Validity of patients' online reviews at direct-to-consumer teleconsultation platforms

Xie, Yunyun; He, Wenjun; Wan, Yuting; Luo, Huanyuan; Cai, Yiyuan; Gong, Wenjie; Liu, Siyuan; Zhong, Dongmei; Hu, Wenping; Zhang, Lanping; Li, Jiaqi; Zhao, Qing; Lv, Sensen; Li, Chunping; Zhang, Zhang; Li, Changchang; Chen, Xiaoshan; Huang, Wangqing; Wang, Yutong; Xu, Dong

DOI:
10.1136/bmjopen-2023-071783

BMJ Open

Validity of patients' online reviews at direct-to-consumer teleconsultation platforms: a protocol for a cross-sectional study using unannounced standardised patients

Yunyun Xie,[1] Wenjun He,[2] Yuting Wan,[1] Huanyuan Luo,[3,4,5] Yiyuan Cai,[6,7] Wenjie Gong ,[8,9] Siyuan Liu,[2,5] Dongmei Zhong,[2,5] Wenping Hu,[10] Lanping Zhang,[1,5] Jiaqi Li,[1,5] Qing Zhao,[5] Sensen Lv,[1,5] Chunping Li,[1,5] Zhang Zhang ,[11] Changchang Li,[3] Xiaoshan Chen,[1,5] Wangqing Huang,[1,5] Yutong Wang,[12] Dong Xu [4,5,13]

To cite: Xie Y, He W, Wan Y, *et al*. Validity of patients' online reviews at direct-to-consumer teleconsultation platforms: a protocol for a cross-sectional study using unannounced standardised patients. *BMJ Open* 2023;**13**:e071783. doi:10.1136/bmjopen-2023-071783

► Prepublication history for this paper is available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2023-071783).

YX and WH are joint first authors.

Received 17 January 2023
Accepted 12 April 2023

For numbered affiliations see end of article.

Correspondence to
Dong Xu; roman.xu@gmail.com

ABSTRACT

Introduction As direct-to-consumer teleconsultation (hereafter referred to as 'teleconsultation') has gained popularity, an increasing number of patients have been leaving online reviews of their teleconsultation experiences. These reviews can help guide patients in identifying doctors for teleconsultation. However, few studies have examined the validity of online reviews in assessing the quality of teleconsultation against a gold standard. Therefore, we aim to use unannounced standardised patients (USPs) to validate online reviews in assessing both the technical and patient-centred quality of teleconsultations. We hypothesise that online review results will be more consistent with the patient-centred quality, rather than the technical quality, as assessed by the USPs.

Methods and analysis In this cross-sectional study, USPs representing 11 common primary care conditions will randomly visit 253 physicians via the three largest teleconsultation platforms in China. Each physician will receive a text-based and a voice/video-based USP visit, resulting in a total of 506 USP visits. The USP will complete a quality checklist to assess the proportion of clinical practice guideline-recommended items during teleconsultation. After each visit, the USP will also complete the Patient Perception of Patient-Centeredness Rating. The USP-assessed results will be compared with online review results using the intraclass correlation coefficient (ICC). If ICC >0.4 ($p<0.05$), we will assume reasonable concordance between the USP-assessed quality and online reviews. Furthermore, we will use correlation analysis, Lin's Coordinated Correlation Coefficient and Kappa as supplementary analyses.

Ethics and dissemination This study has received approval from the Institutional Review Board of Southern Medical University (#Southern Medical Audit (2022) No. 013). Results will be actively disseminated through print and social media, and USP tools will be made available for other researchers.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ We will assess the validity of physician online reviews from the three largest teleconsultation platforms, accounting for almost 80% of the total teleconsultation visits.
⇒ We will use unannounced standardised patients (USPs), representing 11 common primary care conditions as the 'gold standard' of quality assessment.
⇒ Our sampling of physicians simulates an actual consultation process, both technical quality and patient-centredness will be assessed.
⇒ Our sample will not include individual hospital-based teleconsultation sites, although they represent only a minor proportion of teleconsultation in China.
⇒ A given physician in our study is evaluated only by a single USP case, although we have well-balanced common cases at the aggregated level of analysis.

Trial registration The study has been registered at the China Clinical Trials Registry (ChiCTR2200062975).

INTRODUCTION

Chinese government policies and the COVID-19 pandemic have stimulated the rapid growth of Direct-to-consumer (DTC) telemedicine since its initiation at the end of the 20th century in China.[1 2] Unlike traditional medical models, DTC telemedicine enabling patients to receive medical advice and treatment through electronic media (such as computers, phones or smartphones) without establishing a prior doctor–patient relationship.[3] The Institute of Medicine defines telemedicine as 'the use of electronic information and communication technologies to provide and support healthcare when

the participants are at a distance from each other'.[4] DTC teleconsultation (hereafter 'teleconsultation') is a basic form of DTC telemedicine.[5] Teleconsultation is an electronic approach for clinician–patient communication for follow-up care or patients' inquiries about their health,[5–7] using text, pictures, telephone or videos. Patients have increasingly used teleconsultation as an alternative to in-person care.[8–10] In China, the number of teleconsultation users increased from 214.8 million in December 2020 to 239.33 million in June 2021, a growth rate of 11.42%.[11] At Haodf.com alone, one of the popular telemedicine platforms, more than 20 000 doctors provide 200 000 teleconsultations every day.[12]

Teleconsultation allows patients to 'rate a doctor' after an online visit, similar to a customer on an online shopping site.[13] While patients have few means to gauge the physician service quality in brick-and-mortar facilities, online reviews create an easily accessible digital 'word of mouth' for patient-perceived care quality. A recent national survey in the USA showed that 59% of participants referred to online reviews when choosing a physician.[14] Meanwhile, online reviews are a means of audit and feedback for teleconsultation,[15] thus with the potential to improve the quality of healthcare.[16]

However, can online physician reviews accurately reflect the quality of teleconsultation? Given the 'information asymmetry' in healthcare, a fundamental challenge is whether patients can assess the quality of care. And if they can, what aspects of quality are they assessing? Many studies used questionnaires, qualitative interviews and text mining of the online review content to understand the nature of online reviews.[17–20] They often focused on the patient-centred quality aspects such as patient outcomes,[21] patient satisfaction[22 23] and doctor–patient relationship[24–26] rather than the technical quality of care (physician adherence to guideline recommendations, diagnostic accuracy and treatment appropriateness). Even for patient-centred quality, such as patient satisfaction, studies from Widmer *et al*[27] and Ryan *et al*[28] found no significant correlation between physicians' online review ratings and patient satisfaction. In contrast, Reimann *et al* argued that one of the important goals of collecting online reviews is to present information on patient satisfaction.[29] Robert *et al* found that the validity of online reviews varied by medical specialty.[30] Timothy *et al* compared online reviews of physician performance scores and found that online reviews could not measure the actual performance of physicians in terms of quality of service and peer review.[27]

One major reason for the ambiguity about the validity of online reviews may be attributed to the lack of benchmarking the online review results with a 'gold standard' for quality assessment. Unannounced standardised patients (USPs) have been used in many studies to measure the quality of healthcare.[31–33] The USP has many advantages in assessing the quality of care, including the reduced Hawthorne effect,[34 35] minimised recall bias, and standardisation of patent level confounding.[34–39] Therefore, this study will use USPs to evaluate the quality of teleconsultation and then analyse the concordance of the assessment results between online reviews and USP visits in teleconsultations. The higher concordance will indicate higher criterion validity of online reviews as a quality indicator. We further hypothesise that online patient reviews may better align with patient-centred care aspects (ie, patient-perceived care) than the technical quality of care (effectiveness and safety) due to information asymmetry. It should be noted that USP is the gold standard for evaluating the quality of teleconsultations, but it cannot measure patient reviews. This study examines the degree of concordance between online reviews and the quality of medical services.

We thus hypothesise that online patient reviews may be better aligned with the quality components regarding patient-centred care but not with the technical quality of care (effectiveness and safety). Testing of those two hypotheses may advance our understanding of online reviews.

METHODS

Study design

We will use a cross-sectional design. This design aims to evaluate the concordance between patients' online reviews and USP-assessed qualities. This protocol follows the Strengthening the Reporting of Observational Studies in Epidemiology reporting guidelines for observational studies[40] (online supplemental Addendum 1).

Study setting

We will select Ping An Healthcare And Technology Company Limited (hereafter 'Ping An Good Doctor'), Haodf.com and WeDoctor—the three largest teleconsultation platforms for our study. They are teleconsultation versions of amazon.com and taobao.com for eCommerce,[36] accounting for almost 80% of the total market share in patient visits.[41] As of 2019, Ping An Good Doctor has reached 315.2 million registered users, making it the largest mobile medical app in China.[41] What's more, Ping An Good Doctor won the largest share of new users in 2020, with a dominant market share of 66.3%.[41] Haodf.com, founded in 2006, is the earliest teleconsultation platform, hosting nearly 860 000 doctors, among whom nearly 240 000 doctors are Haodf.com authenticated.[42] Haodf.com also hosts the earliest and largest online doctor reviews in China.[43 44] Physicians can register on those sites to deliver teleconsultations. Individual offline hospitals also operate teleconsultation services but at a rather small scale and in a fledging stage. We thus focus on platforms rather than individual hospital teleconsultation sites.

Ping An Good Doctor, Haodf.com, and WeDoctor share similarities in providing online medical consultation services and Artificial Intelligence (AI)-based diagnosis, but differ in their business focus and ownership. Ping An Good Doctor emphasises health management and big

data, while Haodf.com focuses on remote consultation services led by medical experts, and WeDoctor mainly provides mobile healthcare services and owns a pharmacy for drug sales.

Study population

Doctors providing teleconsultations on those three platforms during our study time comprise our study population. Those platforms have slightly different eligibility for doctors to practice on their sites, but generally, physicians must upload copies of their ID cards, professional certificate, medical practitioner's licenses, and medical practitioner's qualification certificates to register for the sites. The sites also use face recognition technology to ensure the authenticity of the application.

Sampling procedures

When patients visit traditional health institutions in China, the 'triage desk' will direct them to appropriate physicians based on their health conditions. All those sites offer similar 'triage desks', which, for instance, is called 'Speedy Match' on Ping An Good Doctor. Thus we will use those 'triage' services to identify doctors for the USP to visit. Those doctors visited by the USP will then be considered part of our sample. We have selected 11 conditions to be role-played by our USPs during their online visit (detailed later).

Sample size calculation

In this study, USP collects multiple measures of the quality of online visits and online review results will be collected from the physician's home page on the teleconsultation platforms. We will use the intracluster correlation coefficient (ICC) to assess the concordance between USP-assessed quality (primarily represented by the proportion of guideline-suggested teleconsultation items) and online review quality (primarily represented by online 'Favourable rate').[45–47] Landis and Koch suggested that ICC from 0 to 0.20 reflects slight concordance, 0.21 to 0.4 fair, 0.41 to 0.60 moderate, 0.61 to 0.8 substantial and above 0.81 almost perfect. Therefore, ICC≥0.4 is set as the effect size for acceptable concordance in this study.

To get the basic data for sample size calculation, we conducted a small pilot on the service quality of teleconsultation by USPs. During the pilot, USPs made 48 visits involving the 11 cases, and we detected an ICC of 0.26. Therefore, we calculated the sample size to ensure that the ICC's lower limit of the one-sided 95% CI is no <0.4 when the experienced value is 0.26 with 80% probability.[46] The sample size was calculated by R package (ICC.Sample.Size) with 80% power and 95% significance. After the calculation, it was recommended 251 physicians, thus, a total of 253 physicians were required for 11 cases in our study, which means 23 physicians were visited for each case at least.[46]

As there are two forms of online consultation, including text consultation and telephone consultation, we decided to use both forms. Therefore, our USPs will need to complete a total of 506 visits to 253 physicians.

Patient and public involvement

No patient was involved.

Variables

Technical quality by USP assessment

The primary measurement in this study will be the quality of teleconsultation assessed by the USP visits and the online reviews, respectively. We adopted the framework developed by the Institute of Medicine (IOM), on quality which includes five domains: effectiveness, efficiency, equity, patient-centeredness, safety and timeliness.[48] We deem effectiveness and safety as the 'technical components' of quality. In this study, the USP will fill out a Quality Checklist after each teleconsultation to track the technical quality. Quality Checklist has explicit quality criteria for collecting data on patient history, physical examination, lab or imaging, diagnosis and treatment. The checklist's development follows an evidence-based and expert opinion augmented protocol that involves gathering, evaluating and selecting evidence and criteria related to clinician actions that are easily evaluated by an USP.[49] Technical quality will be primarily operationalised by the clinician's adherence (a score of 0–1, ie, the proportion of adhered items) to guideline-suggested items for consultation. Specifically, the consultation quality score will be calculated by the items completed by the physician divided by the suggested number of items as assessed by the USP. Second, we will also examine the outcome of diagnostic accuracy and treatment appropriateness. In general, teleconsultation has limitations in reaching definitive diagnoses and treatment plans. However, all our USP cases have a 'known' and 'correct' diagnosis and treatment plan. Therefore, we will evaluate the diagnosis and treatment plan on an ordinal scale of 0 (completely wrong diagnosis), 1 (partially correct diagnosis) and 2 (completely correct diagnosis). All our USP cases are also specifically designed to make teleconsultation possible to make diagnoses and treatment plans (see later sections on the USP measurement tool).

Patient-centeredness by USP assessment

The USP will take a survey using the Patient Perception of Patient-Centeredness Rating Scale-China version (PPPC) after each teleconsultation visit. PPPC covers four domains of patient-centred care: illness experiences, 'the whole person', shared decision-making and physician–patient relations.[50] The PPPC-China version has been validated in our prior study, and the validation results of that study are currently under journal review.[51] To facilitate our concordance analysis, we reversed the original scale of PPPC so that the total scores range from 1 to 4 in ascending order, with higher scores for better patient-centeredness.

Quality by teleconsultation online reviews

The USPs' visits will collect online review information from physicians' home pages. For online reviews by actual

Table 1 Primary outcome variables

Data	Domain	Specific metric	Type	Coding	Specific measurement
Quality evaluated by USP	Technical quality	% of recommended questions asked	Continuous	0–1	SP checklist
		% of recommended exams performed	Continuous	0–1	SP checklist
		Diagnosis quality	Ordinal	0: Incorrect, 1: Partially correct 2: Correct	SP checklist
		Treatment quality	Ordinal	0: Incorrect, 1: Partially correct 2: Correct	SP checklist
	Patient-centredness quality	Patient perception of patient-centredness	Continuous	0–1	PPPC
Quality from online reviews	Patient's review	Favourable rate	Continuous	0–1	Physician Webpage Information
			Ordinal	0 : Extremely low 1 : Low 2 : Medium 3 : High 4 : Extremely high	Physician Webpage Information

PPPC, Patient Perception of Patient-Centeredness; USP, unannounced standardised patient.

patients, we will calculate a quality score operationalised by a percentage of positive reviews, that is, the number of positive reviews divided by the total number of all reviews. For the three platforms, each patient's review was labelled 'good, moderate or poor rating' or the like. We thus define reviews labelled as 'good reviews' as positive reviews. To match the variable type of the USP-assessed quality, we will present the online review score by two forms of expressions: (1) a continuous score of 0–1 based on its original rate; and (2) an ordinal variable with five categories of extremely low, low, medium, high and extremely high with corresponding numeric weights of 0, 1, 2, 3 and 4. We summarise outcome variables in table 1.

Other quality variables

For quality evaluation following the IOM framework,[48] we will also measure the time and efficiency using USP. Equity will be analysed but not measured at each USP visit level. We will report three quality components in the study not used for concordance analysis with the online review score. Online review scores reported more on the overall assessment of patients' perception of technical quality and patient-centeredness but much less on a specific aspect of efficiency, timeliness and equity. The duration of teleconsultation will operationalise time for every USP visit, including the wait time (the time lapse between the USP submitting the teleconsultation appointment request and the actual start time of the consultation), and the total consultation time (the time from the beginning of consultation to the end of consultation). Teleconsultation in China often takes the form of text-based consultation rather than real-time virtual meetings. In text-based consultation, physicians may have several intermittent interactions by text with the patients rather than a single continuous and real-time consultation as in in-person care. The total consultation time includes the time-lapse between each text communication. For efficiency, we will record all expenses related to each USP visit, including the consultation fee and other expenses, such as medical expenses. Other indicators, data collection methods and collection stages are summarised in table 2.

Other reviews variables

Through our pilot, we have identified the following variables as presented on the online platform relevant to our analysis of online reviews :

1. The physicians' offline organisational affiliation, including hospital department and hospital level and grade.
2. The overall response speed of the physicians, the total number of patients served and the total number of teleconsultation visits.
3. Other summary metrics for physician performance on the teleconsultation platforms. For instance, each physician's personal site on Haodf.com displays metrics on 'Peer Recommendation' and 'Patient Recommendation'. Both are ordinal variables of physicians' Haodf.com homepage, indicating the overall extent

Table 2 Secondary outcome variables

Domain	Variable name	Type	Coding	Source
Timeliness	Wait time	Continuous	Minutes	SP checklist
	Consultation time	Continuous	Minutes	SP checklist
Efficiency	Registration cost	Continuous	RMB	SP checklist
	Medication cost	Continuous	RMB	SP checklist
	Total cost	Continuous	RMB	SP checklist

RMB, Renminbi.

of physician peers' and patient's recommendations for this physician, respectively. In general, the 'Peer Recommendation' is a physician peer's rating based on physician academic exchange and peer referrals. For 'Patient Recommendation', a Haodf.com manager assigns a score based on an overall analysis of the online patient reviews and the service data of the physician. However, the platforms hold those ranking methods proprietary and thus lack complete transparency in calculating those metrics. Haodf.com is the earliest and largest online doctor review website in China,[43 44] so it shows the review in more forms. We will also collect similar data from the other two platforms if they are available.

4. The online review metrics for in-person services. The teleconsultation platform presents online reviews in two sections, one for teleconsultation and the other for in-person offline services of the same physician. In this study, we focus on the validity of online reviews for teleconsultation, but we will also collect relevant online review metrics on in-person services as displayed on the physician's personal site, including the total number of reviews, the number of reviews with doctor's response, the number of reviews for visits within 2 years, the patient's satisfaction with the treatment results, and the patient's satisfaction with the physicians' attitude.
5. The number of 'heart-warming gifts' (a form of patient's tips to the physician) and monetary rewards received by physicians. The heart-warming gifts are indicated on the physician's personal site as digital flowers. Patients can opt to give those flowers to the physicians to appreciate their services. The flower has a monetary value (the patient can decide the exact amount) and can be cashed by the physician.

Those review variables are shown in table 3.

Demographic variables

The study will collect the following variables to describe the overall teleconsultation service: (1) social demographic profile information of the physicians, including titles, specialties, and gender; and (2) USP profile information (age, sex, education, etc).

Table 3 Other review variables

Domain	Variable name	Type	Source
Career	The physicians' hospital department	Unordered	Physicians' homepage
	The physicians' hospital grade	Ordinal	Physicians' homepage
Serve	The duration of the visit	Continuous	Physicians' homepage
	The speed of the physicians' response	Continuous	Physicians' homepage
	Total number of patients visited	Continuous	Physicians' homepage
	The number of teleconsultation visits	Continuous	Physicians' homepage
Recommended degree	Peer-recommended degree	Continuous	Physicians' homepage
	Patient-recommended degree	Continuous	Physicians' homepage
	Comprehensive recommendation degree	Continuous	Physicians' homepage
	Medication rationalisation evaluation	Continuous	Physicians' homepage
Offline service reviews	The total number of reviews	Continuous	Physicians' homepage
	The number of reviews with doctor's response	Continuous	Physicians' homepage
	The number of reviews of visits within 2 years	Continuous	Physicians' homepage
	The patient's satisfaction with the treatment results	Continuous	Physicians' homepage
	The patient's satisfaction with the physicians' attitude	Continuous	Physicians' homepage
Gifts	Heart-warming gifts	Continuous	Physicians' homepage
	Monetary reward	Continuous	Physicians' homepage

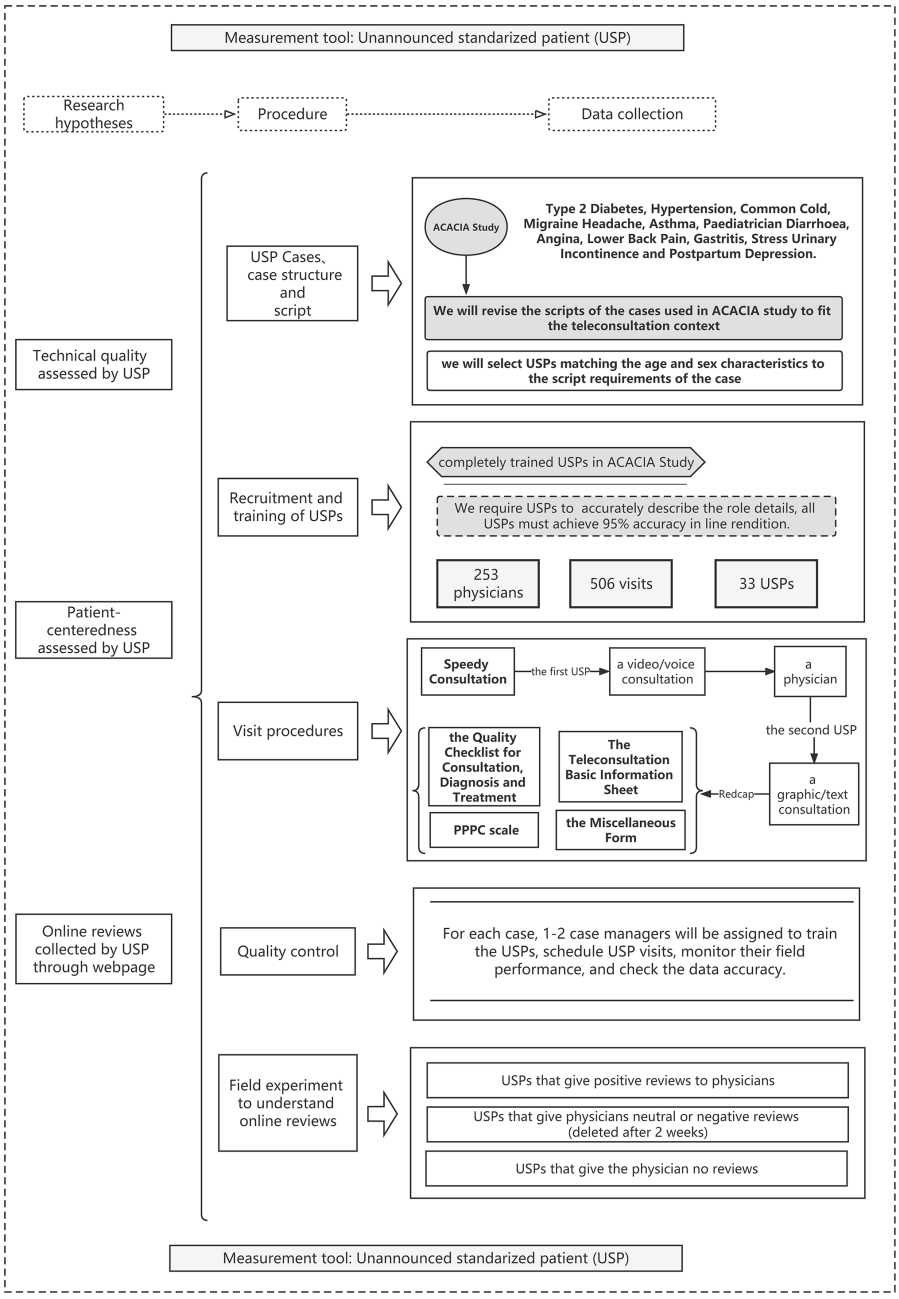

Figure 1 Field implementation procedure.

Measurement tool: USP

We will use USPs to visit online healthcare physicians to collect quality information on teleconsultation. Standardised patients are persons trained to portray a specific patient case in a standardised way,[37] thus controlling for all patient-level factors confounding the quality assessment. When SPs make unannounced visits to physicians, they further minimise Hawthorne effects.[34 35 52–54] USPs are particularly suitable for measuring online services as faking symptoms and signs and avoiding invasive procedures are easier online than the in-person implementation of SP.[53 55] We outlined our USP procedure in the following sections (figure 1).

USP cases

Each USP will play one or two of the 11 cases: type 2 diabetes, hypertension, common cold, migraine headache, asthma, paediatric diarrhoea, angina, lower back pain, gastritis, stress urinary incontinence and postpartum depression. The cases have all been used in our earlier studies to evaluate in-person primary care services and will be slightly adapted to suit online environments. We selected those tracer conditions for the following reasons: (1) they represent a diverse range of common primary care conditions[49 54]; (2) they are validated for face validity, content validity and criterion validity and field used in our prior ACACIA (Primary Health Care Duality Cohort) study (a large USP effort

to assess the quality of in-person primary healthcare in seven provinces of China[49 54 56]; (3) they involve symptoms and signs that can be feasibly and realistically presented during teleconsultation video/voice consultation; and (4) our team has experienced players of these cases who have been trained, validated and used in the ACACIA study.[49 54 56]

Case structure and script

Our USP cases have standardised structures (the script) that consist of case scenarios, patient social and demographic profiles, lines and quality assessment checklists. The USP must adhere to the lines in their conversation with physicians during the teleconsultation that may relate to medical history-taking, physical examination, laboratory and imaging, diagnosis, treatment or management. We will revise the scripts, particularly the lines of the cases used in the ACACIA study to fit the teleconsultation context.[49 54] We will select USPs matching the age and sex characteristics to the script requirements of the case.

Recruitment and training of USPs

As mentioned above, we will arrange two USP visits to the same physician to ensure a more reliable quality assessment. As a result, we will recruit at least 2–3 USPs for each case. Each physician receives one telephone consultation and one text consultation with different USPs. Under the ACACIA study, we have recruited, trained, validated and fielded USPs for all selected cases.[49 54] This study will retain USPs from the ACACIA study. Those USPs are familiar with the rendition of the signs, symptoms and case scripts and are experienced in completing various survey forms such as quality checklists. All USPs will receive refresher training to improve the accuracy of USP performance, mainly focusing on additional or revised scenarios, procedures and scripts as required by the online contexts, filling quality checklist and the procedures of using teleconsultation services on the selected platforms. All USPs must achieve 95% accuracy inline rendition to qualify for official fielding.

Visit procedures

All USP visits will take place between February 2023 and May 2023. Each selected physician will receive two USP visits, one for voice/video-based and the other for graphic/text consultations. Those are the two primary forms of teleconsultation in China. For video/voice consultation, the USP will follow the 'Speedy Match'—the triage process of the platform, to identify physicians to be visited. If an USP is assigned to a physician that the USP has visited before, they will ask the triage to select another physician. For any physician who have received a video/voice visit, a second USP (with an identical medical condition) will directly select them for a graphic/text consultation. We will schedule the visits to the same physician by the two USPs of identical medical conditions 30 days apart to reduce the possibility of physician suspicion of USP activity.

To minimise the risk of leaving a fake medical record in the teleconsultation platform, a case associate will register to the teleconsultation platform and seek care on behalf of the USP. The case association and the USP will work in tandem in interacting with the physician. The associate will inform the physician that he/she is a friend of the USP who is unfamiliar with the internet, so the association is seeking care on his/her behalf. During the phone or video conversation, the USP will be called on to join the associate in the communication. In this way, the physician will be asked to make a note in the teleconsultation system that the medical interaction with the associate is not for him/herself. Further, the associate will later call the teleconsultation platform to ask for removing the information from the system. We have confirmed with the teleconsultation platform that the clients have the right to remove their information.

The USP must complete four surveys by the end of each teleconsultation: (1) the Teleconsultation Basic Information Sheet (physician information, teleconsultation costs and time spent), (2) the PPPC scale,[50] (3) the Quality Checklist for Consultation, Diagnosis and Treatment and (4) the Miscellaneous Form to capture other important information such as lines during the USP–physician conversation not included in our script. As mentioned earlier, the checklist consists of benchmarks for patient history, physical examination, lab/imaging, diagnosis and treatment. The development process was systematic, relying on evidence and expert input to determine criteria related to clinician actions that USPs can conveniently evaluate.[49] The USP and the case associate must record the entire teleconsultation process with phone screenshots of the text communication and voice/video recordings. All data, including the recordings, will be entered and saved online in the Research Electronic Data Capture (REDCap) platform. Redcap is a data collection and sharing platform that ensure data security and also facilitates data tracking, auditing and access control.[57] The REDCap platform provides customisable tools for collecting, storing and sharing data with a high level of security.[58] In this study, all information collected during the USP visits will be directly entered and stored in the REDCap system using smartphones. Once uploaded and confirmed, the USPs will be instructed to destroy any local data. They will also receive training on proper procedures for destroying local data. Access to the data in REDCap will be strictly restricted, and only fully deidentified data will be presented to the analysts. All analyses will be performed at the aggregate level without exposing individual information. This approach is to ensure that the privacy and confidentiality of patients and healthcare providers are protected throughout the study.

Quality control

For each case, one to two case managers will be assigned to train the USPs, schedule USP visits, monitor their

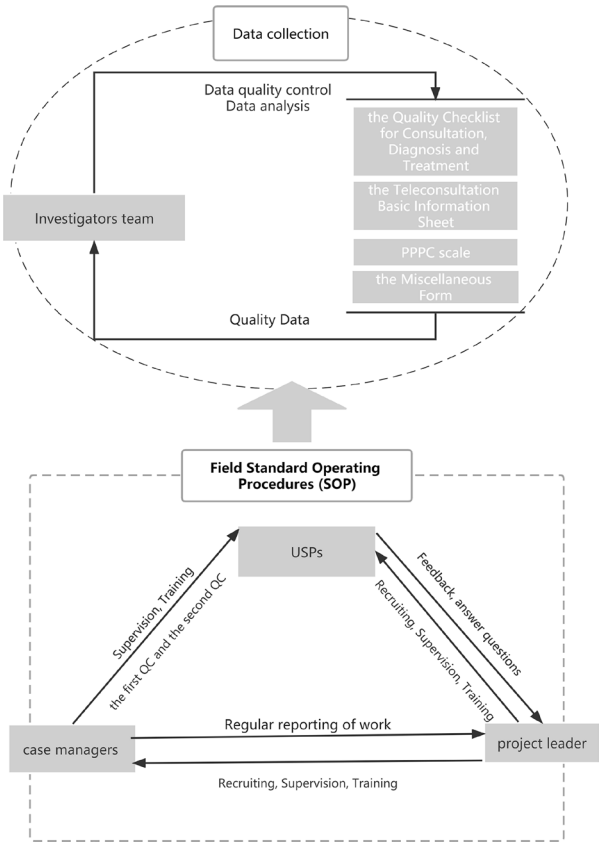

Figure 2 Personnel division. PPPC, Patient Perception of Patient-Centeredness; USP, unannounced standardised patient.

field performance and check the data accuracy. The case manager will log onto the REDCap system to check the voice recording or screenshots of the USP visit for the accuracy of the rendition of the USP for the case script and the information entered by the USP. Particular attention will be given to the accuracy of the completed quality checklist. Any discrepancies will be discussed between the case manager and the USPs. The unresolved dispute will be brought to the case developer for a resolution. The division of personnel for field implementation is shown in figure 2.

Field experiment to understand online reviews

The teleconsultation platform displays several summary metrics of the online review for a given physician on their physician-specific homepage, such as the 'Overall Number of Positive Reviews', 'Degree of Overall Recommendations', 'Degree of Physician Peer Recommendations' and 'Degree of Overall Patient Recommendation' on Haodf. com. We will carefully examine the website and smartphone applications to ascertain the exact definition of those metrics and the background agorism to generate them. If the website and applications lack transparency in explaining those summary metrics, we will call the help desk and ask for clarification. We will conduct a field experiment to understand those metrics if they provide insufficient clarification. In the experiment, three physicians from our research team will open an account on the teleconsultation platform. Nine USPs, divided into three groups, will schedule visits with the physician to leave positive, neutral/negative and no reviews, respectively. The three USPs in the no-review group did not review the doctor. In the experiment, we will examine the following: (1) how soon the review text will be displayed and made publicly available? (2) how and how soon do the individual reviews affect the aforementioned summary metrics? and (3) is there any follow-ups or feedback from the teleconsultation to the physician or patients regarding the reviews? The three physicians will also call the service desk to inquire about the mechanisms of generating his/her summary metrics. The USP will contact customer service after 2 weeks to request deletion of the review on the grounds of 'accidental error' or other similar excuses.

Statistical analysis

We will first conduct a descriptive analysis and present the data in summary tables. Continuous variables will be reported as numbers of observed and missing values, mean, SD, median and range. Categorical variables will be described as frequencies and percentages. We will also report information concerning the execution of the USP data collection process, such as fidelity of the USP rendition of their roles and the accuracy of Quality Checklist completion.

Our primary analysis will focus on the level of concordance, or agreement, between online reviews and USP-assessed quality. We will primarily use the ICC[45] to measure the concordance between the quality as assessed by online reviews and USPs. The value of ICC ranges from 0 to 1, with the greater value indicating a higher correlation of responses within a cluster. Specific to our study, the higher ICC suggests greater concordance between the quality assessed by online reviews and USPs. When the results of ICC are >0.4, we will regard the actual healthcare quality of physicians as consistent with the online reviews of patients.

Additionally, we will also use correlation analysis, Lin's Concordance Correlation Coefficient (CCC) and Kappa as secondary analysis. Correlation is only concerned with whether they have similar patterns of variation and not bias, while CCC measures both correlation and bias. When both variables are normally distributed, we will use the Pearson correlation coefficient or will use Spearman Rank Correlation Coefficient. Kappa will be used for two classification variables. For all statistical tests $p<0.05$ will be considered significant. R software V.4.2.2 was used.

DISCUSSION

This study will examine the concordance between the teleconsultation qualities assessed by online reviews and USPs (the gold standard of quality in this study). As far as we know, this is the first study comparing online reviews to a gold standard. Our research design and methodology

will allow testing of the hypotheses that online reviews are more in line with the quality component related to patient-centred care but not the technical component of quality. Our study may offer important insights for the growing number of teleconsultation consumers to make informed choices of physicians. The study may also guide policymakers and platform operators to identify appropriate approaches to evaluate online consultation quality appropriately.

The strengths of the study include (1) the use of the gold-standard quality measurement tool of USPs, which has been validated in our prior studies; (2) the assessment of both technical and patient-centred parts of quality; (3) unobjective or fake review of the samples for both the cases and physicians; (4) the sampling of physicians (such as using the 'Speedy Match' triage model) closely simulate real patient's experiences of seeking online care; and (5) our attempt with experiment to open the black box of the summary quality metrics on the teleconsultation platform to gain a deeper understanding of online reviews; and finally (7) our use of concordance rather than correlation analysis to evaluate the agreement of two measures for the same quality variable.

The study will have three foreseeable limitations. First, due to resource constraints, we will only be able to visit the three largest teleconsultation platforms, excluding the smaller platforms and teleconsultation services directly provided by hospital-operated online sites. However, we should note that the three platforms accounted for almost 80% of the total teleconsultation visits.[41] Second, we use only 11 cases, while online teleconsultation involves numerous more cases. However, those 11 common cases in primary healthcare provide good tracer conditions on the online quality of care for common primary care conditions. Third, at the individual physician level, the online reviews are generated by patients seeking various medical conditions within that physician's expertise, whereas a given physician in our study is evaluated only by a single USP case. Ideally, we should identify the review information for the same medical condition from each physician for the comparative analysis, which is however not practical. Lastly, as the PPPC-China Version is currently undergoing peer review for publication, we acknowledge that the study's use of this tool may be subject to change. In the event that the PPPC-China Version does not pass peer review, we will search for and utilise the next best available instrument for measuring patient-centred care.

ETHICS AND DISSEMINATION

This protocol received ethical approval from the institutional review board (IRB) of Southern Medical University, with a waiver of informed consent from each participating general practitioner (Approval #Southern Medical Audit (2022) No. 013). It is important to note that all analyses will be conducted at the aggregated level only, creating minimal risk to individual physicians. We have obtained the consent waiver as it will ensure the soundness of our USP methodology to prevent the self-selection of physicians into the study and the exposure of the true identity of the USPs.[34 35] Prior ethical analysis has indicated the justification of those waivers.[59] The study has been registered at the China Clinical Trials Registry (ChiCTR2200062975).

Author affiliations
[1]School of Health Management, Southern Medical University, Guangzhou, China
[2]School of Public Health, Southern Medical University, Guangzhou, Guangdong, China
[3]Dermatology Hospital, Southern Medical University, Guangzhou, Guangdong, China
[4]Southern Medical University Institute for Global Health (SIGHT), Dermatology Hospital of Southern Medical University, Guangzhou, China
[5]Acacia Lab for Implementation Science, School of Health Management and Dermatology Hospital, Southern Medical University, Guangzhou, China
[6]Department of Epidemiology and Medical Statistic, School of Public Health, Sun Yat-Sen University, Guangzhou, Guangdong, China
[7]Department of Epidemiology and Medical Statistics, School of Public Health, Guizhou Medical University, Guiyang, Guizhou, China
[8]School of Public Health, Central South University, Changsha, China
[9]Institute of Applied Health Research, University of Birmingham, Birmingham, UK
[10]Department of Social Medicine and Health Management, Lanzhou University, Lanzhou, Gansu Province, China
[11]Gillings School of Global Public Health, The University of North Carolina at Chapel Hill Gillings School of Global Public Health, Chapel Hill, North Carolina, USA
[12]Department of Epidemiology and Biostatistics, School of Public Health, Xi'an Jiaotong University, Xi'an, Shaanxi, China
[13]Center for World Health Organization Studies and Department of Health Management, School of Health Management, Southern Medical University, Guangzhou, China

Twitter Xiaoshan Chen @13066948007 and Dong Xu @romanxu

Acknowledgements Zhang Zhang received support from the CPC NICHD-NRSA Population Research Training grant (T32 HD007168).

Contributors DX conceived the project concept. YX developed the first protocol draft, along with WenjunH. DX, YX and WenjunH developed the sampling design. YX, WenjunH and HL wrote the section on samples and performed the sample size calculation. WG and YC provided original data of the previous studies for the sample size estimation and calculated some summary statistics. SL, WenpingH, LZ, YW and DZ worked on the SP field execution. ZZ reviewed the content and edited the manuscript and GC reviewed the statistical plan. JL, SL, JL, ChunpingL, QZ, XC, WangqingH, YW and ChangchangL reviewed and commented on the design and methods. All coauthors participated in the revision and approved this manuscript.

Funding This project is funded through the following competitive grants: National Natural Science Foundation of China (#71974211) The Swiss Agency for Development and Cooperation (#81067392).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting or dissemination plans of this research.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iDs
Wenjie Gong http://orcid.org/0000-0002-7943-4041
Zhang Zhang http://orcid.org/0000-0001-9344-6578
Dong Xu http://orcid.org/0000-0001-7438-632X

REFERENCES

1 National Health Commission Office of China. 2022. Notice of the National health Commission office on doing a good job in Internet diagnosis and consultation services in epidemic prevention and control
2 Zheng Y, Lin Y, Cui Y. Teledermatology in China: history, current status, and the next step. *J Investig Dermatol Symp Proc* 2018;19:S71–3.
3 Bollmeier SG, Stevenson E, Finnegan P, *et al*. Direct to consumer telemedicine: is healthcare from home best? *Mo Med* 2020;117:303–9.
4 Angaran DM. Telemedicine and telepharmacy: current status and future implications. *Am J Health Syst Pharm* 1999;56:1405–26.
5 Almathami HKY, Win KT, Vlahu-Gjorgievska E. Barriers and facilitators that influence telemedicine-based, real-time, online consultation at patients' homes: systematic literature review. *J Med Internet Res* 2020;22:e16407.
6 Wu H, Lu N. Online written consultation, telephone consultation and offline appointment: an examination of the channel effect in online health communities. *Int J Med Inform* 2017;107:107–19.
7 Wu H, Lu N. Service provision, pricing, and patient satisfaction in online health communities. *Int J Med Inform* 2018;110:77–89.
8 Cao B, Huang W, Chao N, *et al*. Patient activeness during online medical consultation in China: multilevel analysis. *J Med Internet Res* 2022;24:e35557.
9 Jiang S. Talk to your doctors online: an Internet-based intervention in China. *Health Communication* 2021;36:405–11.
10 Chen S, Guo X, Wu T, *et al*. Exploring the online doctor-patient interaction on patient satisfaction based on text mining and empirical analysis. *Information Processing & Management* 2020;57:102253.
11 China Internet Network Information Center (CNNIC). The 48th statistical report on the development of China's Internet. 2022.
12 China Central Television Caijing Channel. :"Half an Hour of Economy" Program: "Remote Diagnosis and Treatment" in the Context of the Epidemic. CCTV Program Official Website-CCTV-2-CCTV.com..
13 Li J, Zhang Y, Ma L, *et al*. The impact of the Internet on health consultation market concentration: an econometric analysis of secondary data. *J Med Internet Res* 2016;18:e276.
14 Hanauer DA, Zheng K, Singer DC, *et al*. Public awareness, perception, and use of online physician rating sites. *JAMA* 2014;311:734–5.
15 PatientEngagementHIT. Online reputation management tips for healthcare facilities. SRJ Digital, 2022.
16 Ivers N, Jamtvedt G, Flottorp S, *et al*. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012;2012:CD000259.
17 Huiqian H, Ruiqiu Z, Sun W, *et al*. Analysis and research on influencing factors of Internet medical patient satisfaction. *Packaging Engineering* 2021;42:204–11.
18 Liu Y, Lei F, Zhou J, *et al*. An analysis of the satisfaction of hospitals' online inquiry service. *China Sanitary Standards Management* 2020;11:24–6.
19 Liu J, Zhang W, Jiang X, *et al*. Data mining of the reviews from online private doctors. *Telemed J E Health* 2020;26:1157–66.
20 Peng L, Wang Y, Chen J. Consequences of gift giving in online health communities on physician service quality: empirical text mining study. *J Med Internet Res* 2020;22:e18569.
21 Nilsson M, Rasmark U, Nordgren H, *et al*. The physician at a distance: the use of videoconferencing in the treatment of patients with hypertension. *J Telemed Telecare* 2009;15:397–403.
22 Wu QL, Tang L. What satisfies parents of pediatric patients in China: a grounded theory building analysis of online physician reviews. *Health Commun* 2022;37:1329–36.
23 Champion A, Congiusta A, Yagnik A. Comparison of patient satisfaction measures between in-person and telemedicine postoperative appointments following third molar surgery. *Int J Oral Maxillofac Surg* 2021;50:830–4.
24 López A, Detz A, Ratanawongsa N, *et al*. What patients say about their doctors online: a qualitative content analysis. *J Gen Intern Med* 2012;27:685–92.
25 Brown-Connolly NE. Patient satisfaction with telemedical access to specialty services in rural California. *J Telemed Telecare* 2002;8 Suppl 2:7–10.
26 Popeski N, McKeen C, Khokhar B, *et al*. Perceived barriers to and facilitators of patient-to-provider e-mail in the management of diabetes care. *Can J Diabetes* 2015;39:478–83.
27 Daskivich TJ, Houman J, Fuller G, *et al*. Online physician ratings fail to predict actual performance on measures of quality, value, and peer review. *J Am Med Inform Assoc* 2018;25:401–7.
28 Ryan T, Specht J, Smith S, *et al*. Does the press ganey survey correlate to online health grades for a major academic otolaryngology department? *Otolaryngol Head Neck Surg* 2016;155:411–5.
29 Reimann S, Strech D. The representation of patient experience and satisfaction in physician rating sites. A criteria-based analysis of english- and german-language sites. *BMC Health Serv Res* 2010;10:332.
30 McGrath RJ, Priestley JL, Zhou Y, *et al*. The validity of online patient ratings of physicians: analysis of physician peer reviews and patient ratings. *Interact J Med Res* 2018;7:e8.
31 Woodward CA, McConvey GA, Neufeld V, *et al*. Measurement of physician performance by standardized patients. Refining techniques for undetected entry in physicians' offices. *Med Care* 1985;23:1019–27.
32 Beaulieu M-D, Rivard M, Hudon E, *et al*. Using standardized patients to measure professional performance of physicians. *Int J Qual Health Care* 2003;15:251–9.
33 Das J, Holla A, Das V, *et al*. In urban and rural India, a standardized patient study showed low levels of provider training and huge quality gaps. *Health Aff (Millwood)* 2012;31:2774–84.
34 Leonard KL, Masatu MC. Using the Hawthorne effect to examine the gap between a doctor's best possible practice and actual performance. *Journal of Development Economics* 2010;93:226–34.
35 King JJC, Das J, Kwan A, *et al*. How to do (or not to do) … using the standardized patient method to measure clinical quality of care in LMIC health facilities. *Health Policy Plan* 2019;34:625–34.
36 Li J, Liu M, Li X, *et al*. Developing embedded taxonomy and mining patients' interests from web-based physician reviews: mixed-methods approach. *J Med Internet Res* 2018;20:e254.
37 Epstein RM, Hundert EM. Defining and assessing professional competence. *JAMA* 2002;287:226–35.
38 Onishi J, Gupta S, Peters DH. Comparative analysis of exit interviews and direct clinical observations in pediatric ambulatory care services in Afghanistan. *Int J Qual Health Care* 2011;23:76–82.
39 Peabody JW, Luck J, Glassman P, *et al*. Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. *JAMA* 2000;283:1715–22.
40 von Elm E, Altman DG, Egger M, *et al*. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Int J Surg* 2014;12:1495–9.
41 Trustdata. Trustdata: national ecological traffic reached 67 million in the first quarter of 2020. 2022.
42 Haodf.com. About Us-Haodf.com. 2022.
43 Patel RN, Antonarakis GS. Factors influencing the adoption and implementation of teledentistry in the UK, with a focus on orthodontics. *Community Dent Oral Epidemiol* 2013;41:424–31.
44 Ibrahim A-M, Kathleen G, Reeva L. *Online medical consultation: a review of literature and practice*. 2015.
45 McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychological Methods* 1996;1:30–46.
46 Zou GY. Sample size formulas for estimating intraclass correlation coefficients with precision and assurance. *Stat Med* 2012;31:3972–81.
47 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159.
48 Hanefeld J, Powell-Jackson T, Balabanova D. Understanding and measuring quality of care: dealing with complexity. *Bull World Health Organ* 2017;95:368–74.
49 Xu DR, Hu M, He W, *et al*. Assessing the quality of primary healthcare in seven Chinese provinces with unannounced standardised patients: protocol of a cross-sectional survey. *BMJ Open* 2019;9:e023997.
50 Stewart M, Brown JB, Donner A, *et al*. The impact of patient-centered care on outcomes. *J Fam Pract* 2000;49:796–804.
51 Stewart M, Meredith L, Ryan BL, *et al*. *The patient perception of patient centeredness questionaire (PPPC) # 04-1*. Centre for Studies in Family Medicine - Western University, 2022.
52 Daniels B, Dolinger A, Bedoya G, *et al*. Use of standardised patients to assess quality of healthcare in Nairobi, Kenya: a pilot, cross-sectional study with international comparisons. *BMJ Glob Health* 2017;2:e000333.

53 Luck J, Peabody JW. Using standardised patients to measure physicians' practice: validation study using audio recordings. *BMJ* 2002;325:679.
54 Xu DR, Cai Y, Wang X, *et al*. Improving data surveillance resilience beyond COVID-19: experiences of primary health care quality cohort in China (Acacia) using unannounced standardized patients. *Am J Public Health* 2022;112:913–22.
55 Sylvia S, Xue H, Zhou C, *et al*. Tuberculosis detection and the challenges of integrated care in rural China: a cross-sectional standardized patient study. *PLoS Med* 2017;14:e1002405.
56 Kwan A. Implementing quality of care measures: lessons from a standardized patient study in seven provinces of China. *Am J Public Health* 2022;112:818–20.
57 Garcia KKS, Abrahão AA. Research development using redcap software. *Healthc Inform Res* 2021;27:341–9.
58 Harris PA, Taylor R, Thielke R, *et al*. Research electronic data capture (redcap) -- a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377–81.
59 Rhodes KV, Miller FG. Simulated patient studies: an ethical analysis. *Milbank Q* 2012;90:706–24.