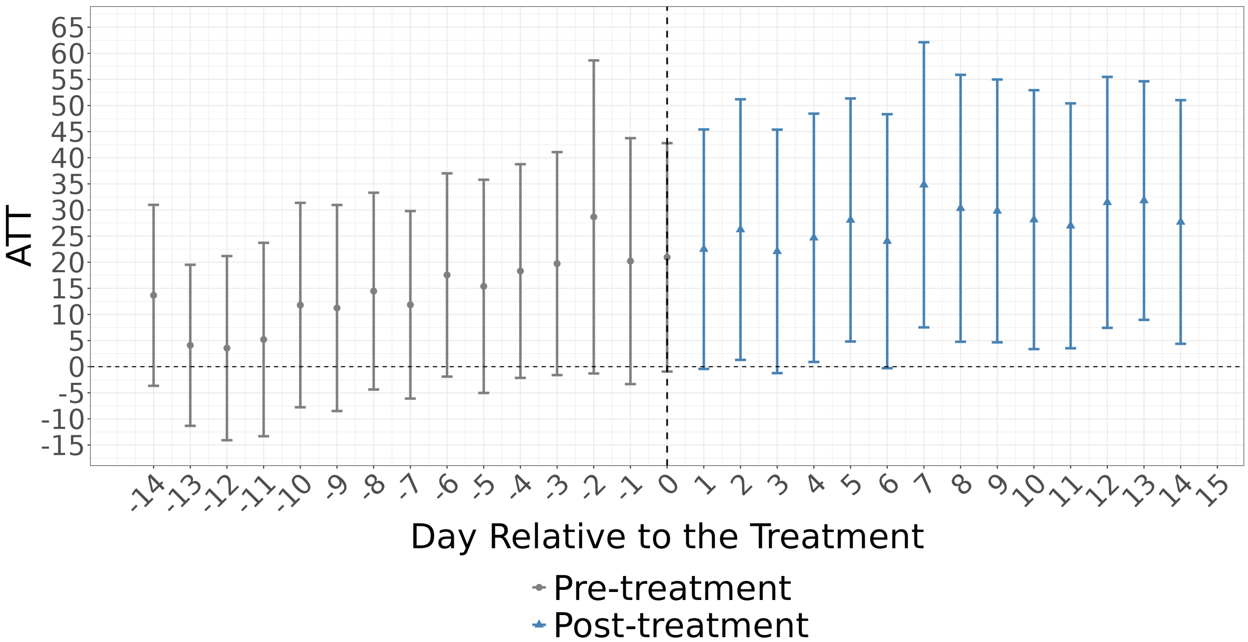


Figure W21: Robustness of the Impact of Live Streaming on Fund Trading Volume

Notes: This figure shows the estimated ATT on the fund trading volume and their 95% confidence intervals for each trading day before and after the live streaming mention date. The sample includes live stream sessions in which only one ETF is mentioned. The estimation includes fund fixed effects and trading-day fixed effects. The standard errors are based on parametric bootstraps of 1,000 times.

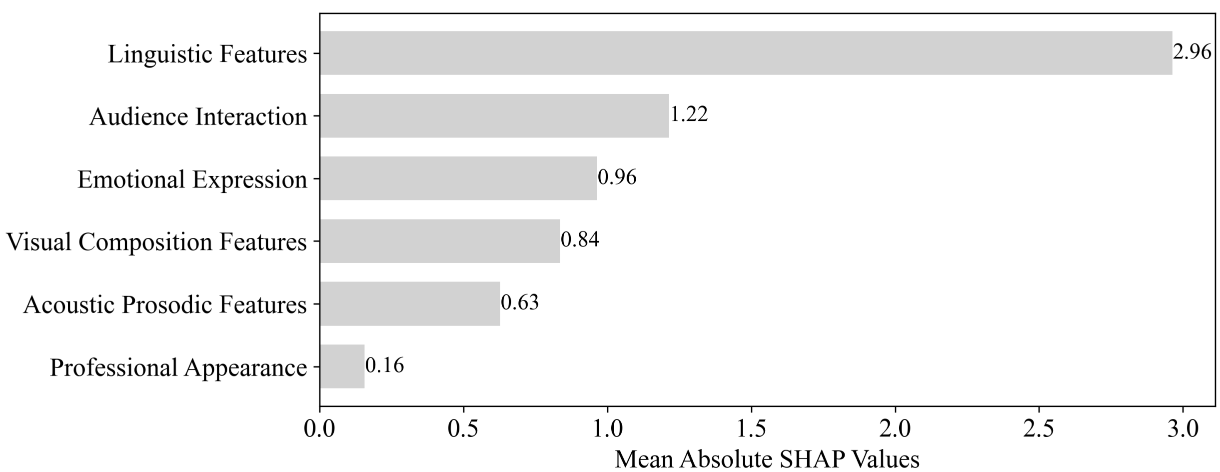
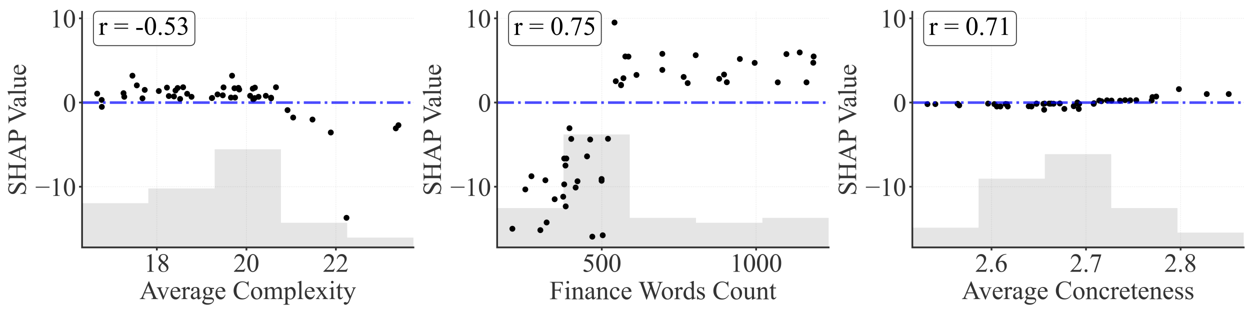
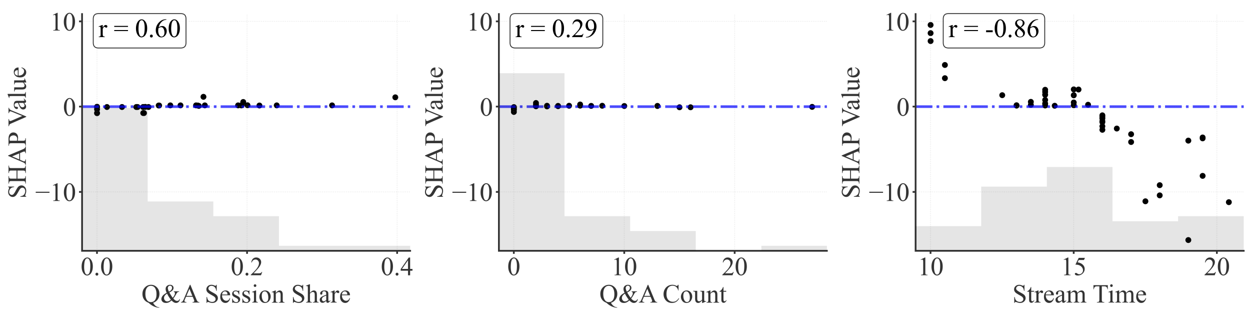


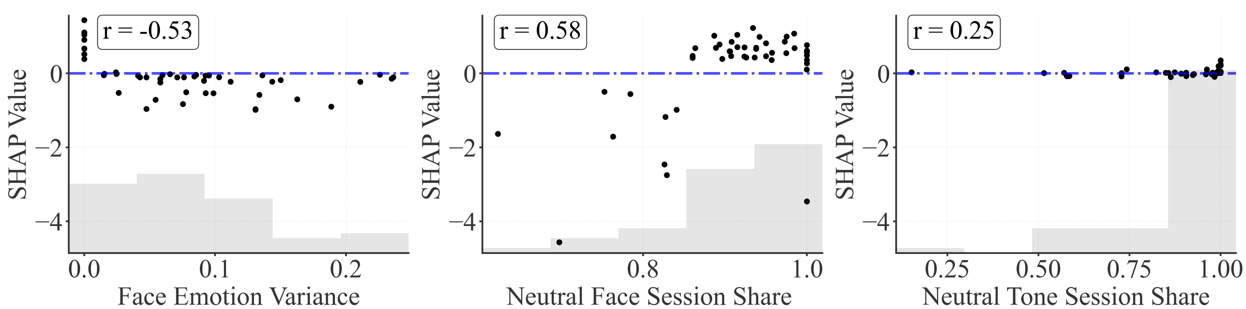
Figure W22: Feature Group Importance for Fund Manager Live Streaming



(a) Linguistic Features



(b) Audience Interaction



(c) Emotional Expression

Figure W23: Individual Feature Importance within Feature Group for Fund Manager Live Streaming

Notes: Scatter points represent individual sample observations in the test set, while the shaded background histograms illustrate the probability density of feature distributions. The X-axis denotes feature values, the Y-axis represents the corresponding SHAP values.

Web Appendix J: Qualitative Evidence from Physician Interviews

We conduct ten in-depth interviews with physicians who provide services on the Haodf platform. The interviews cover physicians' motivations, experiences, and concerns regarding participation in online consultations and live streaming. Basic information on these physicians, including their professional title, specialty, whether they work in tertiary hospitals, and whether they have conducted live streaming, is reported in Table W4.

Table W4: List of Interviewed Physicians

No.	Professional Title	Specialty	Tertiary Hospital	Live Streaming
1	Chief	Ob&Gyn	No	Yes
2	Resident	Dentistry	Yes	No
3	Resident	Internal Medicine	Yes	No
4	Chief	Proctology	No	Yes
5	Resident	Ob&Gyn	Yes	No
6	Chief	Pediatrics	Yes	Yes
7	Chief	Ob&Gyn	Yes	Yes
8	Attending	Internal Medicine	Yes	No
9	Associate Chief	Orthopedics	Yes	No
10	Associate Chief	Thoracic Surgery	Yes	Yes

Notes: Tertiary Hospital indicates whether the physician is affiliated with a tertiary-level hospital. Live Streaming indicates whether the physician conducts live streaming medical consultations.

For each interview, we adopt a semi-structured format. The interview guide includes the following core questions, while allowing interviewees to elaborate freely based on their experiences:

1. Why did you choose to provide services on Haodf? What is your primary motivation?

For physicians who had adopted live streaming, we asked the following additional questions:

1. Why did you choose to conduct live streaming on Haodf? What is your primary motivation?
2. How many times have you conducted live streaming sessions?
3. Did live streaming meet your expectations? Do you plan to conduct more live streaming sessions in the future?

4. How long does a typical live streaming session last? How much time do you usually spend preparing for a live streaming session?

For physicians who had not adopted live streaming, we ask the following additional questions:

1. Were you aware that physicians could conduct live streaming on the platform? If so, why did you choose not to adopt it?

2. Do you plan to conduct live streaming in the future? What concerns, if any, do you have?

3. What potential benefits do you think live streaming could offer you?

Each interview lasts approximately 20 to 30 minutes. The complete set of interview transcripts is available in Mandarin upon request. In what follows, we present English translations of interview responses organized along three dimensions: motivations for providing online consultations, experiences and outcomes related to live streaming, and reasons for not adopting live streaming.

Motivations for Online Consultation

Interview responses indicate that physicians participate in online healthcare platforms for a combination of altruistic and practical reasons. Across interviews, the most frequently cited motivations include improving patient welfare, expanding access to care, and earning supplementary income.

- Interviewee 1: I participate to increase my income and to help patients.

- Interviewee 2: I began during the COVID period to make use of leisure time and to help returning patients plan offline visits in advance. Online consultation also provides additional income.

- Interviewee 3: I started during COVID when my existing patients could not visit in person. I believe online healthcare represents a future trend and wanted to engage early. Online consultation also improves efficiency, as patients do not need to make multiple hospital visits, which saves time for both physicians and patients.

- Interviewee 4: I want to serve the public. Haodf collaborates with Alipay and attracts substantial traffic, which allows me to reach a large number of patients. I feel a sense of fulfillment when serving patients online. As a physician from a top-tier hospital, I also want to serve patients from less developed regions. Additional income is another motivating factor.

- Interviewee 5: My motivation is to help more patients, gain exposure to a wider range of medical cases, and increase my income.

- Interviewee 6: I provide free consultations to give patients an opportunity to communicate with a physician. I prioritize my primary clinical duties and use Haodf mainly during my leisure time.

- Interviewee 7: My primary goal is to serve patients who are not able to attend office visits. Some patients live far from hospitals and face high travel costs to see physicians. Through online consultations, I can conduct preliminary screening and provide medical advice on whether an offline visit is worthwhile and what materials might be useful to bring.

- Interviewee 8: My primary motivation is to earn additional income.

- Interviewee 9: My motivations are to increase my income and to help address patients' medical concerns. Online consultations also facilitate subsequent in-person visits for some patients.

- Interviewee 10: My main objective is to provide health education. In addition, I aim to attract new patients and better serve my existing patient base.

Live Streaming Experience and Outcomes

We asked physicians who had conducted live streaming about their experiences, satisfaction, and perceived outcomes. Overall, respondents viewed live streaming as a valuable complement to online consultations, particularly for patient education, communication, and outreach.

- Interviewee 1 (Streamer): I have streamed multiple times and plan to continue because I want to help patients. My live streams typically last about one hour, and I do not require preparation time. The main benefits are accumulating followers and helping more patients. I initially hoped live streaming would increase my appointments, but since I only recently started, I have not yet observed such conversion.

- Interviewee 4 (Streamer): I plan to continue streaming. I started because I view online healthcare as a future trend. My primary goal is to address patients' medical problems and relieve their pain. However, my expectations were not fully met. As a proctologist, I need in-person examinations and visual information. Many patients ask relatively simple questions, but I cannot adequately address individual conditions through verbal communication alone. I usually stream for 30 minutes to one hour without preparation. Despite these limitations, live streaming has increased my appointments. I am semi-retired, so I view this activity primarily as public service rather than a source of income.

- Interviewee 6 (Streamer): I have conducted two live streams focused on educational content related to specific diseases. As a pediatrician, I treat conditions that most parents encounter. My goal was to educate parents about these common issues. However, the volume of questions was overwhelming, and I could not answer all of them. I usually stream for nearly one hour and require no preparation. Live streaming has increased my appointments. In addition, physicians learn from each other's live streams and share knowledge with colleagues.

- Interviewee 7 (Streamer): I have conducted more than three live streams and plan to continue if time permits. I enjoy the process and am satisfied with the outcomes. Live streaming reduces communication costs, decreases unnecessary consultations, and lowers my work pressure. I typically stream for about one hour without preparation. Live streaming has clearly increased my appointments, and patients often leave comments asking about my clinic hours and location.

- Interviewee 10 (Streamer): To date, I have conducted approximately five to six sessions, each lasting about one hour and requiring no preparation. I plan to continue depending on my schedule, likely one to two times per month. Compared with recording short videos, I find live streaming easier and more effective for physician–patient communication. Live streaming also increases consultation volume, as communication during sessions is limited and patients often seek follow-up consultations either online or in person.

Reasons for not Live Streaming

We ask physicians who had not conducted live streaming about their reasons for non-adoption and any concerns they might have. Overall, respondents cite time constraints, uncertainty about returns, institutional barriers, and personal discomfort with live formats as key reasons for not engaging in live streaming. At the same time, most physicians acknowledge the potential benefits of live streaming for patient education, professional development, and patient acquisition, and several indicated openness to future adoption under more favorable conditions.

- Interviewee 2 (Non-streamer): I am aware of live streaming but currently have no plans to adopt it. I am already very busy with offline work, do not see clear returns from live streaming, and lack the time and energy required. I also have an introverted personality and feel uncomfortable speaking in front of a large audience. I am uncertain how to structure a live streaming session and worry about awkward situations, such as a lack of audience interaction. There has been no formal training on live streaming. That said, I recognize that for physicians who operate private clinics, live streaming can help build personal reputation and attract traffic. I would like to learn from successful streamers and may consider trying it in the future once I better understand how to do it effectively.

- Interviewee 3 (Non-streamer): I am aware of live streaming but do not currently have sufficient time. Physicians who engage in live streaming tend to be specialists, whereas my work as a general physician focuses on chronic disease management and initial consultations. I therefore do not expect strong demand from live streaming sessions. My priority is to fulfill my offline responsibilities. Nevertheless, I recognize that live streaming can better serve patients, as it allows for immediate responses, whereas text-based consultations are often conducted in fragmented time.

- Interviewee 5 (Non-streamer): I am aware of live streaming but have not adopted it because participation requires approval from my hospital, and the approval process is lengthy and burdensome. However, I believe live streaming can help patients better understand healthcare processes and medical knowledge. It can also increase physicians' visibility and generate traffic, which may in turn lead to more online consultations.

- Interviewee 8 (Non-streamer): I am aware of live streaming, but my daytime work schedule and family responsibilities in the evening leave me with limited time and energy to engage in it. I plan to explore live streaming during my child's school holidays. I am also concerned about the unpredictability of live streaming, such as patients repeatedly asking the same questions. At the same time, I believe live streaming can support professional development by helping physicians consolidate their knowledge and can serve as an effective channel for patient education. Many patients lack basic medical knowledge, which can hinder effective physician–patient communication.

- Interviewee 9 (Non-streamer): I am aware of live streaming, but I am concerned that my communication skills may not be well suited to a live streaming setting. In addition, I do not have sufficient spare time and am unfamiliar with the platform's rules and procedures. Although I believe live streaming has the potential to convert viewers into patients, I am currently unsure how to achieve such conversion effectively.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller (2010), "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program," *Journal of the American Statistical Association*, 105 (490), 493–505.

Abdelrahman, Ahmed A, Thorsten Hempel, Aly Khalifa, Ayoub Al-Hamadi, and Laslo Dinges "L2cs-net: Fine-grained gaze estimation in unconstrained environments," "2023 8th International Conference on Frontiers of Signal Processing (ICFSP)," pages 98–102, IEEE (2023).

Blackburn, Justin, Constantin T Yiannoutsos, Aaron E Carroll, Paul K Halverson, and Nir Menachemi (2021), "Infection fatality ratios for COVID-19 among noninstitutionalized persons 12 and older: results of a random-sample prevalence study," *Annals of Internal Medicine*, 174 (1), 135–136.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess (2024), "Revisiting event-study designs: robust and efficient estimation," *Review of Economic Studies*, 91 (6), 3253–3285.

Callaway, Brantly and Pedro HC Sant'Anna (2021), "Difference-in-differences with multiple time periods," *Journal of Econometrics*, 225 (2), 200–230.

Chang, Hannah H, Anirban Mukherjee, and Amitava Chattopadhyay (2023), "More voices persuade: The attentional benefits of voice numerosity," *Journal of Marketing Research*, 60 (4), 687–706.

Goodman-Bacon, Andrew (2021), "Difference-in-differences with variation in treatment timing," *Journal of Econometrics*, 225 (2), 254–277.

Grewal, L, A Stephen, and Y Bart (2025), "Online venting: the impact of temporal proximity cues and emotionality on perceptions of negative online review value," *Journal of Marketing Research*, forthcoming.

Pennebaker, JW, RJ Booth, RL Boyd, and ME Francis (2015), "Linguistic inquiry and word count: LIWC2015. 2015," *Austin, TX: Pennebaker Conglomerates (www. LIWC. net).*

Sun, Liyang and Sarah Abraham (2021), "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of Econometrics*, 225 (2), 175–199.

Wang, Yang and Alexander Chaudhry (2018), "When and how managers' responses to online reviews affect subsequent reviews," *Journal of Marketing Research*, 55 (2), 163–177.

Xie, Zhipeng and Ran Bi (2022), "Construction and inference technique of large-scale Chinese concreteness lexicon," *Acta Scientiarum Naturalium Universitatis Pekinensis*, 58 (1), 1–6.

Xu, Haifeng, Yi Ding, Yu Ding, Qi Zhang, and Cheng Zhang (2025), "Whispers in your mind: the role of voice features in customer acquisition and retention," *Journal of Marketing*, forthcoming.

Yang, Tsun-Yi, Yi-Ting Chen, Yen-Yu Lin, and Yung-Yu Chuang (2019), "Fsa-net: Learning fine-grained structure aggregation for head pose estimation from a single image," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1087–1096.

Zhan, Jiajia, Xu Zhang, and Hongqiao Fu (2026), "Information disclosure via platform endorsement in online healthcare," *Journal of Marketing Research*, forthcoming.