Supplementary Materials

A Appendix Tables

Table A1: Example of case diagnosis and management standards

	Correct Diagnosis	Correct Management
Unstable Angina	One of the following diagnoses: (1) Unstable angina (2) Acute coronary syndromes (3) coronary heart disease	Either (1) or (2), and no antibiotics (1) One of the correct medicines (2) Referral Correct medicines are: (1) Antiplatelet (2) Nitrate Esters (3) Statins (4) β-blockers (5) Angiotensin-converting enzyme (ACE)
Child viral diarrhea	One of the following diagnoses: (1) Viral diarrhea (2) Virus	ORS
Pulmonary Tuberculosis	One of the following diagnoses: (1) Pulmonary tuberculosis (2) Suspected tuberculosis (3) Tuberculosis	One of four management: (1) Chest X-ray + referral (2) Sputum test (3) Referral (4) HRZE

Note: This table provides the definition of correct case diagnosis and management from the online SP project. Standards used in the other SP projects are similar.

Table A2: Quality Outcomes across three disease cases

Panel A: Angina

	Angina				
	Rural	**County**	**Migrant**	**CHC**	**Online**
	Mean (SD) / n(%)				
	95% CI				
Process quality					
Proportion of recommended questions and examinations	0.245 (0.127)	0.254 (0.142)	0.196 (0.096)	0.305 (0.160)	0.278 (0.124)
	[0.23 to 0.26]	[0.19 to 0.31]	[0.17 to 0.22]	[0.27 to 0.34]	[0.25 to 0.31]
Median	0.22	0.22	0.17	0.40	0.27
IQR	(0.17 to 0.33)	(0.17 to 0.39)	(0.11 to 0.28)	(0.20 to 0.40)	(0.18 to 0.36)
Diagnosis quality					
Correct diagnosis	58 (19.7%)	4 (19.0%)	10 (19.6%)	33 (42.3%)	29 (46.0%)
	[0.15 to 0.24]	[0.02 to 0.36]	[0.09 to 0.31]	[0.31 to 0.53]	[0.34 to 0.58]
Case management					
Referred patients	179 (60.9%)	0 (0%)	38 (74.5%)	20 (25.6%)	61 (96.8%)
	[0.55 to 0.66]	[0.00 to 0.00]	[0.62 to 0.87]	[0.16 to 0.35]	[0.92 to 1.01]
Medication					
Medications prescribed	96 (32.7%)	8 (38.1%)	13 (25.5%)	17 (21.8%)	8 (12.7%)
	[0.27 to 0.38]	[0.17 to 0.59]	[0.13 to 0.38]	[0.13 to 0.31]	[0.04 to 0.21]
Correct medications, if any	18 (18.8%)	2 (25.0%)	0 (0%)	13 (76.5%)	4 (50.0%)
	[0.11 to 0.27]	[-0.07 to 0.57]	[0.00 to 0.00]	[0.56 to 0.97]	[0.13 to 0.87]
Correct medications	18 (6.1%)	2 (9.5%)	0 (0%)	13 (16.7%)	4 (6.3%)
	[0.03 to 0.09]	[-0.03 to 0.22]	[0.00 to 0.00]	[0.08 to 0.25]	[0.00 to 0.12]
Unnecessary antibiotics prescribed, if any	18 (18.8%)	0 (0%)	3 (23.1%)		0 (0%)
	[0.11 to 0.27]	[0.00 to 0.00]	[-0.01 to 0.47]		[0.00 to 0.00]
Unnecessary antibiotics prescribed	18 (6.1%)	0 (0%)	3 (5.9%)		0 (0%)
	[0.03 to 0.09]	[0.00 to 0.00]	[-0.01 to 0.12]		[0.00 to 0.00]
Correct case management	190 (64.6%)	2 (9.5%)	38 (74.5%)	29 (37.2%)	61 (96.8%)
	[0.59 to 0.70]	[-0.03 to 0.22]	[0.62 to 0.87]	[0.26 to 0.48]	[0.92 to 1.01]
Observations	294	21	51	78	63

Note: Data are sample mean (SD) for numerical variables or n (%) for binary variables. Process quality for rural and urban providers are measured as the proportion of recommended questions and examinations. Online process quality is measured as the proportion of recommended questions.

Table A2: Quality outcomes across three disease cases

Panel B: Viral Diarrhea

	Diarrhea			
	Rural	**County**	**Migrant**	**Online**
	Mean (SD) / n(%)			
	95% CI			
Process quality				
Proportion of recommended questions and examinations	0.194 (0.114)	0.250 (0.160)	0.173 (0.089)	0.362 (0.085)
	[0.18 to 0.21]	[0.18 to 0.32]	[0.15 to 0.20]	[0.33 to 0.39]
Median	0.18	0.24	0.18	0.35
IQR	(0.12 to 0.24)	(0.12 to 0.41)	(0.12 to 0.24)	(0.35 to 0.44)
Diagnosis quality				
Correct diagnosis	14 (5.3%)	2 (10.0%)	1 (2.0%)	16 (59.3%)
	[0.03 to 0.08]	[-0.03 to 0.23]	[-0.02 to 0.06]	[0.40 to 0.78]
Case management				
Referred patients	27 (10.2%)	0 (0%)	6 (11.8%)	21 (77.8%)
	[0.07 to 0.14]	[0.00 to 0.00]	[0.03 to 0.21]	[0.62 to 0.94]
Medication				
Medications prescribed	191 (72.1%)	14 (70.0%)	34 (66.7%)	26 (96.3%)
	[0.67 to 0.77]	[0.49 to 0.91]	[0.54 to 0.80]	[0.89 to 1.04]
Correct medications, if any	11 (5.8%)	0 (0%)	2 (5.9%)	13 (50.0%)
	[0.02 to 0.09]	[0.00 to 0.00]	[-0.02 to 0.14]	[0.30 to 0.70]
Correct medications	11 (4.2%)	0 (0%)	2 (3.9%)	13 (48.1%)
	[0.02 to 0.07]	[0.00 to 0.00]	[-0.01 to 0.09]	[0.29 to 0.67]
Unnecessary antibiotics prescribed, if any	74 (38.7%)	2 (14.3%)	18 (52.9%)	1 (3.8%)
	[0.32 to 0.46]	[-0.05 to 0.33]	[0.36 to 0.70]	[-0.04 to 0.11]
Unnecessary antibiotics prescribed	74 (27.9%)	2 (10.0%)	18 (35.3%)	1 (3.7%)
	[0.23 to 0.33]	[-0.03 to 0.23]	[0.22 to 0.49]	[-0.04 to 0.11]
Correct case management	51 (19.2%)	5 (25.0%)	4 (7.8%)	14 (51.9%)
	[0.14 to 0.24]	[0.06 to 0.44]	[0.00 to 0.15]	[0.33 to 0.71]
Observations	265	20	51	27

Note: Data are sample mean (SD) for numerical variables or n (%) for binary variables. Process quality for rural and urban providers are measured as the proportion of recommended questions and examinations. Online process quality is measured as the proportion of recommended questions.

Table A2: Quality outcomes across three disease cases

Panel C: Tuberculosis

	TB			
	Rural	**County**	**Migrant**	**Online**
	Mean (SD) / n(%)			
	95% CI			
Process quality				
Proportion of recommended questions and examinations	0.175 (0.087)	0.147 (0.070)	0.171 (0.094)	0.181 (0.086)
	[0.17 to 0.19]	[0.12 to 0.18]	[0.15 to 0.20]	[0.16 to 0.20]
Median	0.16	0.16	0.19	0.17
IQR	(0.13 to 0.23)	(0.10 to 0.19)	(0.10 to 0.26)	(0.14 to 0.24)
Diagnosis quality				
Correct diagnosis	38 (13.2%)	6 (28.6%)	2 (3.8%)	19 (27.1%)
	[0.09 to 0.17]	[0.09 to 0.48]	[-0.01 to 0.09]	[0.17 to 0.38]
Case management				
Referred patients	57 (19.8%)	1 (4.8%)	16 (30.2%)	62 (88.6%)
	[0.15 to 0.24]	[-0.05 to 0.14]	[0.18 to 0.43]	[0.81 to 0.96]
Medication				
Medications prescribed	206 (71.5%)	6 (28.6%)	34 (64.2%)	25 (35.7%)
	[0.66 to 0.77]	[0.09 to 0.48]	[0.51 to 0.77]	[0.24 to 0.47]
Correct medications, if any	0 (0%)	0 (0%)	0 (0%)	0 (0%)
	[0.00 to 0.00]	[0.00 to 0.00]	[0.00 to 0.00]	[0.00 to 0.00]
Correct medications	0 (0%)	0 (0%)	0 (0%)	0 (0%)
	[0.00 to 0.00]	[0.00 to 0.00]	[0.00 to 0.00]	[0.00 to 0.00]
Unnecessary antibiotics prescribed, if any	170 (82.5%)	4 (66.7%)	29 (85.3%)	17 (68.0%)
	[0.77 to 0.88]	[0.25 to 1.08]	[0.73 to 0.97]	[0.49 to 0.87]
Unnecessary antibiotics prescribed	170 (59.0%)	4 (19.0%)	29 (54.7%)	17 (24.3%)
	[0.53 to 0.65]	[0.02 to 0.36]	[0.41 to 0.68]	[0.14 to 0.34]
Correct case management	100 (34.7%)	19 (90.5%)	19 (35.8%)	62 (88.6%)
	[0.29 to 0.40]	[0.78 to 1.03]	[0.23 to 0.49]	[0.81 to 0.96]
Observations	288	21	53	70

Note: Data are sample mean (SD) for numerical variables or n (%) for binary variables. Process quality for rural and urban providers are measured as the proportion of recommended questions and examinations. Online process quality is measured as the proportion of recommended questions.

Table A3: Summary of physician characteristics

	Full Sample	Rural	County	Migrant	CHC	Online
Male physician, 1=yes	980 (77.5%)	727 (85.9%)	42 (67.7%)	83 (70.3%)	42 (53.9%)	86 (53.8%)
Physician Age						
Physician age <50 years	908 (72.0%)	588 (69.5%)	47 (75.8%)	80 (69.6%)	54 (69.2%)	139 (86.9%)
Physician Certificate						
Physician with Practicing Physician Certificate, 1=yes	673 (51.7%)	388 (45.8%)		61 (39.4%)	66 (84.6%)	158 (98.8%)
With Assistant Practicing Physician Certificate, 1=yes	190 (14.6%)	171 (20.2%)	NA	15 (9.7%)	4 (5.1%)	0 (0%)
With Rural Physician Certificate, 1=yes	266 (20.4%)	233 (27.5%)		33 (21.3%)	0 (0%)	0 (0%)
Other or missing, 1=yes	173 (13.3%)	55 (6.5%)	62 (100.0%)	46 (29.7%)	8 (10.3%)	2 (1.3%)
Observations	1302	847	62	155	78	160

Note: Data are sample n (%). Other title/certificate means that the physician title was recorded as other title/certificate in the survey. Physician title in county hospital is missing.

Table A4: Examples of checklists

Panel A: Angina

Questions
Pain type
Pain severity
Triggers
Pain radiation
Pain alleviation
Pain duration
First timing of symptom onset
Recent timing of symptom onset
Shortness of breath
Nausea/vomiting
Sweating
Frequency
Risk factors
Living habits
Family history
Exams
ECG (EKG)
Coronary angiography
Blood Pressure
Pulse
Auscultation (front & back)
Temperature

Note: This table provides the checklists from the online SP project. Checklists used in the other SP projects are similar.

Table A4: Examples of checklists

Panel B: Viral Diarrhea

Questions
Nature of stool
Stool frequency
Blood or mucus in the stool
Stool volume
Urine (including color)
Time of last urination
Child age
Fever
Stool smell
There are insects in the stool
Abdominal pain
Vomiting
Children's diet
Children are weak (mental state)
Hygienic habits especially hand washing
Is there a similar patient (environment) around
Are there tears when crying
Exams
Stool inspection
Mucosal humidity check
Body temperature
Body weight
Abdominal palpation

Note: This table provides the checklists from the online SP project. Checklists used in the other SP projects are similar.

Table A4: Examples of checklists

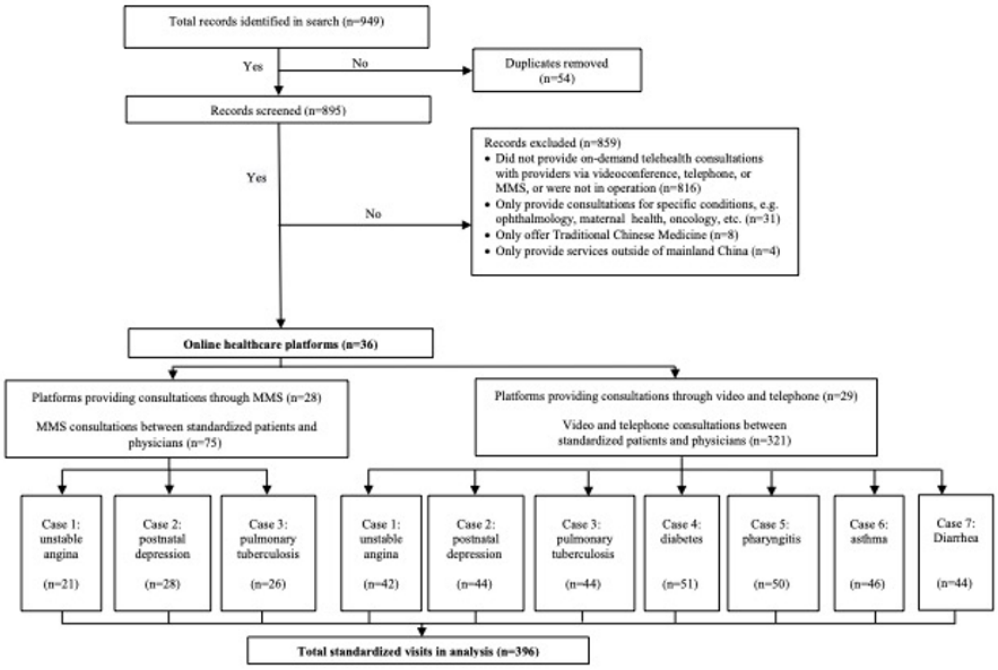
Panel C: Tuberculosis

Questions
Cough duration
(Produced) Sputum
Past tuberculosis
Family tuberculosis
Night sweats
Lost weight
Fever type
Blood in sputum
Fever (duration)
Chest pain
(Loss of) Appetite
Breathing difficulty
Wheezing
Smoking
Diabetes
High blood pressure or hypertension
HIV/AIDS
Alcohol
Age
Family symptoms
Cough throughout the day
Weakness
Exams
Chest radiograph
Sputum smear test (Sputum AFB)
Mantoux Tuberculin Skin Test (TST)
Sputum culture test
Weight
Temperature
Auscultation

Note: This table provides the checklists from the online SP project. Checklists used in the other SP projects are similar.

Figure A1: Flowchart of the sampling process and SP visits from the Three-Province SP project

Figure A2: Flowchart of the sampling process and SP visits from the Online SP project



B Supplementary Materials on Methods

B.1 Sample Selection

Rural Clinics

The data pertaining to rural clinics and county hospitals, henceforth referred to as the "rural sample" and the "county sample" respectively, were collated through the "3-Province SP project". This project was executed in the provinces of Shaanxi, Sichuan, and Anhui in 2015. The rural sample is comprised of data gathered from 847 SP visits to 209 village clinics (VCs) and 209 township health centers (THCs) situated in rural areas. The county sample encapsulates data gathered from 62 SP visits to 21 county hospitals, typically located in the county seat. These data sets were gathered from prefectures predominantly housing rural populations, encompassing approximately 12.23 million residents.[7,38]

The sample was selected to be representative of health systems in three city-level prefectures, one in each province. The prefectures were selected for having a predominantly rural population and in consultation with local authorities. We used the following procedure to sample health facilities: First, across the three prefectures, we randomly sampled 21 of 24 rural counties. Next, 10 township health centers (THCs) were randomly sampled within each county, excluding those located in urban areas of the county (often at the county seats). Since one county only had 9 rural townships, our sample contained 209 (out of the total 311) THCs in the 21 sampled counties. Finally, we randomly selected one village clinic (VC) associated with each sampled THC, which yielded a total of 209 VCs. This village was selected from an administrative list of villages in each township. In the case that the village originally selected did not have a village clinic, we replaced that village with a randomly selected backup village from the same township. Out of the 209 originally-sampled villages, 22% had no village clinic and was replaced with the backup. Only one SP was sent to each village clinic (which are much smaller in size) to minimize the risk that SPs were identified as fake patients.

Migrant Clinics

Data pertaining to migrant community clinics, subsequently referred to as the "migrant sample", were derived from the "Migrant SP project". This initiative took place in the Shaanxi province in 2015 and covered 53 clinics across 25 urban migrant communities, collectively housing approximately 475,000 migrants, in an undisclosed prefecture.[27,39] Our study incorporates information gathered from 155 SP visits to these clinics.

We randomly selected 25 communities out of a list of migrant communities suggested by locals in two of nine total urban districts in the selected prefecture. To generate a list of candidate clinics, we used a random sampling procedure to identify 10 sample households

in each of the 25 migrant communities, and asked them to name the clinics or hospitals they most often visited when they were sick. Based on the clinic list, we selected two or three clinics with the highest frequency of responses in each community as our sample. Notably, clinics only served Chinese medicine and specialty clinics were excluded from sample selection. In total, 53 clinics were included in this project.

Urban Community Health Centers

The data pertaining to urban community health centers (CHCs), referred to hereafter as the "urban sample", was amassed through the "Xi'an CHCs SP project." This project took place in the urban region of Xi'an, the capital city of Shaanxi province, in 2017. The sample comprises 78 SP visits to 63 CHCs, which cater to an estimated urban populace of 6.5 million residents.

This project covers all urban districts of Xi'an, the capital of Shaanxi Province. We invited all the 65 CHCs in the area to participate. Two of them were then excluded from the study because they only provided public health, not basic medical health. Thus, the study consisted of 63 CHCs (53 public CHCs vs. 10 private CHCs).

Online Platforms

Data related to online platforms, henceforth referred to as the "online sample," were gathered through the "Online SP project," implemented in 2019. This project encompassed 177 SP visits to 36 direct-to-consumer telemedicine platforms offering real-time consultations.[28]

The Online SP project aimed to enumerate all direct-to-consumer (DTC) telemedicine platforms operating in mainland China. These platforms provided synchronous or asynchronous consultations via videoconference or SMS/MMS text. To develop a comprehensive list of such platforms, we conducted a systematic survey of telemedicine platforms in China from August to October 2019. The survey was completed in three steps.

First, we conducted a systematic web search for telemedicine platforms. We used twenty-one keywords (see Table S1) to search on Baidu, Google, and six application stores (Apple, Google Play, Tencent, Baidu, Huawei, Xiaomi). The search was implemented by six research assistants divided into three groups; each group responsible for seven keywords. The two research assistant in each group conducted the search and listed all records found from each keyword combination independently. We then compared the two lists and removed the duplicates. This process identified 949 records in total.

Second, the research team excluded 54 duplicates found across groups and 859 records that failed to meet the study inclusion criteria (See Fig. 1 in manuscript). The excluded records consisted of 635 platforms that did not provide on-demand telehealth consultations to patients, 118 platforms without a webpage or an application, 63 platforms not in operation,

and 31 platforms focusing on specific conditions beyond research scope (ophthalmology, cosmetics, oncology, etc.), eight platforms that only provided Traditional Chinese Medicine consultations, and four platforms that only provided services outside of mainland China. In total, 36 platforms were included in the study.

Finally, data on platform characteristics and structure were extracted from the websites of each platform. If the information was unavailable on the platform, research assistants were instructed to locate this information from other sources. This was done independently by two research assistants for each platform.

B.2 Data Collection

The 3-Provinces SP project comprised three waves (Fig A1). Wave one involved an initial facility and doctor survey for village and township-level providers (excluding county hospitals) in June 2015, with physician consent for SP visits obtained. Wave two involved unannounced SP visits in August 2015, with SPs presenting one of three disease cases: unstable angina, pulmonary tuberculosis (TB), or child with viral gastroenteritis (presented by a parent). Wave three was a follow-up survey in September 2015, inquiring whether providers detected SPs.

The Migrant SP project mirrored the first study's procedures and had three waves in July and October 2015. The first wave of data collection consisted of an initial short facility survey that was conducted in early July 2015. Physician consent for SP visits was obtained at this time. Actual SP visits started ten days after the initial survey in late July 2015. Finally, in October 2015, we conducted a follow-up survey with those physicians visited by SPs. In this survey, we asked physicians whether they had detected any SPs and administered a detail facility and physician survey. This survey collected information including physician characteristics.

The CHCs SP project conducted SP visits between August 2017 and July to August 2018. SPs presented unstable angina or asthma. Written consent was obtained from physicians and CHC directors. Face-to-face surveys were completed approximately three months before SP visits.

The Online SP project had two phases (Fig A2). The first surveyed direct-to-consumer (DTC) telemedicine platforms, with platform design and characteristics collected from webpages and public documentation. The second phase involved SP visits to providers, presenting seven disease cases (including unstable angina, diarrhea, and TB) via videoconferencing, telephone, text, or chat-apps (MMS or SMS). Physicians were unaware that SPs were not real patients.

B.3 Standardized Patients

Description of SP Cases

The research team trained the standardized patients to present one of the three disease cases during their visits. In each case, the SP began the interaction with a physician using an opening statement, including the chief complaint. For TB case, the opening statement is: "Doctor, I have a cough that is not improving and a fever." For diarrhea case, the opening statement is "Doctor, my child has diarrhea", and for angina case with "Doctor, I have chest pain recently". The SP then answers any questions asked by the providers and receives any (non- invasive) exams. Upon appropriate questioning by the provider, the SP reveals symptoms consistent with a classic case of three disease case. For TB, these symptoms include cough durations for 2-3 weeks, fever with night sweats, and loss of appetite and weight, which are consistent with a classic case of presumed TB and, according to national guidelines, should not be prescribed appropriate antibiotics until confirmation. For diarrhea, appropriate questioning would reveal symptoms of a two-year old child consistent with a viral infection (including watery diarrhea without blood or mucus, no fever or vomiting and little change in behavior). The chief complaint of the angina case, chest-pain, is less likely consistent with a disease requiring antibiotics but provides insight into antibiotic prescription for less ambiguous cases.

SP script development

The SP script used in this study were adapted from earlier validation studies in India1 and previous studies in China.[2–4] Previous studies have demonstrated several advantages. First, participation in the study had minimal to no risk for the SPs or health care providers. Second, the likelihood of SP detection among visited providers was low, confirming that SPs were considered real by health providers who were visited. Third, SP deviations from the case scripts were rare. Additionally, because the SPs paid fees requested, there is no loss to provider income from participation in the study.

We adapted the SP scripts thoroughly to the Chinese context with the help of an advisory panel consisting of medical experts in China. Adaptation of the scripts included: (1) ensuring that the clinical presentation of the SP would be interpreted clearly given telemedicine consultation context; (2) ensuring that SP responses were prepared for any likely questioning by providers; (3) developing SP interaction protocols given clinical settings in China; (4) developing detailed background histories for SPs to minimize the threat of SP detection as fake patients. Additionally, for the 3-provinces SP project, because local dialects varied across the three regions, scripts with alternative phrasing to match the local dialect (where appropriate) were developed from a version in standard Mandarin. These were small alterations to phrasing or vocabulary and were chosen to convey the exact information as the standard script. SP script adaptations took place as an iterative process including field pretesting with pre-trained SPs.

The SP Scripts (including background and dialog) are available for download here: https://ssylvia.io/sp-cases.

SP recruitment and characteristics

For all in-person SP visits, we recruited standardized patients from local area, in an effort to minimize the risk of being detected as fake patients and to maintain a high quality of training. This meant that SPs were similar to patients typically seen by the clinics in terms of language (dialect), mannerisms, and dress.

In the 3-Provinces SP project and Migrant SP project, a total of 63 individuals (42 males and 21 females) from an initial group of 21 were hired and trained as SPs from three provinces (21 from each province). The SPs, although recruited specifically to fit the three disease cases in terms of health and physical characteristics, differed in age, gender height and weight. Our recruitment standard for SPs was that they be average weight and height, and healthy with no obvious signs of illness or other conditions that could prejudice diagnoses. The average age of recruited SPs for TB was 37, the youngest was 28 and the oldest 43; the average age of recruited SPs for diarrhea was 29, the youngest was 25 and the oldest 38; the average age of recruited SPs for angina was 41, the youngest was 34 and the oldest 50.

In the CHCs SP project, 12 SPs were selected, including eight female SPs and two male SPs in 2017, and seven female SPs and one male SPs in 2018. It should be emphasized that 6 of these SPs (5 female SPs and a male SP) in 2017 also participated in the survey of 2018.

In the Online SP project, SPs were recruited from Sun Yat-Sen University in China through a multi-stage process. We received 107 applications that meet the requirements of SPs. After screening, interview, and training, 52 undergraduate and graduate students (43 females and ten males) were selected out of 107 to participate in telemedicine consultation as SPs. All SPs were between 20 and 28 years of age. For cases predicting an older patient, SPs presented cases as the child of the patient, a common practice in China particularly for online consultations.

SP training

The aim of SP training was to ensure that they (a) correctly presented the cases in a consistent way, (b) correctly recalled the interaction with providers, (c) avoided detection and (d) for the 3-provinces SP project, Migrant SP project and CHCs SP project, SPs would be able to complete interactions safely without being exposed to invasive tests or procedures.

These aims were achieved through classroom training in case presentation and testing of recall, as well as mock interviews and dry runs that were supervised in the field. The training started with a focus on the cases and the development of scripts and proceeded to

memorization and appropriate role-playing. SPs were taught to internalize completely the characters and the details of their background stories. Mock interviews were conducted with trainers as well as providers working with the research team. These mock interviews initially asked only potentially clinically relevant questions, but in later rounds, additional questions about family or neighborhood were added to ensure that SPs could answer appropriately and convincingly. Mock interviews also simulated "threats" of invasive procedures.

In the 3-Provinces SP project and Migrant SP project, all SPs underwent a centralized intensive two-week training. In the final week of training, SPs conducted supervised dry runs in clinics nearby the training site. Since it is common in China that more than one patient is present in the examination room at the same time, dry runs were conducted in which a supervisor was present and thus could watch the interaction and offer corrections later. Dry runs were also conducted during a two-day practice round in local areas to ensure that the SPs were accustomed to local conditions before starting data collection. Enumerators accompanying SPs also attended the full training in order to familiarize themselves with the survey process and the SP visit.

In the CHCs SP project, all SPs participated in a three-day training conducted by a team consisting of professors, medical experts, and investigators. The training included the following: 1) the details of case scenarios and scripts were explained to the SPs; 2) the recordings of interactions between SPs and physicians obtained from our pilot study were presented to the SPs; 3) role-playing and one-on-one training were used to help SPs understand and memorize the scripts and portray the cases; 4) some principles and scripts for responding to the physician's questions were provided to help the SPs avoid examinations; and 5) an assessment of the SPs' performance.

In the Online SP project, all SPs participated in a centralized intensive three-day training. SPs also received training as enumerators and trained to accompany other SPs in consultations with physicians. During interactions these enumerators would sit out of sight of the camera and take notes on the interaction following a standardized form. In the afternoon of the third training day, all trained SPs participated in an evaluation by the research team members. The evaluation included seven aspects: 1) the proficiency of script, 2) familiarity of communication rules with physicians, 3) the procedure of making appointments with physicians at platforms, 4) the performance during mock interviews; 5) understanding of exit questionnaire, 6) attitudes, and 7) collaboration ability. Only SPs who met the requirements of all seven aspects are selected for participation.

B.3.1 Assignment of SPs to providers

In the 3-Provinces SP project and Migrant SP project, each survey team (comprising of three SPs for different cases and two enumerators) were randomly assigned to two counties within each province. Within each county, teams were assigned a random half of sample townships. SPs were never assigned to their home township where they would risk being recognized.

Each survey team first visited their assigned townships (both township health centers and village clinics in each township) in their first assigned county and then traveled to their second assigned county. County hospitals were visited by the second team assigned to each county.

Within each facility, SPs visited the doctor following the normal procedures for any walk-in patient. Given a choice of which doctor to visit, SPs randomly chose a doctor following a pre-determined randomization protocol. In county hospitals, where patients can choose doctors by specialty, SPs visited generalists. Our results therefore approach the care a walk-in patient would receive at each of the sampled facilities.

In the CHCs SP project, four SPs were randomly assigned to and independently visited each CHC (two SPs portraying unstable angina and two SPs portraying asthma). The SPs could not choose the physicians, and they must be seen by whoever would have seen them had they been a common patient once they entered the practical setting. The SPs wore a concealed recording device to record the interactions between physicians and the SPs.

In the Online SP project, all SPs were randomly assigned to telemedicine platforms. For video and phone telemedicine consultations, two SPs present each disease case were assigned to each platform offering these consultation modes. One SP presenting each of three disease cases were assigned to each platform offering text-based consultations.

When booking consultations, the SP followed the routine procedures for any telemedicine patient. First, the SP will choose the specific department for the presented case and then choose the first physician based on the platforms' default listing. SPs made appointments with physicians meeting two criteria: 1) they provided the type of consultation assigned to SPs; and 2) they had an available appointment slot in the following week. If the first listed physician did not meet these criteria, they would move to the next. If the SP was assigned to use video or phone consultation, but the physician only provided text and graph service, the SP would skip this physician and make an appointment with the next available physician who provides a video or phone consultation. The SP would choose a telephone call if the video consultation was not available. At the end of the appointment procedure, the SP paid the consultation fee and provided required information about the patient, e.g., age, gender.

Upon each consultation with the physician, SPs made an opening statement of the disease cases' primary symptom(s). Then SPs responded to all questions by physicians following a predetermined script. An enumerator paired with the SP recorded the full consultation, including diagnosis and case management using a structured questionnaire. After the consultation, the SP answered additional questions and cross-checked the interaction records with the pared enumerator.

Ethical Approval

In the 3-provinces SP project and Migrant SP project, approvals from the institutional review boards of Stanford University, USA (protocol number: 25904) and Sichuan University, China (protocol number: K2015025) were obtained. Informed consent was obtained verbally from all providers participating in the study. To prevent influence on the study, both IRBs approved a procedure whereby providers consented to SP visits "at some point in the next six months." Consent from village and township providers was obtained as part of the facility survey approximately 5 weeks before SP visits. Consent for county providers was obtained through communications with providers. It was approved to record the interactions between physicians and SPs using a concealed recording device. All individuals who participated as SPs were trained to protect themselves from any invasive tests or procedures.

In the CHCs SP project, the ethics approval was obtained by the Ethics Committee of Xi'an Jiaotong University Health Science Center (approval number: 2015-406). It was approved to record the interactions between physicians and SPs using a concealed recording device. Written consents were obtained from physician and the director of each CHC along with a face-to-face survey approximately three months prior to SP visits.

In the Online SP project, the requirement of prior informed consent was waived based on minimal risk and that no individually identifiable information on physicians was recorded.

Evaluation of the process quality, diagnosis quality and case management

We assessed process quality by grading recorded interactions against 'clinical checklists' of recommended questions to be asked of the patient and any physical exams to be performed. These recommended condition-specific checklists were based on national and international guidelines 5–10 that were tested in a pilot study in our rural Chinese context 2-3 as well as checklists used in the Das and colleagues in a study of quality of care in rural India.1 The same approach was used for diagnosis and case management standards. Table S2 presents checklist items (recommended questions and examinations), and standards for diagnosis and management for three disease cases.

Drug Identification

To get full information about the drugs and cost for each interaction, SPs purchased any medications prescribed and paid providers their usual fee. After each visit, enumerators packed all the drugs for each case in an individual bag (for the Online SP project, recorded all the drugs) and labeled all the information related to drugs in it.

In the case of drugs prescribed to be taken intravenously, the protocol was designed to allow SPs to avoid being administered the IV while also recording the drugs to be administered.

If an IV was prescribed, SPs paid for the IV and took the written prescription but left before being administered the IV (indicating that they would soon return). If the written prescription was illegible, SPs asked pharmacy staff the contents of the IV. If there were no pharmacy staff, SPs asked providers the contents directly. IVs were prescribed in 11% of village clinic interactions, and 28% in township health centers.

All labelled medicines prescribed by the pharmacies were digitized and stored and then coded by enumerators with the assistance of consulting doctors and pharmacologists. Blinded from any provider details, they identified and categorized medicines as antibiotics, broad-spectrum antibiotics, anti-TB drugs, or other medicines. Loose or unlabeled pills were dispensed in 10% of THC interactions and 37% of village clinic interactions. For loose or unlabeled pills, a group of 3 consulting pharmacologists worked together to identify drugs. Unless all pharmacologists agreed, we coded these drugs as unidentified. Less than 5% of these drugs could not be identified (we did not perform chemical testing).

For the Online SP project, all medicines prescribed were digitized and stored and then coded by enumerators with the assistance of consulting doctors and pharmacologists. Blinded from any provider details, they identified and categorized medicines as correct medicines, harmful medicines to each disease, and Chinese medicines.