BMJ Open

Assessing Quality of Primary Health Care in 7 Chinese Provinces with Unannounced Standardized Patients: Protocol of a Cross-sectional Survey

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-023997
Article Type:	Protocol
Date Submitted by the Author:	22-May-2018
Keywords:	standardized patients, unannounced standardized patients, quality of primary health care, patient-centered care

Complete List of Authors:

Xu, Dong; Sun Yat-sen University, Sun Yat-sen Global Health Institute (SGHI),School of Public Health and Institute of State Governance
Hu, Mengyao; University of Michigan, Survey Research Center, Institute for Social Research
He, Wenjun; Sun Yat-sen University, Department of Biostatistics and Epidemiology, School of Public Health
Liao, Jing; Sun Yat-sen University, Sun Yat-sen Global Health Institute (SGHI), School of Public Health and Institute of State Governance
Cai, Yiyuan; Guizhou Medical University, School of Public Health; Sun Yat-Sen University, School of Public Health
Sylvia, Sean ; University of North Carolina at Chapel Hill, Department of Health Policy and Management, Gillings School of Global Public Health
Hanson, Kara; London School of Hygiene & Tropical Medicine, Department of Global Health and Development, Faculty of Public Health and Policy
Chen, Yao-Long; Lanzhou University, Evidence-Based Medicine Center, School of Basic Medical Sciences
Pan, Jay; Sichuan University, West China School of Public Health
Zhou, Zhongliang; Xi'an Jiaotong University, School of Public Policy and Administration
Zhang, Nan; Inner Mongolia Medical University, Department of Health Management, School of Health Management
Tang, Chengxiang; Guangzhou University, School of Public Administration
Wang, Xioahui; Lanzhou University, Evidence Based Medicine Center, School of Basic Medical Sciences
Rozelle, Scott; Stanford University, Freeman Spogli Institute for International Studies
He, Hua; Tulane University, Department of Epidemiology, School of Public Health and Tropical Medicine
Wang, Hong; Bill & Melinda Gates Foundation, Health Economics, Financing & Systems
Chan, Gary; University of Washington, Department of Biostatistics
Melipillán, Edmundo; University of Michigan, Survey Research Center, Institute for Social Research
Zhou, Wei; Central South University, Xiangya school of public health,
Gong, Wenjie; Central South University, School of Public Health; University of Rochester Medical Center, Department of Psychi

Assessing Quality of Primary Health Care in 7 Chinese Provinces with Unannounced Standardized Patients: Protocol of a Cross-sectional Survey

Dong Roman Xu,[1] Mengyao Hu,[2] Wenjun He,[3] Jing Liao,[1] Yiyuan Cai,[4,3,1] Sean Sylvia,[5] Kara Hanson,[6] Yaolong Chen,[7] Jay Pan,[8] Zhongliang Zhou,[9] Nan Zhan,[10] Chengxiang Tang,[11] Xiaohui Wang,[7] Scott Rozelle,[12] Hua He,[13] Hong Wang,[14] Gary Chan,[15] Edmundo Roberto Melipillán,[2] Wei Zhou,[16] Wenjie Gong[17*]

[1] Sun Yat-sen Global Health Institute (SGHI), School of Public Health and Institute of State Governance, Sun Yat-sen University, No.74 Zhongshan 2nd Road, Guangzhou, P.R. China, 510080

[2] Survey Research Center, Institute for Social Research, University of Michigan, 426 Thompson St, Ann Arbor, MI 48104, USA

[3] Department of Biostatistics and Epidilogoy, School of Public Health, Sun Yat-sen University, No.74 Zhongshan 2nd Road, Guangzhou, P.R. China, 510080

[4] School of Public Health, Guizhou Medical University, Guizhou, China, 550025

[5] Department of Health Policy and Management, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, CB#7411, Chapel Hill, NC 27599, USA

[6] Department of Global Health and Development, Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine, WC1H 9SH, London, United Kingdom

[7] Evidence Based Medicine Center, School of Basic Medical Sciences, Lanzhou University, No. 199 Donggang West Rd, Lanzhou City, Gansu Province, 730000 China

[8] West China School of Public Health, Sichuan University, No. 17, Ren Min Nan Road, Chengdu, China

[9] School of Public Policy and Administration, Xi'an Jiaotong University, No. 28 Xianning West Road, Xi'an, Shaanxi, 710049, China.

[10] Department of Health Management, School of Health Management, Inner Mongolia Medical University, Jinshan Development Zone, Hohhot, Inner Mongolia, P.R. China, 010110

[11] School of Public Administration, Guangzhou University, Guangzhou, Guangdong, 510320, China.

[12] Freeman Spogli Institute for International Studies, Stanford University, Encina Hall East, E407, Stanford, CA 94305-6055, USA

[13] Department of Epidemiology, School of Public Health and Tropical Medicine, Tulane University, New Orleans, Louisiana, USA

[14] Health Economics, Financing & Systems, Bill & Melinda Gates Foundation, PO Box 23350, Seattle WA, 98102, USA

[15] Department of Biostatistics, University of Washington, Seattle, Washington, USA

[16] Hospital Administration Institute, Xinagya Hospital, Central South University, No. 87 Xiangya Road, Changsha, Hunan, China

[17] Xiangya school of public health, Central South University, Changsha, Hunan, China

Corresponding author: Wenjie Gong, gongwenjie@csu.edu.cn, Xiangya school of public health, Central South University, Xiangya school of public health, Central South University, China*

Abstract

Introduction: Primary health care (PHC) serves as the cornerstone for the attainment of universal health coverage (UHC). Efforts to promote UHC should focus not only on the expansion of access but also on healthcare quality. However, robust quality evidence has remained scared in China. Common quality assessment methods such as chart abstraction, patient rating, and clinical vignette use indirect information that may not represent real practice. This study instead will send standardized patients (SP or healthy person trained to consistently simulate the medical history, physical symptoms, and emotional characteristics of a real patient) unannounced to PHC providers to collect quality information.

Methods and Analysis: 1981 SP-clinician visits will be made to a random sample of PHC providers across 7 provinces in China. SP cases will be developed for 10 tracer conditions in PHC. Each case will include a standard script for the SP to use and a quality checklist that the SP will complete after the clinical visit to indicate diagnostic and treatment activities performed by the clinician. The patient-centeredness will be assessed by Patient Perception of Patient-centeredness (PPPC) rating scale by the SP. The SP cases and the checklist will be developed through a standard protocol and will be validated for validity and reliability before its full use. The usual descriptive analysis will be performed for the survey results such as a tabulation of quality scores across geographies and provider types. Several hypotheses will also be tested including the effect of facility ownership on PHC quality.

Ethics and dissemination: The study has been reviewed and approved by the Institutional Review Board of the School of Public Health of Sun Yat-sen University (#SYSU 2017-011). The

results will be actively disseminated through print and social media, and the SP tools will be made available for other researchers.

Keywords: standardized patients; unannounced standardized patients; quality of primary health care; patient-centered care

Strengths:

- We will assess the quality of care with a random sample of primary health care providers in 7 provinces in China.
- We will use unannounced standardized patient (USP) - the "gold standard" of quality assessment.
- Both technical quality and patient-centeredness will be assessed.

Limitations:

- USPs are not suitable for certain health conditions.
- 7 provinces are not randomly selected, although we intend them to represent different health development conditions of China's provinces.

Assessing Quality of Primary Health Care in 7 Chinese Provinces with Unannounced Standardized Patients: Protocol of a Cross-sectional Survey

Background

In 2015, all 191 UN member states adopted the Sustainable Development Goals (SDGs), aiming to achieve universal health coverage (UHC) – the access to high-quality health care services without incurring financial hardship – by 2030.[1] As previous literature emphasized, efforts to promote UHC should focus not only on the expansion of access but also on healthcare quality.[2] Healthcare quality is defined philosophically by the WHO as "responsiveness" of the health system,[3] or as the instrumental goals on structure, process and outcome in the Donabedian Framework,[4] or as the six comprehensive aims (effectiveness, efficiency, equity, patient-centeredness, safety and timeliness) put forth by the Institute of Medicine (IOM).[5] In this study, we take the IOM definition of the quality.

Primary health care (PHC) serves as the cornerstone for the attainment of UHC.[6] China's new round of health reform since 2009 has invested heavily in strengthening PHC. There have been some efforts to assess the quality of PHC in China: patients were interviewed with a Primary Care Assessment Tool (PCAT) questionnaire in Guangdong, Shanghai, and Hongkong;[7-9] comprehensiveness of the service provision was used as a proxy for quality through clinician interviewing;[10] PHC clinicians' adherence to clinical guideline was assessed with a self-report questionnaire.[11] However, assessment of the quality of PHC has largely remained scanty in China, and the assessment tools are indirect and prone to bias.[12] A number of studies have found quality of PHC to be low in other low and middle income countries (LMICs)[6 13-18], where robust evidence remains scarce.[19] Commonly-used methods of measuring technical quality of care

include chart abstraction, patient rating of care, and using a clinical vignette to test clinician knowledge. Those methods use indirect information that may not represent real practice. This study instead will use unannounced standardized patients (USP) to measure the quality of real practice. The Standardized Patient (SP) is a healthy person (or occasionally a real patient) trained to consistently simulate the medical history, physical symptoms, and emotional characteristics of a real patient. The SP, particularly when their visit is unannounced, has several reported advantages: (1) reliability in measurement and cross-provider comparison because the same patient is presented to all providers, (2) elimination of the Hawthorne effect (i.e., that the study itself may change doctors' behavior) due to the nature of disguised and unannounced visit by SPs,[20-22] and (3) reduced recall bias.[23 24]

Despite these advantages, the application of SP in China has concentrated mainly in medical education.[25] An ongoing systematic review identified four papers only on the use of SP for quality assessment in China,[14 26-28], and 44 in other LMICs. Those projects, often based on a small convenience sample, tended to target a limited number of conditions (approximately 70% on family planning services, childhood infectious diseases, sexually transmitted infections, and respiratory tract infection). In this study, we intend to assess the quality of PHC with a probability sample of PHC visits in seven Chinese provinces, using USP for 10 commonly seen conditions in the PHC setting.

Methods

Survey Design

The purpose of the sample design is to create a representative sample of China's primary health care (PHC) providers so that healthcare quality can be assessed based on USP visits to those providers.

Survey Population/Frame

We would consider creating nationally representative probability sample, but at this stage, we have selected seven provinces to "represent" China due to feasibility considerations. These provinces represent five levels of average life expectancies across China's provinces (Figure 1). Those provinces have the similar life expectancy to five countries ranging from low income to high income.[29] We intend to create a probability sample that represents primary health care in these seven provinces. For the survey population, we intend to include (1) licensed physician and licensed assistant physician at community/township health centers/stations and urban health stations, (2) certified village doctors (a terminology in China that refers to village clinicians who have village-level practice privilege even without a medical license) and village sanitarians (referring to un-certified village doctors who are supposed to work under the supervision of the village doctors) at village clinics; and (3) clinicians with a license notation for general practice, internal medicine, obstetrics/gynecology, and pediatrics at the level 1 and level 2 hospitals. We exclude level III hospitals, which provide more specialized care, and specialty hospitals. The clinician meeting those criteria will constitute the "sampling frame".

Sampling Procedures

The sample will be selected using a multi-stage, clustered sample design covering all eligible clinicians of the seven provinces (Figure 2). In the first stage, stratification will be based on the provinces. Due to the high number of visits in the seven capital cities, we will sample each capital city with certainty. Each province is thus divided into two strata consisting of the provincial capital city and other prefecture-level municipalities, leading to 14 strata in total. We will use proportionate allocation (in terms of the number of eligible clinicians) of sample size for each stratum. For each stratum, five rural townships or urban sub-districts (the primary sampling

unit/PSU) will be selected using probability proportional to size (PPS). In the second stage, for each PSU, PHC facilities as afore-defined (Secondary Sampling Unit/SSU) will be selected using PPS systematic sampling. Neighboring village clinics will be grouped as an SSU. The number of SSUs for each stratum will vary depending on the size of the stratum – e.g., more SSUs will be selected in strata with more PHC clinicians. In the final stage, a fixed number of USP visits will be made to each selected facility or the group of facilities in the case of village clinics. The exact number of visits will be determined once we obtain and examine our sampling frame. If multiple clinicians are available in that facility at the time of a particular USP visit (PHC visits in China do not require appointments), the field coordinator will randomly select a clinician by drawing lots onsite.

Sample Size Calculation

Sample size was calculated for the primary purpose of the standard descriptive survey analysis of this survey. The sample size (power) calculation for the other related hypothesis related studies will be described in separate study protocols. The primary statistic of interest in this survey is a latent variable measuring clinician's quality, constructed using the 2-parameter logistic item response theory (IRT) model.[30 31] The model was based on a list of quality checklist items measuring whether doctors asked recommended questions and whether they performed recommended exams (see section on *Scoring Method* below). Survey sample size was calculated based on the desired level of relative precision (coefficient of variation, CV), an estimate for the population element variance for the variable of interest (s^2) from previous study and design effect ($deff$). In this study, our desired level of relative precision (CV) is 0.08. s^2 was estimated to be 4.54, based on Sylvia et al's work on the USP-assessed quality of PHC in three Chinese provinces.[14 27] Design effect is the variance inflation due to cluster sampling. It was calculated

based on intra-class correlation (ICC) (describing the level of homogeneity of the units in a cluster) and cluster sample size: $deff = 1 + \delta(n - 1)$, where δ is the intra-class correlation (ICC) and n is the average size of the cluster. The ICC of 0.0486 was also estimated from Sylvia et al's work, which was 0.0486. Our estimated average cluster size is 27 clinician-SP encounters per PSU. Accordingly, we calculated the total sample size required to be 1981 clinician-SP encounters. The steps of calculating sample size can be found in **Error! Reference source not found.**.

USP Case Development

The development process of a USP case is based on our extensive literature review,[20 32] as well as our own USP experiences in Shaanxi province of China.[14 27] We are concurrently developing smartphone-based virtual standardized patients (VPs) (details described elsewhere). The two projects will share almost identical case scenarios and quality criteria.

Case Selection

Our purpose is to select ten health problems as tracer conditions for PHC in China. Ideally our selected cases should (1) be highly prevalent in PHC settings, (2) carry challenging features in different aspects of PHC (e.g., some cases focus on curative care while others on prevention, disease management, culturally-sensitive care,[33] or misuse of low value tests[34-36]), (3) not involve invasive and painful procedures, (4) not require physical signs that cannot be simulated (e.g., jaundice can be simulated with make-up, but heart murmurs cannot.[23]). We created a list of the top 30 conditions commonly seen in PHC in China, combining the results of two national surveys on PHC.[12] A panel of physicians, public health and health system researchers then applied the principles above and selected a dozen PHC problems for the USP development (Table 1). Ten final conditions will be selected from this list.

Development Team

We have created an overall development team and 10 case-specific development teams. Each team includes case-specific specialists, general practitioners, public health and health system researchers (**Error! Reference source not found.**). A third overall panel consisting of primary care providers at the village, township and community levels will review all cases for contextual appropriateness in the primary care settings. In developing the case, we will follow several principles: (1) limiting case scenarios to those that require definitive clinician action on the first visit to minimize potential "first-visit bias",[37] (2) focusing on the presentation of symptoms for which evidence is well-established for its diagnosis and management, (3) deriving some content of the cases from the actual case history of relevant patient files in real practice.[23]

Case Description

The case description describes the relevant clinical roles and psycho-social biographies of the SP.[38] We used a structured description of the cases as follows:

1. Social and demographical profile: (1) Socioeconomic information: name, gender, age, ethnicity, education, occupation, family structure (e.g., Married and have two children but live alone), dress style (e.g., dressed in jeans, work boots and a well-worn but neat sweater), health insurance or other social program participation; (2) personality that may influence interaction with the clinician (e.g., non-proactive and introverted); (3) lifestyle relevant to health (e.g., smoke one pack of cigarette since age 18, like fried pork but also eat much fruit, exercise regularly, watch TV series a lot in spare time, play mahjong with friends, visit children every week)
2. Medical history: (1) disease information: severity of the condition (e.g., mild or severe depression), duration of the condition (the first onset? Previously

diagnosed/existing (how long)?), comorbidity (any other physical and/or psychological problems?), (2) reason for seeking care for this specific visit (e.g., was feeling down for 2 months but depression worsened last week), (3) treatment/management already or currently received (e.g., a diabetic "patient" took metoprolol for hypertension but does not monitor his glucose / watch his diet/weight).

3. Physical examination: Symptoms the SP will (and will not) portray (e.g., reduced appetite, but not showing agitation), and medical signs the SP has or does not have (e.g., heart murmur).
4. Laboratory and imaging: The laboratory and imaging that a clinician may prescribe for the SP. The laboratory and imaging results of the SP may be generated from those of real typical patients.
5. Diagnosis: The correct diagnosis that the clinician should make based on the information presented by the SP.
6. Treatment and management: the decision of the clinician on what medications, procedures, advice, or referral will be given at the end of the consultation.

Script

Corresponding to the six components of the afore-mentioned case description, we will develop a detailed script for the SP to use in their PHC visit with the clinician. The script ideally should cover all possible questions a clinician may ask as well as the answers during the clinical interaction. Panels of clinicians will be consulted to collect relevant questions that will guide the development of the script. The script will continue to add new questions asked by the clinicians on the SP-clinician interaction. The script will have five sections: (1) an opening – spontaneous information given to the clinician at the start (e.g., Doctor, I have been feeling headache for two

days), (2) the information given only on request, (3) the information for the SP to volunteer even if not asked, (4) language to insist on a diagnosis if not given, and (5) an end.[14 20 39]

Quality Checklist

The checklist consists of explicit quality criteria for history, physical examination, laboratory/imaging, diagnosis and treatment.[14 32] Based on our comprehensive review of 14 literature and the evidence-based clinical guideline development methodology,[40] we have established the principle and a standard protocol for the checklist development. In principle, our process will be (1) evidence-based and augmented by expert opinion,[41] (2) following a systematic procedure to gather, evaluate and select evidence and criteria, (3) selecting criteria related to clinician actions that the SP can easily evaluate,[42] (4) keeping the number of the checklist items under 30 to include high-priority criteria only so that the SP can reliably recall clinician behaviour[42-44]. The details of our checklist development protocol will be described in a separate paper, and key messages are summarized in **Error! Reference source not found.**.

Selecting and Training SPs

We will advertise on social media to recruit SPs. The candidate must be in stable health without confounding symptoms; should match the real patients in age, sex, and physical features; are willing to allow the examinations appropriate to their condition; have the intellectual maturity to present the behavior of the actual patient and complete the checklist.[23 45 46] We may consider recruiting real patients with stable conditions to portray the cases not subject to simulation.[23] The training of the SP will aim at portraying the signs, symptoms, and presentations, completing the checklist, and minimizing detection by the provider.[20] The week-long training will have three stages: classroom instruction, a dress rehearsal, and two field tests.[23]

[46 47] A standardized training manual will be developed to guide the training and appraisal of the SPs.

Fielding SPs

A disguise plan will be developed for each case to minimize physician detection of the SP status (e.g., convincing excuse for seeking care where they do not usually reside). In the pilot (instrument validation) phase, consent will be sought for audio recording (see below); in these cases, fieldwork will start only 3-4 weeks after consent is obtained. We will provide each SP with a calamity letter, explaining the project in case of their identity being exposed.

Variables

Outcome Variables

We will collect a range of quality information and other related explanatory variables. The IOM quality framework (effective, safe, patient-centered, timely, efficient, and equitable) will be used for quality evaluation (Table 2). The **effectiveness** (avoiding underuse and misuse) and **safety** (avoiding harm), the traditional technical quality, will be evaluated through the yes/no checklist discussed above (**Error! Reference source not found.**). **Patient-centeredness** (respectful of and responsive to individual preferences) will be assessed by the 9-item Patient Perception of Patient-centeredness (PPPC) rating scale.[48-50] Using a 4-point Likert scale, PPPC evaluates three dimensions of patient-centeredness: exploring the disease and illness experiences, understanding the whole person, finding common ground.[48] Following a method developed by Pongsupap et al.,[5] we will embed patient-centered standardized questions into the script to elicit clinician response for the PPPC rating. Prior studies have demonstrated the validity of SPs rating clinician communications.[51 52] **Timeliness** will be assessed by analyzing opening hours, waiting time, consultation time, and clinician politeness and friendliness.[5] **Efficiency** (avoiding waste)

will be measured by costs of care of the SP-clinician encounter. **Equity of care** (no variance in quality because of personal characteristics) will be assessed through a sperate but related study in a randomized cross-over trial.

Scoring Method

Technical quality will be reflected by a continuous score ranging from 0-1. We will evaluate further whether to classify checklist items in four categories (essential, important, indicated, and non-contributory) with corresponding numeric weights (3, 2, 1, and 0).[53] Two scoring methods will be used: 1) the simple scoring will use the formula of items performed ÷ total number of items on the checklist for the process scores, whereas 2) the complex method will use an algorism based on item-response-theory (IRT).[30] Using the IRT model approach, we can obtain a latent performance score for each doctor, which has been corrected for measurement error. An ordinal variable will be used for diagnosis and management plans (Table 2). Patient-centeredness will follow the scoring methods of PPPC (possible range of score from 1-4).[50]

Other Variables

We will collect additional information on the predictors, confounders, and effect modifiers to the outcomes in the planned hypothesis testing of the related studies to this survey. The information will include qualification of the clinician and facility information (environment, amenity, size, location, ownership type, and so forth).

Analytical Methods

Survey Descriptive Analysis

Usual descriptive analysis of survey data will be performed we will present characteristics of the providers in tables as well as maps with geospatial analytical tools; results of overall quality and sub-domains will be tabulated in tables and figures across administrative

regions and provider types. Exploratory analyses will also be conducted to identify determinants of quality.

USP Validation

USP validation will be based on a convenience sample of clinicians not included in our final survey sample in the project training and pilot phase. Those SP-clinician interactions in the pilot will be audio recorded and transcribed. The **Validity** is the extent to which an instrument measures what it is supposed to measure. The face validity of the SP assessment depends on (1) SP remaining undetected (detection ratio reported to be 5%-10%[54]), (2) authentically and consistently portraying the clinical features, and (3) accurately completing the checklist.[55] We will send the participating clinician in the pilot a "detection form" to report degrees of their suspicion of any SP visit.[45] The authenticity of the SP presentation will be evaluated by checking the transcribed recording whether a key piece of information was divulged by the SP when appropriately prompted, not divulged when prompted, or volunteered when not prompted. The criterion validity will be assessed through the agreement of the SP-completed checklist against that by a clinician based on the transcript of the visit (i.e., the clinician rating as the "gold standard").[56-59] Checklist items depending on visual observation will be excluded. **Reliability** examines the level of consistency of the repeated measurements. The inter-rater reliability of two SPs on the same condition and context will be assessed with two SPs completing the checklist for the same recorded transcript. Test-retest reliability will be analyzed by the concordance of assessment results of the same SP to score his own recorded encounter weeks later).[57] The agreement will be analyzed with Lin's concordance correlation coefficient (r_c)[60]. r_c indicates how closely pairs of observation fell on a 45° line (the perfect concordance line)

through the origin in addition to their correlation.[60-62] Bland-Altman plot will be used to visualize the concordance.[63 64]

Hypothesis Testing

Several hypothesis-driven analyses will also be conducted. Separate study protocols will be developed to provide detail on the background, theoretical framework, and analytical methods. Among them, we will, in particular, assess whether private providers provide inferior quality of PHC to the public providers. Propensity scores matching will be used as the primary analytical method. A logistic regression model will be used to estimate the propensity score of each SP-clinician visit: including all available variables that are believed to be related to the quality outcome and/or the provider type.[65] The SP visits to the private providers will then be matched to the public ones based on the logit of their propensity scores. After the optimal balance is achieved, quality scores will be compared between the private and public providers. McNemar's test will be used to calculate the statistical significance. The R program's MatchIt package will be used for the statistical analysis.[66].

Ethical Consideration

The study has received ethical approval from the institutional review board (IRB) of Sun Yat-sen University School of Public Health with a waiver of informed consent from each participating clinician. USP studies do not necessarily require the consent if they meet certain conditions.[67 68] Our waiver is granted as (1) our study serves important public good while requiring informed consent may lead to considerable selection bias and greater risk for the detection of the SP; (2) the study does not intend to entrap or reveal identities of any institution or individual and all analyses will be conducted at the broader health system level (after data cleaning all individual identifiers will be destroyed); (3) no audio-visuals will be recorded

during the SP-clinician encounter (however, in the pilot stage, we will seek informed consent from the participating clinicians as we will use a disguised recording for the validation purposes).

Patient and public involvement

We selected the conditions for the USP partly based on the survey results of the common conditions in the context of primary health care as reported by the patient. The USP cases will also be reviewed by a panel that includes patients. The results of the studies will be widely distributed in scientific reports as well as social media to benefit policy-makers, clinicians, and patients.

Discussion

In this study, we will develop, validate and implement methods of assessing the quality of PHC using USPs. Compared to existing studies using USPs,[32] this proposed study has several distinctive features. First, we will establish a large probability random sample so that representative estimates of PHC quality can be achieved in the seven provinces in China. Second, unlike previous studies,[14 27] we do not only include village clinics, township health centers, community health centers but also county hospitals and other level I and level II hospitals in the study. The latter were not officially designated as PHC facilities in China but provided a substantial amount of PHCs. Third, 10 SP cases will be developed through a standardized process using the same template and methodology, and represent common conditions in PHC, while past studies often used 2-3 conditions.[32] Fourth, an evidence-based systematic method will guide the checklist development. In a review, only 12 out of the 29 SP articles reported the procedures of the checklist development and many checklists were developed by expert consensus only.[53] Fifth, in addition to using the checklist to evaluate

technical quality as performed in most other USP studies, we will assess patient-centeredness with a global rating scales. Sixth, we have planned a series of related studies to address the quality of PHC in a concerted effort. Most noteworthy, we are developing 10 identical conditions as smartphone-based virtual patients to assess the competency of PHC providers. Seventh, we used the same case for all levels of providers from village doctors to township health centers to county hospitals, but quality checklists for process, diagnosis and treatment will be tailored to fit the expected roles and responsibilities of the different providers. Eighth, we have secured the understanding and cooperation from the provincial health authorities. Finally, the project has involved researchers from Nepal as well as 20 universities across 19 provinces in China in a USP Network (https://www.researchgate.net/project/Unannounced-Standardized-Patient-USP-and-Virtual-Patient-VP-to-Measure-Quality-of-Primary-Care). The USP resources will be pooled and shared widely within the network first and then with the general public.

We note two particular issues. In high-income settings, logistical arrangements for the SP is complex. A significant challenge is to introduce the SP into medical practice.[23 46 47] However, in China and many other LMICs, enrollment with a clinician is not required, and a walk-in visit to clinicians without an appointment is commonplace. However, village doctors usually know their patients well. For these areas, the SPs in other studies pretended to be tourists or friends visiting the families in the village. We will try other pretenses such as a temporary poverty-relief worker who has just arrived in a nearby village. Those poverty-relief workers are common in remote rural areas in China. On a second issue, assessing quality with USP was reported to incur high cost in the developed countries (estimated to be USD 350-400 per visit).[52 69] We expect the cost in China to be considerably less due to the lower labor cost. We will collect detailed cost information to inform the future application of the USP.

The study has several potential limitations. First of all, the USP method has several technical challenges. If healthy people are used to simulate the patient, it is difficult to achieve complete alignment of patient presentation of signs and symptoms (for instance, it is difficult to fake a sore throat). There are also challenges of obtaining fake laboratory-test results that may be necessary for the diagnosis. Some clinical roles that require the SP to go through invasive investigation may also pose a problem. We will experiment with a real patient with stable conditions to resolve some of those challenges. Second, our judgment of the clinical quality through the first and only visit with the SP may lead to "first-visit bias".[37] The quality of a clinician who spreads out his or her diagnosis and management over several visits may be underestimated. We try to minimize this bias by designing cases that require a definitive decision on the first visit. Lastly, even though we intend to select ten tracer conditions in the context of PHC, we still need to be cautious in generalizing the findings to the overall quality of PHC.

In conclusion, this proposed study may produce a set of validated tools for the assessment of the quality of PHC using the USP and apply it to obtain valuable quality information of China's PHC.

Tables

Table 1 Selected Candidate Conditions

Conditions	Special Focus Areas											
	Chronic Disease Management	Public Health Delivery	Mental Health	Maternal & Child Care	Preventative Care	Referral	Patient-centered care	Older Adults	Low value diagnostic	Antibiotics	Process Traditional Chinese Drug	Injury
Common cold (flu season)					✖					✖	✖	
Hypertension	✖										✖	
T2DM	✖						✖	✖			✖	
Gastritis							✖					
Child diarrhea				✖						✖		
Low back pain (patient requesting low-value test)							✖		✖			
Depression (Maternal care)			✖	✖		✖	✖					
Angina (heavy smoker)					✖	✖	✖				✖	
Headache											✖	
Fall					✖		✖	✖				✖
Asthma												
Tuberculosis		✖			✖	✖						

Table 2 Variables

	Variable name	Type	Coding	Source
1. Effectiveness & Safety				
1.1	% of recommended questions asked	continuous	0-1	SP checklist
1.2	% of recommended exams performed	continuous	0-1	SP checklist
1.3	Diagnosis quality	ordinal	0: incorrect 1:partially correct 2:correct	SP checklist
1.4	Treatment quality	ordinal	0: incorrect 1:partially correct 2:correct	SP checklist
2. Patient-centeredness				
2.1	Patient perception of patient-centeredness	continuous	0-1	PPPC
2.2	Choice of provider	dichotomous	0: no 1: yes	SP checklist
2.3	Ease of navigation in facility	ordinal	0: difficult 1: median 2: easy	SP rating
3. Timeliness				
3.1	Opening hours	continuous	hours	SP checklist
3.2	Wait time	continuous	minutes	SP checklist
3.3	Consultation time	continuous	minutes	SP checklist
4. Efficiency				
4.1	Total cost	continuous	RMB	SP checklist
4.2	Medication cost	continuous	RMB	SP checklist
4.3	Laboratory/imaging cost	continuous	RMB	SP checklist
5. Equity				
5.1	To be analyzed in a separate cross-over trial			

Figures

Figure 1 Selected seven sample provinces on the map of China with referencing countries of equivalent life expectancy in the bracket

Figure 2 Sampling Procedure

Author contribution

DX conceived the project concept and developed the first protocol draft along with WG. DX, MH and WH developed the sampling design; and MH, WH and EM wrote the section on samples and performed the sample size calculation. SS provided original data of the previous studies for the sample size estimation and calculated some summary statistics. JL and YC worked on the SP case templates. YC and XW developed the guideline for the development of the quality checklist. KH reviewed the content and edited the manuscript. HH and GC reviewed the statistical plan. SR, JP, HW, ZZ, CT, WZ reviewed and commented on the design and methods. All co-authors participated in the revision and approved this draft.

Competing interests

We report no competing interests.

Funding

The project is funded through a competitive grant from the China Medical Board (grant number: CMB GNL 16-260), National Natural Science Foundation of China (grant number: 81773446) and Young Scientists Fund of the National Natural Science Foundation of China (grant number: 81402690).

Data sharing statement

We have not yet started the data collection. However, the data generated from this project and the USP cases and the accompanying user manuals will be made available to other researchers upon request after we complete our primary analysis.

References

1. A/RES/70/1 R. Transforming our world: the 2030 agenda for sustainable development 2015 [Available from: http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E accessed Feburary 17, 2018 2018.
2. Hanefeld J, Powell-Jackson T, Balabanova D. Understanding and measuring quality of care: dealing with complexity. *Bulletin of the World Health Organization* 2017;95(5):368.
3. Murray CJ, Frenk J. A WHO framework for health system performance assessment: Evidence and Information for Policy, World Health Organization 1999.
4. Donabedian A. The quality of care: how can it be assessed? *Archives of pathology & laboratory medicine* 1997;121(11):1145.
5. Pongsupap Y, Lerberghe WV. Choosing between public and private or between hospital and primary care: responsiveness, patient-centredness and prescribing patterns in outpatient consultations in Bangkok. *Tropical Medicine & International Health* 2006;11(1):81-89.
6. Bitton A, Ratcliffe HL, Veillard JH, et al. Primary health care as a foundation for strengthening health systems in low-and middle-income countries. *Journal of general internal medicine* 2017;32(5):566-71.
7. Wei X, Li H, Yang N, et al. Changes in the perceived quality of primary care in Shanghai and Shenzhen, China: a difference-in-difference analysis. *Bulletin of the World Health Organization* 2015;93(6):407-16.
8. Zou Y, Zhang X, Hao Y, et al. General practitioners versus other physicians in the quality of primary care: a cross-sectional study in Guangdong Province, China. *BMC family practice* 2015;16(1):134.
9. Feng S, Shi L, Zeng J, et al. Comparison of Primary Care Experiences in Village Clinics with Different Ownership Models in Guangdong Province, China. *PloS one* 2017;12(1):e0169241.
10. Wong WC, Jiang S, Ong JJ, et al. Bridging the Gaps between patients and primary care in China: a nationwide representative survey. *The Annals of Family Medicine* 2017;15(3):237-45.
11. Zeng L, Li Y, Zhang L, et al. Guideline use behaviours and needs of primary care practitioners in China: a cross-sectional survey. *BMJ open* 2017;7(9):e015379.
12. Li X, Lu J, Hu S, et al. The primary health-care system in China. *The Lancet* 2017;390(10112):2584-94.
13. Das J, Hammer J. Quality of primary care in low-income countries: facts and economics. *Annu Rev Econ* 2014;6(1):525-53.
14. Sylvia S, Shi Y, Xue H, et al. Survey using incognito standardized patients shows poor quality care in China's rural clinics. *Health policy and planning* 2014;30(3):322-33.
15. Berendes S, Heywood P, Oliver S, et al. Quality of private and public ambulatory health care in low and middle income countries: systematic review of comparative studies. *PLoS*

medicine 2011;8(4):e1000433.
16. 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 affairs* 2012;31(12):2774-84.
17. Das J, Gertler PJ. Variations in practice quality in five low-income countries: a conceptual overview. *Health affairs* 2007;26(3):w296-w309.
18. Das J, Hammer J, Leonard K. The quality of medical advice in low-income countries. *Journal of Economic Perspectives* 2008;22(2):93-114.
19. Coarasa J, Das J, Gummerson E, et al. A systematic tale of two differing reviews: evaluating the evidence on public and private sector quality of primary care in low and middle income countries. *Globalization and health* 2017;13(1):24.
20. Glassman PA, Luck J, O'Gara EM, et al. Using standardized patients to measure quality: evidence from the literature and a prospective study. *Joint Commission Journal on Quality and Patient Safety* 2000;26(11):644-53.
21. Leonard K, Masatu MC. Outpatient process quality evaluation and the Hawthorne Effect. *Social science & medicine* 2006;63(9):2330-40.
22. McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. *Journal of clinical epidemiology* 2014;67(3):267-77.
23. Woodward CA, McConvey GA, Neufeld V, et al. Measurement of physician performance by standardized patients: refining techniques for undetected entry in physicians' offices. *Medical care* 1985:1019-27.
24. Das J, Hammer J. Money for nothing: the dire straits of medical practice in Delhi, India. *Journal of Development Economics* 2007;83(1):1-36.
25. 钟玉杰, 王敏, 李勤. 从 10 年文献回顾分析我国标准化病人教学的发展. *中华护理杂志* 2009;44(3):259-61.
26. Currie J, Lin W, Zhang W. Patient knowledge and antibiotic abuse: Evidence from an audit study in China. *Journal of health economics* 2011;30(5):933-49.
27. 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 Medicine* 2017;14(10):e1002405.
28. Li L, Lin C, Guan J. Using standardized patients to evaluate hospital-based intervention outcomes. *International journal of epidemiology* 2013;43(3):897-903.
29. Zhou M, Wang H, Zhu J, et al. Cause-specific mortality for 240 causes in China during 1990–2013: a systematic subnational analysis for the Global Burden of Disease Study 2013. *The Lancet* 2016;387(10015):251-72.
30. Das J, Hammer J. Which doctor? Combining vignettes and item response to measure clinical competence. *Journal of Development Economics* 2005;78(2):348-83.
31. Hambleton RK, Swaminathan H, Rogers HJ. Fundamentals of item response theory: Sage 1991.
32. Rethans JJ, Gorter S, Bokken L, et al. Unannounced standardised patients in real practice: a systematic literature review. *Medical education* 2007;41(6):537-49.
33. Kutob RM, Bormanis J, Crago M, et al. Assessing culturally competent diabetes care with unannounced standardized patients. *Fam Med* 2013;45(6):400-08.
34. Fenton JJ, Kravitz RL, Jerant A, et al. Promoting patient-centered counseling to reduce use of low-value diagnostic tests: a randomized clinical trial. *JAMA internal medicine*

2016;176(2):191-97.
35. May L, Franks P, Jerant A, et al. Watchful waiting strategy may reduce low-value diagnostic testing. *The Journal of the American Board of Family Medicine* 2016;29(6):710-17.
36. ORDERING OF LABS AND TESTS: VARIATION AND CORRELATES OF VALUE-BASED CARE IN AN UNANNOUNCED STANDARDIZED PATIENT VISIT. JOURNAL OF GENERAL INTERNAL MEDICINE; 2016. SPRINGER 233 SPRING ST, NEW YORK, NY 10013 USA.
37. Tamblyn RM, Abrahamowicz M, Berkson L, et al. First-visit bias in the measurement of clinical competence with standardized patients. *Academic Medicine* 1992;67(10):S22-4.
38. Shepherd HL, Barratt A, Trevena LJ, et al. Three questions that patients can ask to improve the quality of information physicians give about treatment options: a cross-over trial. *Patient education and counseling* 2011;84(3):379-85.
39. Peabody JW, Luck J, Jain S, et al. Assessing the accuracy of administrative data in health information systems. *Medical care* 2004;42(11):1066-72.
40. Organization WH. WHO handbook for guideline development: World Health Organization 2014.
41. Campbell S, Braspenning J, Hutchinson A, et al. Research methods used in developing and applying quality indicators in primary care. *Qual Saf Health Care* 2002;11(4):358-64.
42. De Champlain AF, Margolis MJ, King A, et al. Standardized patients' accuracy in recording examinees' behaviors using checklists. *Academic Medicine* 1997;72(10):S85-7.
43. Vu NV, Steward DE, Marcy M. An assessment of the consistency and accuracy of standardized patients' simulations. *Academic Medicine* 1987;62(12):1000-2.
44. Vu NV, Marcy M, Colliver J, et al. Standardized (simulated) patients' accuracy in recording clinical performance check-list items. *Medical Education* 1992;26(2):99-104.
45. Maiburg BH, Rethans JJE, Van Erk IM, et al. Fielding incognito standardised patients as 'known'patients in a controlled trial in general practice. *Medical education* 2004;38(12):1229-35.
46. L. Gorter J-JR, Albert JJA Scherpbier, Sjef van der Linden, Marijke HM van Santen-Hoeufft, Désirée MFM van der Heijde, Harry HML Houben, Cees PM van der Vleuten, Simone. How to introduce incognito standardized patients into outpatient clinics of specialists in rheumatology. *Medical teacher* 2001;23(2):138-44.
47. Siminoff LA, Rogers HL, Waller AC, et al. The advantages and challenges of unannounced standardized patient methodology to assess healthcare communication. *Patient education and counseling* 2011;82(3):318-24.
48. Oates J, Weston WW, Jordan J. The impact of patient-centered care on outcomes. *Fam Pract* 2000;49(9):796-804.
49. Hudon C, Fortin M, Haggerty JL, et al. Measuring patients' perceptions of patient-centered care: a systematic review of tools for family medicine. *The Annals of Family Medicine* 2011;9(2):155-64.
50. Brown J, Stewart M, Tessier S. Assessing communication between patients and doctors: a manual for scoring patient-centred communication. *London: Thames Valley Family Practice Research Unit* 1995
51. Ozuah PO, Reznik M. Can standardised patients reliably assess communication skills in asthma cases? *Medical education* 2007;41(11):1104-05.
52. Zabar S, Ark T, Gillespie C, et al. Can unannounced standardized patients assess

professionalism and communication skills in the emergency department? *Academic Emergency Medicine* 2009;16(9):915-18.

53. Gorter S, Rethans J-J, Scherpbier A, et al. Developing Case-specific Checklists for Standardized-patient—Based Assessments in Internal Medicine: A Review of the Literature. *Academic Medicine* 2000;75(11):1130-37.

54. Franz CE, Epstein R, Miller KN, et al. Caught in the act? Prevalence, predictors, and consequences of physician detection of unannounced standardized patients. *Health services research* 2006;41(6):2290-302.

55. Tamblyn RM. Use of standardized patients in the assessment of medical practice. *CMAJ: Canadian Medical Association Journal* 1998;158(2):205.

56. Swartz MH, Colliver JA, Bardes CL, et al. Validating the standardized-patient assessment administered to medical students in the New York City Consortium. *Academic medicine: journal of the Association of American Medical Colleges* 1997;72(7):619-26.

57. Rethans J, Drop R, Sturmans F, et al. A method for introducing standardized (simulated) patients into general practice consultations. *Br J Gen Pract* 1991;41(344):94-96.

58. Luck J, Peabody JW. Using standardised patients to measure physicians' practice: validation study using audio recordings. *Bmj* 2002;325(7366):679.

59. Shirazi M, Sadeghi M, Emami A, et al. Training and validation of standardized patients for unannounced assessment of physicians' management of depression. *Academic Psychiatry* 2011;35(6):382-87.

60. Lin L. A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometric, 45, 255-268, 1989.

61. Steichen TJ, Cox NJ. A note on the concordance correlation coefficient. *Stata J* 2002;2(2):183-89.

62. Lawrence I, Lin K. Assay validation using the concordance correlation coefficient. *Biometrics* 1992:599-604.

63. Kwiecien R, Kopp-Schneider A, Blettner M. Concordance analysis: part 16 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt International* 2011;108(30):515.

64. Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet* 1986;327(8476):307-10.

65. Austin PC, Grootendorst P, Anderson GM. A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. *Statistics in medicine* 2007;26(4):734-53.

66. Stuart EA, King G, Imai K, et al. MatchIt: nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software* 2011;42(8)

67. Rhodes K. Taking the mystery out of "mystery shopper" studies. *New England Journal of Medicine* 2011;365(6):484-86.

68. Rhodes KV, Miller FG. Simulated patient studies: an ethical analysis. *The Milbank Quarterly* 2012;90(4):706-24.

69. Weiner SJ, Schwartz A. Directly observed care: can unannounced standardized patients address a gap in performance measurement? *Journal of general internal medicine* 2014;29(8):1183-87.

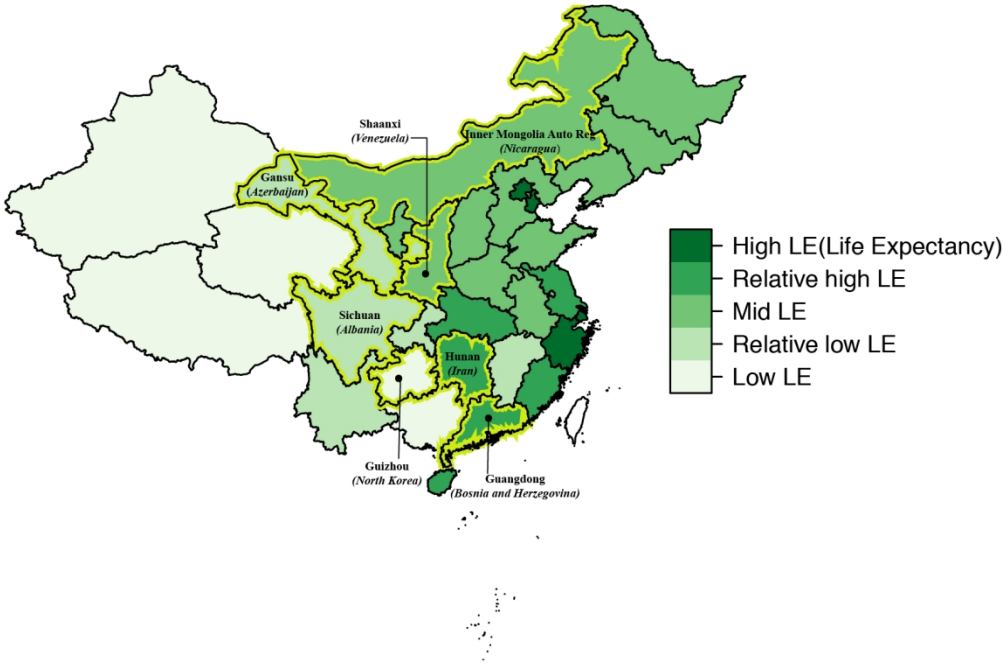

Figure 1 Selected seven sample provinces on the map of China with referencing countries of equivalent life expectancy in the bracket

147x98mm (300 x 300 DPI)

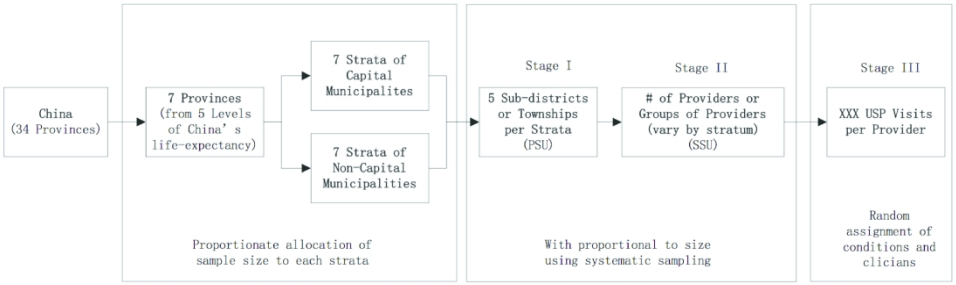

Figure 2 Sampling Procedure

397x122mm (300 x 300 DPI)

Web Appendix

Web Appendix 1 Evidence-based process of developing quality criteria for the SP cases

In partnership with the Lanzhou University Evidence Medicine Center, we have developed a working paper on the results of our review of the literature in quality checklist development and also our recommended protocol of developing those checklists. We provide an abstract of that working paper below and will make available the full paper once it is fully developed.

Abstracts

Objective To explore the procedures and methods for determining the quality checklist for the most common conditions in the context of primary health care, particularly to be used for quality inspection by unannounced standardized patients.

Methods We conducted a systematic search of literature in the subject matter, while adopting the WHO handbook for guideline development. **Results** A total of 14 related articles were included and the methodological aspects were evaluated. Based on this review, we propose five key steps in the checklist development: (1) Forming a multidisciplinary team; (2) Reviewing, evaluating and selecting relevant literature based on evidence-based medicine quality of evidence principles; (3) Extracting

essential quality information to form a pool of quality items; (4) using expert consensus to select candidate quality checklist items from the pool; (5) pre-testing to determine the final items. **Discussion** We recommend a checklist development method based on evidence-based method augmented by expert opinions through a multidisciplinary group discussion. The selection of the items on the checklist will consider their importance and feasibility. Our proposed methods can be mainly applied to common conditions seen in the primary care settings and may not be applied to more complex conditions.

Web Appendix 2 Sample Size Calculation

Compute the sampling variance of the mean: $var(\bar{y})$, based on desired coefficient of variation - 0.08.

$$var(\bar{y}) = se(\bar{y})^2 = (cv * \bar{y})^2 = (0.08 * (-0.9))^2 = 0.0052$$

Estimate number of completed interviews in need for a simple random sample(SRS):n_{srs}

$$n_{srs} = \frac{s^2}{var(\bar{y})} = \frac{4.54}{0.0052} = 875$$

Estimate design effect:

$$d_{eff} = 1 + \delta(n-1) = 1 + 0.0486 * (27-1) = 2.26$$

Multiply n_{srs} by the design effect to account for a complex survey design:

$$n_{complex} = n_{srs} * d_{eff} = 875 * 2.26 \approx 1981$$

BMJ Open

Assessing the Quality of Primary Health Care in 7 Chinese Provinces with Unannounced Standardized Patients: Protocol of a Cross-sectional Survey

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-023997.R1
Article Type:	Protocol
Date Submitted by the Author:	22-Aug-2018
Complete List of Authors:	Xu, Dong; Sun Yat-sen University, Sun Yat-sen Global Health Institute (SGHI),School of Public Health and Institute of State Governance Hu, Mengyao; University of Michigan, Survey Research Center, Institute for Social Research He, Wenjun; Sun Yat-sen University, Department of Biostatistics and Epidemiology, School of Public Health Liao, Jing; Sun Yat-sen University, Sun Yat-sen Global Health Institute (SGHI), School of Public Health and Institute of State Governance Cai, Yiyuan; Guizhou Medical University, School of Public Health; Sun Yat-Sen University, School of Public Health Sylvia, Sean ; University of North Carolina at Chapel Hill, Department of Health Policy and Management, Gillings School of Global Public Health Hanson, Kara; London School of Hygiene & Tropical Medicine, Department of Global Health and Development, Faculty of Public Health and Policy Chen, Yao-Long; Lanzhou University, Evidence-Based Medicine Center, School of Basic Medical Sciences Pan, Jay; Sichuan University, West China School of Public Health Zhou, Zhongliang; Xi'an Jiaotong University, School of Public Policy and Administration Zhang, Nan; Inner Mongolia Medical University, Department of Health Management, School of Health Management Tang, Chengxiang; Guangzhou University, School of Public Administration Wang, Xiaohui; Lanzhou University, Evidence Based Medicine Center, School of Basic Medical Sciences Rozelle, Scott; Stanford University, Freeman Spogli Institute for International Studies He, Hua; Tulane University, Department of Epidemiology, School of Public Health and Tropical Medicine Wang, Hong; Bill & Melinda Gates Foundation, Health Economics, Financing & Systems Chan, Gary; University of Washington, Department of Biostatistics Melipillán, Edmundo; University of Michigan, Survey Research Center, Institute for Social Research Zhou, Wei; Central South University, Xiangya school of public health, Gong, Wenjie; Central South University, School of Public Health; University of Rochester Medical Center, Department of Psychi

Primary Subject Heading:	Health policy
Secondary Subject Heading:	Public health
Keywords:	standardized patients, unannounced standardized patients, quality of primary health care, patient-centered care

Assessing the Quality of Primary Health Care in 7 Chinese Provinces with Unannounced Standardized Patients: Protocol of a Cross-sectional Survey

Dong Roman Xu,[1] Mengyao Hu,[2] Wenjun He,[3] Jing Liao,[1] Yiyuan Cai,[4,3,1] Sean Sylvia,[5] Kara Hanson,[6] Yaolong Chen,[7] Jay Pan,[8] Zhongliang Zhou,[9] Nan Zhang,[10] Chengxiang Tang,[11] Xiaohui Wang,[7] Scott Rozelle,[12] Hua He,[13] Hong Wang,[14] Gary Chan,[15] Edmundo Roberto Melipillán,[2] Wei Zhou,[16] Wenjie Gong[17*]

[1] Sun Yat-sen Global Health Institute (SGHI), School of Public Health and Institute of State Governance, Sun Yat-sen University, No.74 Zhongshan 2nd Road, Guangzhou, P.R. China, 510080

[2] Survey Research Center, Institute for Social Research, University of Michigan, 426 Thompson Street, Ann Arbor, MI 48104, USA

[3] Department of Biostatistics and Epidemiology, School of Public Health, Sun Yat-sen University, No.74 Zhongshan 2nd Road, Guangzhou, P.R. China, 510080

[4] School of Public Health, Guizhou Medical University, Guizhou, P.R. China, 550025

[5] Department of Health Policy and Management, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, CB#7411, Chapel Hill, NC 27599, USA

[6] Department of Global Health and Development, Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine, WC1H 9SH, London, UK

[7] Evidence Based Medicine Center, School of Basic Medical Sciences, Lanzhou University, No. 199 Donggang West Road, Lanzhou, Gansu Province, 730000 P.R. China

[8] West China School of Public Health, Sichuan University, No. 17, Ren Min Nan Road, Chengdu, P.R. China

[9] School of Public Policy and Administration, Xi'an Jiaotong University, No. 28 Xianning West Road, Xi'an, Shaanxi, 710049, P.R. China.

[10] Department of Health Management, School of Health Management, Inner Mongolia Medical University, Jinshan Development Zone, Hohhot, Inner Mongolia, P.R. China, 010110

[11] School of Public Administration, Guangzhou University, Guangzhou, Guangdong, 510320, P.R. China.

[12] Freeman Spogli Institute for International Studies, Stanford University, Encina Hall East, E407, Stanford, CA 94305-6055, USA

[13] Department of Epidemiology, School of Public Health and Tropical Medicine, Tulane University, 1440 Canal Street, New Orleans, LA 70112, USA

[14] Health Economics, Financing & Systems, Bill & Melinda Gates Foundation, PO Box 23350, Seattle WA, 98102, USA

[15] Department of Biostatistics, University of Washington, 1705 NE Pacific Street, #F600, Seattle, WA 98195, USA

[16] Hospital Administration Institute, Xiangya Hospital, Central South University, No. 87 Xiangya Road, Changsha, Hunan, P.R. China

[17] Xiangya school of public health, Central South University, Changsha, Hunan, China

Corresponding author: Wenjie Gong, gongwenjie@csu.edu.cn, Xiangya School of Public Health, Central South University, Xiangya school of public health, Central South University, P.R. China*

Abstract

Introduction: Primary health care (PHC) serves as the cornerstone for the attainment of universal health coverage (UHC). Efforts to promote UHC should focus not only on the expansion of access but also on healthcare quality. However, robust quality evidence has remained scarce in China. Common quality assessment methods such as chart abstraction, patient rating, and clinical vignette use indirect information that may not represent real practice. This study will send standardized patients (SP or healthy person trained to consistently simulate the medical history, physical symptoms, and emotional characteristics of a real patient) unannounced to PHC providers to collect quality information and represent real practice..

Methods and Analysis: 1981 SP-clinician visits will be made to a random sample of PHC providers across seven provinces in China. SP cases will be developed for 10 tracer conditions in PHC. Each case will include a standard script for the SP to use and a quality checklist that the SP will complete after the clinical visit to indicate diagnostic and treatment activities performed by the clinician. Patient-centeredness will be assessed according to the Patient Perception of Patient-centeredness (PPPC) rating scale by the SP. SP cases and the checklist will be developed through a standard protocol and assessed for content, face and criterion validity and test-retest and inter-rater reliability before its full use. Various descriptive analyses will be performed for the survey results such as a tabulation of quality scores across geographies and provider types.

Ethics and dissemination: This study has been reviewed and approved by the Institutional Review Board of the School of Public Health of Sun Yat-sen University (#SYSU 2017-011). Results will be actively disseminated through print and social media, and SP tools will be made available for other researchers.

Keywords: standardized patients; unannounced standardized patients; quality of primary health care; patient-centered care

Strengths:

- We will assess the quality of care with a random sample of primary health care providers in seven provinces in China.
- We will use unannounced standardized patients (USPs) - the "gold standard" of quality assessment.
- Both technical quality and patient-centeredness will be assessed.

Limitations:

- USPs are not suitable for certain health conditions.
- The seven provinces are not randomly selected, although we intend them to represent different health development conditions of China's provinces.

Assessing the Quality of Primary Health Care in 7 Chinese Provinces with Unannounced Standardized Patients: Protocol of a Cross-sectional Survey

Background

In 2015, all 191 UN member states adopted the Sustainable Development Goals (SDGs), aiming to achieve universal health coverage (UHC) – the access to high-quality health care services without incurring financial hardship – by 2030.[1] As previous literature emphasized, efforts to promote UHC should focus not only on the expansion of access but also on healthcare quality.[2] Healthcare quality is variously defined by the WHO as the "responsiveness" of the health care system to meet desired health outcomes,[3] as the instrumental goals on structure, process, and outcome in the Donabedian Framework,[4] and as the six comprehensive aims (effectiveness, efficiency, equity, patient-centeredness, safety, and timeliness) put forth by the Institute of Medicine (IOM).[5] In this study, we take the IOM definition of quality.

Primary health care (PHC) serves as the cornerstone for the attainment of UHC.[6] China's latest round of health care reform since 2009 has invested heavily in strengthening PHC. There have been some efforts to assess the quality of PHC in China: patients were interviewed with a Primary Care Assessment Tool (PCAT) questionnaire in Guangdong, Shanghai, and Hong Kong;[7-9] comprehensiveness of the service provision was used as a proxy for quality through clinician interviewing;[10] and PHC clinicians' adherence to clinical guideline was assessed with a self-report questionnaire.[11] However, assessment of the quality of PHC has largely remained scanty in China, and the assessment tools are indirect and prone to bias.[12] A number of studies have found the quality of PHC to be low in other low and middle income countries (LMICs)[6 13-18], where robust evidence remains scarce.[19] Commonly-used methods of measuring technical

quality of care include chart abstraction, patient rating of care, and using a clinical vignette to test clinician knowledge. Those methods use indirect information that may not represent real practice. This study instead will use unannounced standardized patients (USPs) to measure the quality of real practice. The Standardized Patient (SP) is a healthy person (or occasionally a real patient) trained to consistently simulate the medical history, physical symptoms, and emotional characteristics of a real patient. The SP, particularly when their visit is unannounced, has several reported advantages: (1) reliability in measurement and cross-provider comparison because the same patient is presented to all providers, (2) elimination of the Hawthorne effect (i.e., that the study itself may change doctors' behavior) due to the nature of disguised and unannounced visit by SPs,[20-22] and (3) reduced recall bias.[23 24]

Despite these advantages, the application of SP in China has been concentrated mainly in the area of medical education.[25] An ongoing systematic review identified four papers only on using the SP for quality assessment in China,[14 26-28], and 44 in other LMICs. Those projects, often based on a small convenience sample, tended to target a limited number of conditions (approximately 70% on family planning services, childhood infectious diseases, sexually transmitted infections, and respiratory tract infections). In this study, we intend to assess the quality of PHC with a probability sample of PHC visits in seven Chinese provinces, using USPs for 10 commonly seen conditions in the PHC setting. The project has involved 20 universities across 19 provinces in China as well as researchers from Nepal, US, and UK in a USP Network (https://www.researchgate.net/project/Unannounced-Standardized-Patient-USP-and-Virtual-Patient-VP-to-Measure-Quality-of-Primary-Care). The USP resources will be pooled and shared widely within the network first and then with the general public. This study is the first of a series of studies to be based on quality data collected using USPs. The primary purpose of this study is

to collect and present descriptive data on the quality of China's PHC. We are developing separate protocols for the various hypothesis-driven studies, which will be available elsewhere and from our Network website.[29]

Methods

Survey Design

The purpose of the sample design is to create a representative sample of China's primary health care (PHC) providers so that healthcare quality can be assessed based on USP visits to those providers.

Survey Population/Frame

We considered creating a nationally representative probability sample, but at this stage, we have selected seven provinces to "represent" China due to feasibility considerations. These provinces represent five levels of average life expectancies across China's provinces (Figure 1), which are similar to those of five countries with low to high income levels.[30] We intend to create a probability sample that represents primary health care in these seven provinces. For the survey population, we intend to include (1) licensed physician and licensed assistant physician at community/township health centers/stations and urban health stations, (2) certified village doctors (a terminology in China that refers to village clinicians who have village-level practice privilege even without a medical license) and village sanitarians (referring to un-certified village doctors who are supposed to work under the supervision of the village doctor) at village clinics; and (3) clinicians with a license notation for general practice, internal medicine, obstetrics/gynecology, and pediatrics at the level 1 and level 2 hospitals and the maternal and child care centers. We exclude level 3 hospitals, which provide more specialized care, and specialty hospitals. Clinician meeting those criteria will constitute the "sampling frame".

Sampling Procedures

The sample will be selected using a multi-stage, clustered sample design covering all eligible clinicians in the seven provinces (Figure 2). In the first stage, stratification will be based on the provinces. Due to the high number of visits in the seven capital cities, we will sample each capital city. Each province is thus divided into two strata consisting of the provincial capital city and other prefecture-level municipalities, leading to 14 strata in total. We will use proportionate allocation (in terms of the number of eligible clinicians) of sample size for each stratum. For each stratum, five rural townships or urban sub-districts (the primary sampling unit/PSU) will be selected using probability proportional to size (PPS). In the second stage, for each PSU, PHC facilities as previously-defined (Secondary Sampling Unit/SSU) will be selected using PPS systematic sampling. Neighboring village clinics will be grouped as an SSU. The number of SSUs for each stratum will vary depending on the size of the stratum – for example, more SSUs will be selected in strata with more PHC clinicians. In the final stage, a fixed number of USP visits will be made to each selected facility or the group of facilities in the case of village clinics. The exact number of visits will be determined once we obtain and examine our sampling frame.

Sample Size Calculation

The sample size was calculated for the primary purpose of the standard descriptive survey analysis of this survey. The sample size (power) calculation for the other related hypothesis related studies will be described in separate study protocols. The primary statistic of interest in this survey is a latent variable measuring clinician's quality, constructed using the 2-parameter logistic item response theory (IRT) model.[31 32] The model was based on a list of quality checklist items measuring whether doctors asked recommended questions and whether they performed recommended exams (see the *Scoring Method* section below). Survey sample

size was calculated based on the desired level of relative precision (coefficient of variation, CV), an estimate for the population element variance for the variable of interest (s^2) from previous study and design effect ($deff$). In this study, our desired level of relative precision (CV) is 0.08. s^2 was estimated to be 4.54, based on Sylvia et al's work on the USP-assessed quality of PHC in three Chinese provinces.[14 27] Design effect is the variance inflation due to cluster sampling. This figure was calculated based on intra-class correlation (ICC) (describing the level of homogeneity of the units in a cluster) and cluster sample size: $deff = 1 + \delta(n-1)$, where δ is the intra-class correlation (ICC) and n is the average size of the cluster. The ICC of 0.0486 was also estimated from Sylvia et al's work, which was 0.0486. Our estimated average cluster size is 27 clinician-SP encounters per PSU. Accordingly, we calculated the total required sample size to be 1981 clinician-SP encounters. The steps taken to calculate sample size can be found in Web Appendix 1.

USP Case Development and Implementation

The development process of a USP case is based on our extensive literature review,[20 33] as well as our own USP experiences in Shaanxi Province, China.[14 27] We are concurrently developing smartphone-based virtual standardized patients (VPs) (details described elsewhere). The two projects will share almost identical case scenarios and quality criteria.

Case Selection

Our purpose is to select ten health problems as tracer conditions for PHC in China. Ideally our selected cases should (1) be highly prevalent in PHC settings, (2) carry challenging features in different aspects of PHC (e.g., some cases focus on curative care while others on prevention, disease management, culturally-sensitive care,[34] or misuse of low value tests[35-37]), (3) not involve invasive and painful procedures, (4) not require physical signs that cannot be

simulated (e.g., jaundice can be simulated with make-up, but heart murmurs cannot.[23]). We created a list of the top 30 conditions commonly seen in PHC in China, combining the results of two national surveys on PHC.[12] A panel of physicians and public health and health system researchers then applied the principles above and selected a dozen PHC problems for USP development (Table 1). Ten final conditions will be selected from this list.

Development Team

We have created an overall development team and 10 case-specific development teams. Each team includes case-specific specialists, general practitioners, and public health and health system researchers (Web Appendix 2). A third overall panel consisting of primary care providers at the village, township and community levels will review all cases for contextual appropriateness in primary care settings. In developing the case, we will follow several principles: (1) limiting case scenarios to those that require definitive clinician action on the first visit to minimize potential "first-visit bias",[38] (2) focusing on the presentation of symptoms for which evidence is well-established for diagnosis and management, (3) deriving some content of the cases from the actual case history of relevant patient files in real practice.[23]

Case Description

The case description describes the relevant clinical roles and psycho-social biographies of the SP.[39] We used a structured description of the cases as follows:

1. Social and demographic profile: (1) socioeconomic information: name, gender, age, ethnicity, education, occupation, family structure (e.g., Married and have two children but live alone), dress style (e.g., dressed in jeans, work boots and a well-worn but neat sweater), health insurance or other social program participation; (2) personality that may influence interaction with the clinician (e.g., non-proactive and introverted); (3)

lifestyle relevant to health (e.g., smoked one pack of cigarette since age 18, like fried pork but also eat much fruit, exercise regularly, watch TV a lot in spare time, play mahjong with friends, visit children every week)

2. Medical history: (1) disease information: severity of the condition (e.g., mild or severe depression), duration of the condition (the first onset? Previously diagnosed/existing (how long)?), comorbidity (any other physical and/or psychological problems?), (2) reason for seeking care for this specific visit (e.g., was feeling down for two months but depression worsened last week), (3) treatment/management already or currently received (e.g., a diabetic "patient" took metoprolol for hypertension but does not monitor his glucose / watch his diet/weight).
3. Physical examination: symptoms the SP will (and will not) portray (e.g., reduced appetite, but not showing agitation), and medical signs the SP has or does not have (e.g., heart murmur).
4. Laboratory and imaging: laboratory and imaging that a clinician may prescribe for the SP. The laboratory and imaging results of the SP may be generated from those of real typical patients.
5. Diagnosis: the correct diagnosis that the clinician should make based on the information presented by the SP.
6. Treatment and management: the decision of the clinician on what medications, procedures, advice, or referral will be given at the end of the consultation.

Script

Corresponding to the six components of the afore-mentioned case description, we will develop a detailed script for the SPs to use in their PHC visit with the clinician. The script

ideally should cover all possible questions a clinician may ask as well as the SP's answers during the clinical interaction. Panels of clinicians will be consulted to collect relevant questions that will guide the development of the script. The script will continue to add new questions asked by the clinicians on the SP-clinician interaction. The script will have five sections: (1) an opening: spontaneous information given to the clinician at the start (e.g., Doctor, I have had a headache for two days), (2) the information given only on request, (3) the information for the SP to volunteer even if not asked, (4) the language to insist on a diagnosis if not given, and (5) an ending.[14 20 40]

Quality Checklist

The checklist consists of explicit quality criteria for gathering data on patient history, physical examination, laboratory/imaging, diagnosis and treatment.[14 33] Based on our comprehensive review of 14 articles of literature and evidence-based clinical guideline development methodology,[41] we have established a guiding principle and standard protocol for checklist development. Our process will (1) be evidence-based and augmented by expert opinion,[42] (2) follow a systematic procedure to gather, evaluate and select evidence and criteria, (3) select criteria related to clinician actions that the SP can easily evaluate,[43] (4) keep the number of checklist items under 30 to include high-priority criteria only so that the SP can reliably recall clinician behaviour[43-45]. The details of our checklist development protocol will be described in a separate paper, and key messages are summarized in Web Appendix 2.

Selecting and Training SPs

We will advertise on social media to recruit SPs. The candidate must be in stable health without confounding symptoms; should match the real patients in age, sex, and physical features; are willing to allow the examinations appropriate to their condition; have the intellectual

maturity to present the behavior of the actual patient and complete the checklist.[23 46 47] We may consider recruiting real patients with stable conditions to portray the cases not subject to simulation.[23] The training of the SP will aim at portraying the signs, symptoms, and presentations, completing the checklist, and minimizing detection by the provider.[20] The week-long training will have three stages: classroom instruction, a dress rehearsal, and two field tests.[23 47 48] Each case will have three SPs who will be trained according to a standardized training manual that will be developed to guide the training and appraisal of the SPs.

Fielding and Implementing SPs

A disguise plan will be developed for each case to minimize physician detection of the SP status (e.g., convincing excuse for seeking care where they do not usually reside). In the pilot (instrument validation) phase, consent will be sought for audio recording (see below); in these cases, fieldwork will start only three to four weeks after consent is obtained. We will provide each SP with a calamity letter, explaining the project in case of their identity being exposed.

After the facilities are selected, and the number of visits per facility is determined, each of the planned visits will be given a unique identifier (e.g., facility A-1, facility A-2, facility B-1), which will then be randomly ordered to form a random sequence numbered from 1 to 1981 consecutively. One of the 10 SP cases will be randomly assigned to each number on this random sequence. The seven SPs per case will be dispatched to the seven provinces concurrently, one SP per province. If multiple clinicians are available in that facility at the time of a particular SP visit (PHC visits in China do not require appointments), the field coordinator will randomly select a clinician by drawing lots onsite. Each SP is expected to make a total of approximately 30 visits. We plan to complete those SP visits over a three-month time span.

In a separate but related study, a week after the visit of the SP, the same clinician will perform the same consultation but with a standardized virtual patient on a smartphone.[29] We will use this opportunity to administer a detection questionnaire to the clinician, asking whether they suspect they had any visit from an SP over the past week. The detected cases will be treated as missing data in the data analysis.

Variables

Outcome Variables

We will collect a variety of quality of care information and other related explanatory variables. The IOM quality framework (effective, safe, patient-centered, timely, efficient, and equitable) will be used for quality evaluation (Table 2). E**ffectiveness** (avoiding underuse and misuse) and **safety** (avoiding harm), traditional technical goals of quality of care, will be evaluated through the yes/no checklist discussed above (Web Appendix 2). **Patient-centeredness** (respectful of and responsive to individual preferences) will be assessed by the Patient Perception of Patient-centeredness (PPPC) rating scale.[49-51] Using a 4-point Likert scale, PPPC evaluates three dimensions of patient-centeredness: exploring the disease and illness experiences, understanding the whole person, finding common ground.[49] Prior studies have demonstrated the validity of SPs rating clinician communications.[52 53] A separate study will be conducted to test the validity of PPPC scale. **Timeliness** will be assessed by analyzing opening hours, waiting time, and consultation time.[5] **Efficiency** (avoiding waste) will be measured by costs of care of the SP-clinician encounter. **Equity of care** (no variance in quality because of personal characteristics) will be assessed through a separate but related study in a randomized cross-over trial.

Scoring Method

Technical quality of care will be reflected by a continuous score ranging from 0-1. We will evaluate further whether to classify checklist items in four categories (essential, important, indicated, and non-contributory) with corresponding numeric weights (3, 2, 1, and 0).[54] Two scoring methods will be used: 1) the simple scoring method will use the formula of items performed divided by the total number of items on the checklist for the process scores, whereas 2) the complex method will use an algorism based on item-response-theory (IRT).[31] Using the IRT model approach, we can obtain a latent performance score for each doctor, which has been corrected for measurement error. An ordinal variable will be used for diagnosis and management plans (Table 2), while patient-centeredness will follow the scoring methods of PPPC (possible range of score from 1-4).[51]

Other Variables

We will collect additional information on the predictors, confounders, and effect modifiers to the outcomes in the planned hypothesis testing of the related studies to this survey. The information will include qualifications of the clinician and facility information (environment, amenity, size, location, ownership type, and so forth).

Analytical Methods

USP Validation

USP validation will be based on a convenience sample of clinicians not included in our final survey sample in the project training and pilot phase. Those SP-clinician interactions in the pilot will be audio recorded and transcribed. **Validity** is the extent to which an instrument measures what it is supposed to measure. We will assess content, face, and criterion validity of the cases. Content validity will be assessed by an expert panel who will use a 4-point Likert scale

to evaluate the appropriateness of the written content of the cases. The face validity of the SP assessment depends on (1) the SP remaining undetected (detection ratio reported to be 5%-10%[55]), and (2) authentically and consistently portraying the clinical features of the case. We will send the participating clinician in the pilot a "detection form" to report their degrees of suspicion of any SP visit.[46] The authenticity of the SP presentation will be evaluated by checking the transcribed recording to discover whether a key piece of information was divulged by the SP when appropriately prompted, not divulged when prompted, or volunteered when not prompted. Criterion validity will be assessed through the agreement of the SP-completed checklist against that completed by a clinician based on the transcript of the visit (i.e., the clinician rating as the "gold standard").[56-59] Checklist items which depend on visual observation will be excluded.

Reliability examines the level of consistency of the repeated measurements. The inter-rater reliability of two SPs on the same condition and context will be assessed with two SPs completing the checklist for the same recorded transcript. Test-retest reliability will be analyzed by the concordance of assessment results of the same SP to score his or her own recorded encounter a month later).[57] The agreement will be analyzed with Lin's concordance correlation coefficient (r_c)[60]. r_c indicates how closely pairs of observation fell on a 45° line (the perfect concordance line) through the origin in addition to their correlation.[60-62] Bland-Altman plot will be used to visualize the concordance.[63 64] Table 3 summarizes our methods of validation.

Survey Analysis

We will focus on descriptive analysis to present the quality of PHC in the seven provinces. Hypothesis-driven analyses will be described in separate study protocols. For descriptive analysis, we will first present clinician and facility profiles in tables for all seven provinces and by each province. The clinician profile will include socio-demographic

information (age, gender, and ethnicity), professional qualification (general and medical education, licensure, and professional ranks), and service information (volume of visits, number of support personnel). The Facility profile will include information on operation and management (years in operation, ownership types, accreditation, level of hospitals, affiliation with medical universities, revenue, health insurance contracting, payment methods), clinical services (annual number of inpatient and outpatient visits, number of clinical departments), personnel (number of physicians, nurses, and attrition ratio), and equipment. Second, we will tabulate results of overall quality and sub-domains across administrative regions and provider types. Third, we will map out the locations of the facilities along with their quality scores with geospatial analytical tools. Finally, a T-test/Wilcoxon test or Chi-square test will be employed to compare quality differences between public versus private providers, primary care clinics/centers versus hospital outpatient services, care in rural versus urban areas, and across different conditions, clinician educational levels, and payment mechanisms.

Related Studies

This study protocol mainly deals with the descriptive analysis and presentation of the data to be collected by the USPs. Using the USP survey data, we have planned several related studies that will be covered by separate study protocols with details on the background, theoretical framework, and analytical methods. To summarize those related studies, we will assess (1) the effect of ownership types of the PHC providers (i.e., private versus public) on quality of PHC (study protocol under revision), (2) the know-do gap between the assessment results by a smartphone-based virtual standardized patient and USP (protocol already published),[29] (3) the effect of using smartphone-based virtual patient in improving clinician performance, (4) the effect of types of insurance carried by a patient on quality of care, (5) the

impact of gatekeeping by primary care providers on quality of TB care – a mathematical modeling study, and (6) clinician skills in handling low-value or harmful patient requested services, particularly antibiotics and some processed traditional Chinese medicine.

Ethical Considerations

This study has received ethical approval from the institutional review board (IRB) of the Sun Yat-sen University School of Public Health with a waiver of informed consent from each participating clinician. USP studies do not necessarily require the consent if they meet certain conditions.[65 66] Our waiver has been granted for the following reasons: (1) our study serves important public good while requiring informed consent may lead to considerable selection bias and greater risk for the detection of the SP; (2) this study does not intend to entrap or reveal identities of any institution or individual and all analyses will be conducted at the broader health system level (after data cleaning all individual identifiers will be destroyed); (3) no audio-visuals will be recorded during the SP-clinician encounter (however, in the pilot stage, we will seek informed consent from participating clinicians as we will use a disguised recording for the validation purposes).

Patient and public involvement

We selected the conditions for the USP partly based on results from surveys on common conditions in the context of primary health care as reported by the patient. The USP cases will also be reviewed by a panel that includes patients. The results of the studies will be widely distributed in scientific reports as well as social media to benefit policy-makers, clinicians, and patients.

Discussion

In this study, we will develop, validate and implement methods of assessing the quality of PHC using USPs. Compared to existing studies using USPs,[33] this proposed study has several distinctive features. First, we will establish a large probability random sample so that representative estimates of PHC quality can be achieved in the chosen seven provinces in China. Second, unlike previous studies,[14 27] we include not only village clinics, township health centers, and community health centers but also county hospitals and other level I and level II hospitals in the study. The latter were not officially designated as PHC facilities in China but provided a substantial amount of PHCs. Third, 10 SP cases will be developed through a standardized process using the same template and methodology and will represent common conditions in PHC, while past studies often used two to three conditions.[33] Fourth, an evidence-based systematic method will guide checklist development. In a review, only 12 out of 29 SP articles reported the procedures of checklist development and many checklists were developed by expert consensus only.[54] Fifth, in addition to using the checklist to evaluate technical quality of care as performed in most other USP studies, we will assess patient-centeredness with a global rating scales. Sixth, we have planned a series of related studies to address the quality of PHC in a concerted effort. Most noteworthy, we are developing ten identical conditions as smartphone-based virtual patients to assess the competency of PHC providers. Seventh, we used the same case for all levels of providers from village doctors to township health centers to county hospitals, but quality checklists for process, diagnosis and treatment will be tailored to fit the expected roles and responsibilities of the different providers. Finally, we have secured the understanding and cooperation of the provincial health authorities.

We note two particular issues. In high-income settings, logistical arrangements for the SP is complex. A significant challenge is to introduce the SP into medical practice.[23 47 48] However, in China and many other LMICs, enrollment with a clinician is not required, and a walk-in visit to clinicians without an appointment is commonplace. However, village doctors usually know their patients well. For these areas, the SPs in other studies pretended to be tourists or friends visiting the families in the village. We will try other pretenses such as a temporary poverty-relief worker who has just arrived in a nearby village. Those poverty-relief workers are common in remote rural areas in China. On a second issue, assessing quality with USP was reported to incur high cost in the developed countries (estimated to be USD 350-400 per visit).[53 67] We expect the cost in China to be considerably less due to the lower labor cost. We will collect detailed cost information to inform the future application of the USP.

The study has several potential limitations. Most important, even though the assessment of SP is considered the gold standard for measuring clinician performance, and in this study we have further expanded the use of SPs to evaluate other elements of quality in the IOM framework such as patient-centeredness, timeliness, and efficiency, we recognize that those quality of care elements are still largely clinician-related, and other important quality aspects such as the quality of laboratory testing cannot be assessed by our SPs. In addition, the USP method has several technical challenges. If healthy people are used to simulate the patient, it is difficult to achieve complete alignment of patient presentation of signs and symptoms (for instance, it is difficult to fake a sore throat). There are also challenges of obtaining fake laboratory-test results that may be necessary for the diagnosis. Some clinical roles that require the SP to go through invasive investigation may also pose a problem. We will experiment with a real patient in stable conditions to resolve some of those challenges. Next, our judgment of the clinical quality

through the first and only visit with the SP may lead to "first-visit bias".[38] The quality of care provided by a clinician who spreads his or her diagnosis and management over several visits may be underestimated. We try to minimize this bias by designing cases that require a definitive decision on the first visit. Last, even though we intend to select ten tracer conditions in the context of PHC, we still need to be cautious in generalizing the findings to the overall quality of PHC.

In conclusion, this proposed study may produce a set of validated tools for the assessment of the quality of PHC using USP and apply it to obtain valuable quality of care information on primary health care in China.

Tables

Table 1 Selected Candidate Conditions

Conditions	Special Focus Areas											
	Chronic Disease Management	Public Health Delivery	Mental Health	Maternal & Child Care	Preventative Care	Referral	Patient-centered care	Older Adults	Low value diagnostic	Antibiotics	Process Traditional Chinese Drug	Injury
Common cold (flu season)					✖					✖	✖	
Hypertension	✖										✖	
T2DM	✖						✖	✖			✖	
Gastritis							✖					
Child diarrhea				✖						✖		
Low back pain (patient requesting low-value test)							✖		✖			
Depression (Maternal care)			✖	✖		✖	✖					
Angina (heavy smoker)					✖	✖	✖				✖	
Headache											✖	
Fall					✖		✖	✖				✖
Asthma												
Tuberculosis		✖			✖	✖						

Table 2 Variables

	Variable name	Type	Coding	Source
1. Effectiveness & Safety				
1.1	% of recommended questions asked	continuous	0-1	SP checklist
1.2	% of recommended exams performed	continuous	0-1	SP checklist
1.3	Diagnosis quality	ordinal	0: incorrect 1:partially correct 2:correct	SP checklist
1.4	Treatment quality	ordinal	0: incorrect 1:partially correct 2:correct	SP checklist
2. Patient-centeredness				
2.1	Patient perception of patient-centeredness	continuous	0-1	PPPC
2.2	Choice of provider	dichotomous	0: no 1: yes	SP checklist
2.3	Ease of navigation in facility	ordinal	0: difficult 1: median 2: easy	SP rating
3. Timeliness				
3.1	Opening hours	continuous	hours	SP checklist
3.2	Wait time	continuous	minutes	SP checklist
3.3	Consultation time	continuous	minutes	SP checklist
4. Efficiency				
4.1	Total cost	continuous	RMB	SP checklist
4.2	Medication cost	continuous	RMB	SP checklist
4.3	Laboratory/imaging cost	continuous	RMB	SP checklist
5. Equity				
5.1	To be analyzed in a separate cross-over trial			

Table 3 Methods of Validation for the USP cases

Domain	Indicator	Data Collection		Statistical Analysis
		Phase	Method	
Content Validity	Content Validity Index (CVI)	USP case review	Expert panel review of SP cases, measured by a 4-point Likert scale (1=lowest, 4=highest)	CVI for SP case and for specific USP, where CVI = number of raters giving a rating of 3 or 4÷total number of raters
Face Validity	Authenticity of SP role-play	Validation Study	Transcripts of the recording of the USP-clinician encounter to be assessed by a member of the project team for accuracy of portraying the clinical case by a 5-point Likert scale (1=100% inaccurate, 5=100% accurate)	Accuracy score = percent of positive evaluations (i.e., evaluation ≥4)
	Detection Ratio		Clinicians receiving an SP visit to complete a "detection form" afterwards to report any suspected USP visits: 0=not suspected; 1=somehow suspected; 2=suspected with certainty)	Detection ratio = number of detected USP visit ÷ total number of USP visits (for case-specific detection ratio and all-case detection ratio, respectively). Detection ration of 10% and less are considered acceptable
Criterion Validity	Lin's Concordance correlation coefficient (rc); Kappa statistic	Validation Study	SP-completed checklist against that by a clinician based on the transcript of the visit (i.e., the clinician rating as the "reference standard")	The concordance of the quality scores based on SP-completed checklist against that based on the reference standard. rc used for continuous process quality scores, and Kappa for dichotomous diagnoses and treatment & management measures
Test-retest Reliability	Lin's Concordance correlation coefficient (rc); Kappa statistic	Validation Study	The same SP to score his own recorded encounter in a month	The concordance to be examined by rc for continuous process quality scores, fees charged (yuan) and time spent (min); and Kappa for dichotomous diagnoses and treatment & management measures
Inter-rater Reliability			Multiple SPs to complete the checklist for the same recorded transcript	

Figures

Figure 1 Seven selected seven sample provinces on the map of China with referencing countries of equivalent life expectancy in brackets

Figure 2 Sampling Procedure

Author contribution

DX conceived the project concept and developed the first protocol draft along with WG. DX, MH, and WH developed the sampling design; and MH, WH and EM wrote the section on samples and performed the sample size calculation. SS provided original data of the previous studies for the sample size estimation and calculated some summary statistics. JL and YC worked on the SP case templates. YC and XW developed the guideline for the development of the quality checklist. KH reviewed the content and edited the manuscript. HH and GC reviewed the statistical plan. SR, JP, HW, ZZ, CT, NZ, and WZ reviewed and commented on the design and methods. All co-authors participated in the revision and approved this draft.

Competing interests

There are no competing interests.

Funding

This project is funded through following competitive grants: China Medical Board (grant number: CMB GNL 16-260); National Natural Science Foundation of China (grant number: 81773446); Young Scientists Fund of the National Natural Science Foundation of China (grant number: 81402690).

Data sharing statement

We have not yet started data collection. However, the data generated from this project and the USP cases and accompanying user manuals will be made available to other researchers upon request after we complete our primary analysis.

References

- 1. A/RES/70/1 R. Transforming our world: the 2030 agenda for sustainable development 2015 [Available from: http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E accessed Feburary 17, 2018 2018.
- 2. Hanefeld J, Powell-Jackson T, Balabanova D. Understanding and measuring quality of care: dealing with complexity. *Bulletin of the World Health Organization* 2017;95(5):368.
- 3. Murray CJ, Frenk J. A WHO framework for health system performance assessment: Evidence and Information for Policy, World Health Organization 1999.
- 4. Donabedian A. The quality of care: how can it be assessed? *Archives of pathology & laboratory medicine* 1997;121(11):1145.
- 5. Pongsupap Y, Lerberghe WV. Choosing between public and private or between hospital and primary care: responsiveness, patient-centredness and prescribing patterns in outpatient consultations in Bangkok. *Tropical Medicine & International Health* 2006;11(1):81-89.
- 6. Bitton A, Ratcliffe HL, Veillard JH, et al. Primary health care as a foundation for strengthening health systems in low-and middle-income countries. *Journal of general internal medicine* 2017;32(5):566-71.
- 7. Wei X, Li H, Yang N, et al. Changes in the perceived quality of primary care in Shanghai and Shenzhen, China: a difference-in-difference analysis. *Bulletin of the World Health Organization* 2015;93(6):407-16.
- 8. Zou Y, Zhang X, Hao Y, et al. General practitioners versus other physicians in the quality of primary care: a cross-sectional study in Guangdong Province, China. *BMC family practice* 2015;16(1):134.
- 9. Feng S, Shi L, Zeng J, et al. Comparison of Primary Care Experiences in Village Clinics with Different Ownership Models in Guangdong Province, China. *PloS one* 2017;12(1):e0169241.
- 10. Wong WC, Jiang S, Ong JJ, et al. Bridging the Gaps between patients and primary care in China: a nationwide representative survey. *The Annals of Family Medicine* 2017;15(3):237-45.
- 11. Zeng L, Li Y, Zhang L, et al. Guideline use behaviours and needs of primary care practitioners in China: a cross-sectional survey. *BMJ open* 2017;7(9):e015379.
- 12. Li X, Lu J, Hu S, et al. The primary health-care system in China. *The Lancet* 2017;390(10112):2584-94.
- 13. Das J, Hammer J. Quality of primary care in low-income countries: facts and economics. *Annu Rev Econ* 2014;6(1):525-53.
- 14. Sylvia S, Shi Y, Xue H, et al. Survey using incognito standardized patients shows poor quality care in China's rural clinics. *Health policy and planning* 2014;30(3):322-33.
- 15. Berendes S, Heywood P, Oliver S, et al. Quality of private and public ambulatory

health care in low and middle income countries: systematic review of comparative studies. *PLoS medicine* 2011;8(4):e1000433.

- 16. 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 affairs* 2012;31(12):2774-84.
- 17. Das J, Gertler PJ. Variations in practice quality in five low-income countries: a conceptual overview. *Health affairs* 2007;26(3):w296-w309.
- 18. Das J, Hammer J, Leonard K. The quality of medical advice in low-income countries. *Journal of Economic Perspectives* 2008;22(2):93-114.
- 19. Coarasa J, Das J, Gummerson E, et al. A systematic tale of two differing reviews: evaluating the evidence on public and private sector quality of primary care in low and middle income countries. *Globalization and health* 2017;13(1):24.
- 20. Glassman PA, Luck J, O'Gara EM, et al. Using standardized patients to measure quality: evidence from the literature and a prospective study. *Joint Commission Journal on Quality and Patient Safety* 2000;26(11):644-53.
- 21. Leonard K, Masatu MC. Outpatient process quality evaluation and the Hawthorne Effect. *Social science & medicine* 2006;63(9):2330-40.
- 22. McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. *Journal of clinical epidemiology* 2014;67(3):267-77.
- 23. Woodward CA, McConvey GA, Neufeld V, et al. Measurement of physician performance by standardized patients: refining techniques for undetected entry in physicians' offices. *Medical care* 1985:1019-27.
- 24. Das J, Hammer J. Money for nothing: the dire straits of medical practice in Delhi, India. *Journal of Development Economics* 2007;83(1):1-36.
- 25. 钟玉杰, 王敏, 李勤. 从 10 年文献回顾分析我国标准化病人教学的发展. *中华护理杂志* 2009;44(3):259-61.
- 26. Currie J, Lin W, Zhang W. Patient knowledge and antibiotic abuse: Evidence from an audit study in China. *Journal of health economics* 2011;30(5):933-49.
- 27. 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 Medicine* 2017;14(10):e1002405.
- 28. Li L, Lin C, Guan J. Using standardized patients to evaluate hospital-based intervention outcomes. *International journal of epidemiology* 2013;43(3):897-903.
- 29. Liao J, Chen Y, Cai Y, et al. Using smartphone-based virtual patients to assess the quality of primary healthcare in rural China: protocol for a prospective multicentre study. *BMJ open* 2018;8(7):e020943.
- 30. Zhou M, Wang H, Zhu J, et al. Cause-specific mortality for 240 causes in China during 1990–2013: a systematic subnational analysis for the Global Burden of Disease Study 2013. *The Lancet* 2016;387(10015):251-72.
- 31. Das J, Hammer J. Which doctor? Combining vignettes and item response to measure clinical competence. *Journal of Development Economics* 2005;78(2):348-83.
- 32. Hambleton RK, Swaminathan H, Rogers HJ. Fundamentals of item response theory: Sage 1991.
- 33. Rethans JJ, Gorter S, Bokken L, et al. Unannounced standardised patients in real

practice: a systematic literature review. *Medical education* 2007;41(6):537-49.

- 34. Kutob RM, Bormanis J, Crago M, et al. Assessing culturally competent diabetes care with unannounced standardized patients. *Fam Med* 2013;45(6):400-08.
- 35. Fenton JJ, Kravitz RL, Jerant A, et al. Promoting patient-centered counseling to reduce use of low-value diagnostic tests: a randomized clinical trial. *JAMA internal medicine* 2016;176(2):191-97.
- 36. May L, Franks P, Jerant A, et al. Watchful waiting strategy may reduce low-value diagnostic testing. *The Journal of the American Board of Family Medicine* 2016;29(6):710-17.
- 37. ORDERING OF LABS AND TESTS: VARIATION AND CORRELATES OF VALUE-BASED CARE IN AN UNANNOUNCED STANDARDIZED PATIENT VISIT. JOURNAL OF GENERAL INTERNAL MEDICINE; 2016. SPRINGER 233 SPRING ST, NEW YORK, NY 10013 USA.
- 38. Tamblyn RM, Abrahamowicz M, Berkson L, et al. First-visit bias in the measurement of clinical competence with standardized patients. *Academic Medicine* 1992;67(10):S22-4.
- 39. Shepherd HL, Barratt A, Trevena LJ, et al. Three questions that patients can ask to improve the quality of information physicians give about treatment options: a cross-over trial. *Patient education and counseling* 2011;84(3):379-85.
- 40. Peabody JW, Luck J, Jain S, et al. Assessing the accuracy of administrative data in health information systems. *Medical care* 2004;42(11):1066-72.
- 41. Organization WH. WHO handbook for guideline development: World Health Organization 2014.
- 42. Campbell S, Braspenning J, Hutchinson A, et al. Research methods used in developing and applying quality indicators in primary care. *Qual Saf Health Care* 2002;11(4):358-64.
- 43. De Champlain AF, Margolis MJ, King A, et al. Standardized patients' accuracy in recording examinees' behaviors using checklists. *Academic Medicine* 1997;72(10):S85-7.
- 44. Vu NV, Steward DE, Marcy M. An assessment of the consistency and accuracy of standardized patients' simulations. *Academic Medicine* 1987;62(12):1000-2.
- 45. Vu NV, Marcy M, Colliver J, et al. Standardized (simulated) patients' accuracy in recording clinical performance check-list items. *Medical Education* 1992;26(2):99-104.
- 46. Maiburg BH, Rethans JJE, Van Erk IM, et al. Fielding incognito standardised patients as 'known'patients in a controlled trial in general practice. *Medical education* 2004;38(12):1229-35.
- 47. L. Gorter J-JR, Albert JJA Scherpbier, Sjef van der Linden, Marijke HM van Santen-Hoeufft, Désirée MFM van der Heijde, Harry HML Houben, Cees PM van der Vleuten, Simone. How to introduce incognito standardized patients into outpatient clinics of specialists in rheumatology. *Medical teacher* 2001;23(2):138-44.
- 48. Siminoff LA, Rogers HL, Waller AC, et al. The advantages and challenges of unannounced standardized patient methodology to assess healthcare communication. *Patient education and counseling* 2011;82(3):318-24.
- 49. Oates J, Weston WW, Jordan J. The impact of patient-centered care on outcomes. *Fam Pract* 2000;49(9):796-804.
- 50. Hudon C, Fortin M, Haggerty JL, et al. Measuring patients' perceptions of patient-centered care: a systematic review of tools for family medicine. *The Annals of Family*

Medicine 2011;9(2):155-64.

- 51. Brown J, Stewart M, Tessier S. Assessing communication between patients and doctors: a manual for scoring patient-centred communication. *London: Thames Valley Family Practice Research Unit* 1995
- 52. Ozuah PO, Reznik M. Can standardised patients reliably assess communication skills in asthma cases? *Medical education* 2007;41(11):1104-05.
- 53. Zabar S, Ark T, Gillespie C, et al. Can unannounced standardized patients assess professionalism and communication skills in the emergency department? *Academic Emergency Medicine* 2009;16(9):915-18.
- 54. Gorter S, Rethans J-J, Scherpbier A, et al. Developing Case-specific Checklists for Standardized-patient—Based Assessments in Internal Medicine: A Review of the Literature. *Academic Medicine* 2000;75(11):1130-37.
- 55. Franz CE, Epstein R, Miller KN, et al. Caught in the act? Prevalence, predictors, and consequences of physician detection of unannounced standardized patients. *Health services research* 2006;41(6):2290-302.
- 56. Swartz MH, Colliver JA, Bardes CL, et al. Validating the standardized-patient assessment administered to medical students in the New York City Consortium. *Academic medicine: journal of the Association of American Medical Colleges* 1997;72(7):619-26.
- 57. Rethans J, Drop R, Sturmans F, et al. A method for introducing standardized (simulated) patients into general practice consultations. *Br J Gen Pract* 1991;41(344):94-96.
- 58. Luck J, Peabody JW. Using standardised patients to measure physicians' practice: validation study using audio recordings. *Bmj* 2002;325(7366):679.
- 59. Shirazi M, Sadeghi M, Emami A, et al. Training and validation of standardized patients for unannounced assessment of physicians' management of depression. *Academic Psychiatry* 2011;35(6):382-87.
- 60. Lin L. A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometric, 45, 255-268, 1989.
- 61. Steichen TJ, Cox NJ. A note on the concordance correlation coefficient. *Stata J* 2002;2(2):183-89.
- 62. Lawrence I, Lin K. Assay validation using the concordance correlation coefficient. *Biometrics* 1992:599-604.
- 63. Kwiecien R, Kopp-Schneider A, Blettner M. Concordance analysis: part 16 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt International* 2011;108(30):515.
- 64. Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet* 1986;327(8476):307-10.
- 65. Rhodes K. Taking the mystery out of "mystery shopper" studies. *New England Journal of Medicine* 2011;365(6):484-86.
- 66. Rhodes KV, Miller FG. Simulated patient studies: an ethical analysis. *The Milbank Quarterly* 2012;90(4):706-24.
- 67. Weiner SJ, Schwartz A. Directly observed care: can unannounced standardized patients address a gap in performance measurement? *Journal of general internal medicine* 2014;29(8):1183-87.

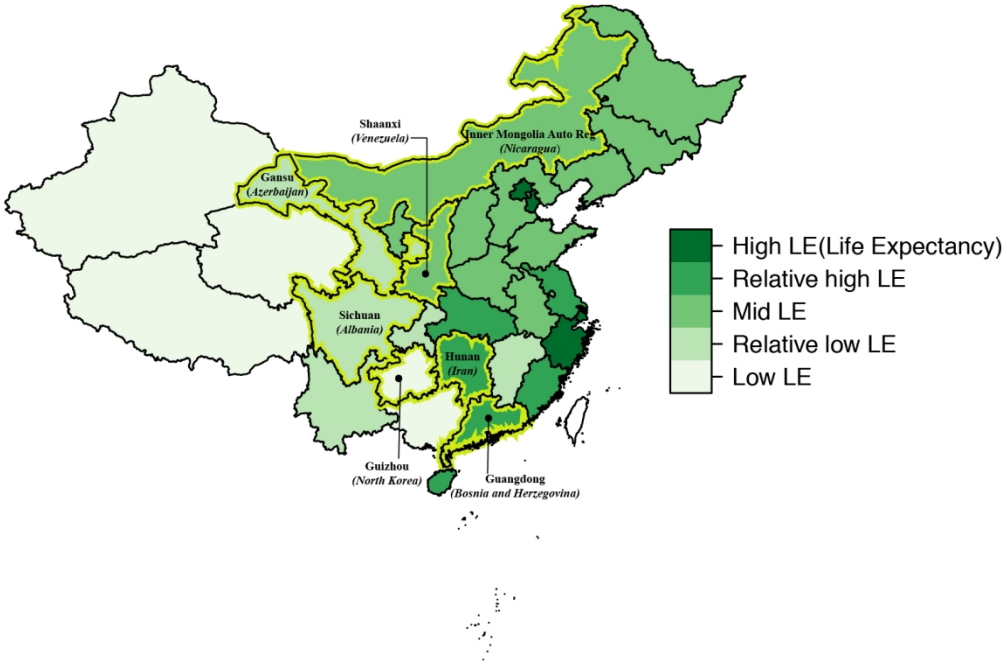

Figure 1 Selected seven sample provinces on the map of China with referencing countries of equivalent life expectancy in the bracket

147x98mm (300 x 300 DPI)

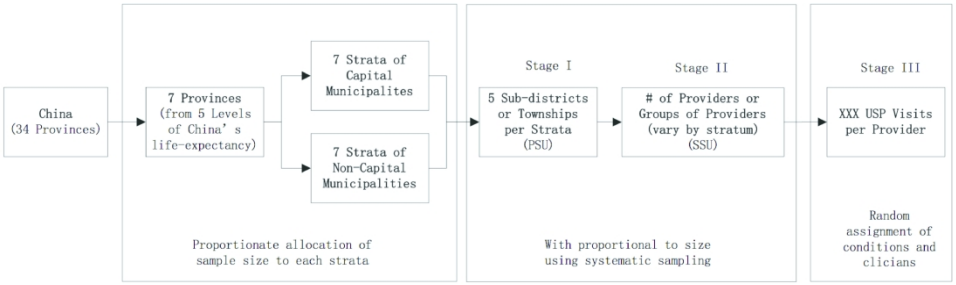

Figure 2 Sampling Procedure

397x122mm (300 x 300 DPI)

Web Appendix

Web Appendix 1 Sample Size Calculation

Compute the sampling variance of the mean: $var(\bar{y})$, based on desired coefficient of variation - 0.08.

$$var(\bar{y}) = se(\bar{y})^2 = (cv * \bar{y})^2 = (0.08 * (-0.9))^2 = 0.0052$$

Estimate number of completed interviews in need for a simple random sample(SRS):n_{srs}

$$n_{srs} = \frac{s^2}{var(\bar{y})} = \frac{4.54}{0.0052} = 875$$

Estimate design effect:

$$d_{eff} = 1 + \delta(n-1) = 1 + 0.0486 * (27-1) = 2.26$$

Multiply n_{srs} by the design effect to account for a complex survey design:

$$n_{complex} = n_{srs} * d_{eff} = 875 * 2.26 \approx 1981$$

Web Appendix 1 Evidence-based process of developing quality criteria for the SP cases

In partnership with the Lanzhou University Evidence Medicine Center, we have developed a working paper on the results of our review of the literature in quality checklist development and also our recommended protocol of developing

those checklists. We provide an abstract of that working paper below and will make available the full paper once it is fully developed.

Abstracts

Objective To explore the procedures and methods for determining the quality checklist for the most common conditions in the context of primary health care, particularly to be used for quality inspection by unannounced standardized patients.

Methods We conducted a systematic search of literature in the subject matter, while adopting the WHO handbook for guideline development. **Results** A total of 14 related articles were included and the methodological aspects were evaluated. Based on this review, we propose five key steps in the checklist development: (1) Forming a multidisciplinary team; (2) Reviewing, evaluating and selecting relevant literature based on evidence-based medicine quality of evidence principles; (3) Extracting essential quality information to form a pool of quality items; (4) using expert consensus to select candidate quality checklist items from the pool; (5) pre-testing to determine the final items. **Discussion** We recommend a checklist development method based on evidence-based method augmented by expert opinions through a

multidisciplinary group discussion. The selection of the items on the checklist will consider their importance and feasibility. Our proposed methods can be mainly applied to common conditions seen in the primary care settings and may not be applied to more complex conditions.

BMJ Open

Assessing the Quality of Primary Health Care in 7 Chinese Provinces with Unannounced Standardized Patients: Protocol of a Cross-sectional Survey

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-023997.R2
Article Type:	Protocol
Date Submitted by the Author:	01-Nov-2018
Complete List of Authors:	Xu, Dong; Sun Yat-sen University, Sun Yat-sen Global Health Institute (SGHI),School of Public Health and Institute of State Governance Hu, Mengyao; University of Michigan, Survey Research Center, Institute for Social Research He, Wenjun; Sun Yat-sen University, Department of Biostatistics and Epidemiology, School of Public Health Liao, Jing; Sun Yat-sen University, Sun Yat-sen Global Health Institute (SGHI), School of Public Health and Institute of State Governance Cai, Yiyuan; Guizhou Medical University, School of Public Health; Sun Yat-Sen University, School of Public Health Sylvia, Sean ; University of North Carolina at Chapel Hill, Department of Health Policy and Management, Gillings School of Global Public Health Hanson, Kara; London School of Hygiene & Tropical Medicine, Department of Global Health and Development, Faculty of Public Health and Policy Chen, Yao-Long; Lanzhou University, Evidence-Based Medicine Center, School of Basic Medical Sciences Pan, Jay; Sichuan University, West China School of Public Health Zhou, Zhongliang; Xi'an Jiaotong University, School of Public Policy and Administration Zhang, Nan; Inner Mongolia Medical University, Department of Health Management, School of Health Management Tang, Chengxiang; Guangzhou University, School of Public Administration Wang, Xiaohui; Lanzhou University, Evidence Based Medicine Center, School of Basic Medical Sciences Rozelle, Scott; Stanford University, Freeman Spogli Institute for International Studies He, Hua; Tulane University, Department of Epidemiology, School of Public Health and Tropical Medicine Wang, Hong; Bill & Melinda Gates Foundation, Health Economics, Financing & Systems Chan, Gary; University of Washington, Department of Biostatistics Melipillán, Edmundo; University of Michigan, Survey Research Center, Institute for Social Research Zhou, Wei; Central South University, Xiangya school of public health, Gong, Wenjie; Central South University, School of Public Health; University of Rochester Medical Center, Department of Psychi

Primary Subject Heading:	Health services research
Secondary Subject Heading:	Public health
Keywords:	standardized patients, unannounced standardized patients, quality of primary health care, patient-centered care

Assessing the Quality of Primary Health Care in 7 Chinese Provinces with Unannounced Standardized Patients: Protocol of a Cross-sectional Survey

Dong Roman Xu,[1] Mengyao Hu,[2] Wenjun He,[3] Jing Liao,[1] Yiyuan Cai,[4,3,1] Sean Sylvia,[5] Kara Hanson,[6] Yaolong Chen,[7] Jay Pan,[8] Zhongliang Zhou,[9] Nan Zhan,[10] Chengxiang Tang,[11] Xiaohui Wang,[7] Scott Rozelle,[12] Hua He,[13] Hong Wang,[14] Gary Chan,[15] Edmundo Roberto Melipillán,[2] Wei Zhou,[16] Wenjie Gong[17*]

[1] Sun Yat-sen Global Health Institute (SGHI), School of Public Health and Institute of State Governance, Sun Yat-sen University, No.74 Zhongshan 2nd Road, Guangzhou, P.R. China, 510080

[2] Survey Research Center, Institute for Social Research, University of Michigan, 426 Thompson Street, Ann Arbor, MI 48104, USA

[3] Department of Biostatistics and Epidemiology, School of Public Health, Sun Yat-sen University, No.74 Zhongshan 2nd Road, Guangzhou, P.R. China, 510080

[4] School of Public Health, Guizhou Medical University, Guizhou, P.R. China, 550025

[5] Department of Health Policy and Management, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, CB#7411, Chapel Hill, NC 27599, USA

[6] Department of Global Health and Development, Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine, WC1H 9SH, London, UK

[7] Evidence Based Medicine Center, School of Basic Medical Sciences, Lanzhou University, No. 199 Donggang West Road, Lanzhou, Gansu Province, 730000 P.R. China

[8] West China School of Public Health, Sichuan University, No. 17, Ren Min Nan Road, Chengdu, P.R. China

[9] School of Public Policy and Administration, Xi'an Jiaotong University, No. 28 Xianning West Road, Xi'an, Shaanxi, 710049, P.R. China.

[10] Department of Health Management, School of Health Management, Inner Mongolia Medical University, Jinshan Development Zone, Hohhot, Inner Mongolia, P.R. China, 010110

[11] School of Public Administration, Guangzhou University, Guangzhou, Guangdong, 510320, P.R. China.

[12] Freeman Spogli Institute for International Studies, Stanford University, Encina Hall East, E407, Stanford, CA 94305-6055, USA

[13] Department of Epidemiology, School of Public Health and Tropical Medicine, Tulane University, 1440 Canal Street, New Orleans, LA 70112, USA

[14] Health Economics, Financing & Systems, Bill & Melinda Gates Foundation, PO Box 23350, Seattle WA, 98102, USA

[15] Department of Biostatistics, University of Washington, 1705 NE Pacific Street, #F600, Seattle, WA 98195, USA

[16] Hospital Administration Institute, Xiangya Hospital, Central South University, No. 87 Xiangya Road, Changsha, Hunan, P.R. China

[17] Xiangya school of public health, Central South University, Changsha, Hunan, China

Corresponding author: Wenjie Gong, gongwenjie@csu.edu.cn, Xiangya School of Public Health, Central South University, Xiangya school of public health, Central South University, P.R. China*

Abstract

Introduction: Primary health care (PHC) serves as the cornerstone for the attainment of universal health coverage (UHC). Efforts to promote UHC should focus not only on the expansion of access but also on healthcare quality. However, robust quality evidence has remained scarce in China. Common quality assessment methods such as chart abstraction, patient rating, and clinical vignette use indirect information that may not represent real practice. This study will send standardized patients (SP or healthy person trained to consistently simulate the medical history, physical symptoms, and emotional characteristics of a real patient) unannounced to PHC providers to collect quality information and represent real practice.

Methods and Analysis: 1981 SP-clinician visits will be made to a random sample of PHC providers across seven provinces in China. SP cases will be developed for ten tracer conditions in PHC. Each case will include a standard script for the SP to use and a quality checklist that the SP will complete after the clinical visit to indicate diagnostic and treatment activities performed by the clinician. Patient-centeredness will be assessed according to the Patient Perception of Patient-centeredness (PPPC) rating scale by the SP. SP cases and the checklist will be developed through a standard protocol and assessed for content, face and criterion validity and test-retest and inter-rater reliability before its full use. Various descriptive analyses will be performed for the survey results such as a tabulation of quality scores across geographies and provider types.

Ethics and dissemination: This study has been reviewed and approved by the Institutional Review Board of the School of Public Health of Sun Yat-sen University (#SYSU 2017-011). Results will be actively disseminated through print and social media, and SP tools will be made available for other researchers.

Keywords: standardized patients; unannounced standardized patients; quality of primary health care; patient-centered care

Strengths:

- We will assess the quality of care with a random sample of primary health care providers in seven provinces in China.
- We will use unannounced standardized patients (USPs) - the "gold standard" of quality assessment.
- Both technical quality and patient-centeredness will be assessed.

Limitations:

- USPs are not suitable for certain health conditions.
- The seven provinces are not randomly selected, although we intend them to represent different health development conditions (by using life expectancy as the proxy) of China's provinces.

Assessing the Quality of Primary Health Care in 7 Chinese Provinces with Unannounced Standardized Patients: Protocol of a Cross-sectional Survey

Background

In 2015, all 191 UN member states adopted the Sustainable Development Goals (SDGs), aiming to achieve universal health coverage (UHC) – the access to high-quality health care services without incurring financial hardship – by 2030.[1] As previous literature emphasized, efforts to promote UHC should focus not only on the expansion of access but also on healthcare quality.[2] Healthcare quality is variously defined by the WHO as the "responsiveness" of the health care system to meet desired health outcomes,[3] as the instrumental goals on structure, process, and outcome in the Donabedian Framework,[4] and as the six comprehensive aims (effectiveness, efficiency, equity, patient-centeredness, safety, and timeliness) put forth by the Institute of Medicine (IOM).[5] In this study, we adopt the IOM definition of quality.

Primary health care (PHC) serves as the cornerstone for the attainment of UHC.[6] China's latest round of health care reform since 2009 has invested heavily in strengthening PHC. There have been some efforts to assess the quality of PHC in China: patients were interviewed with a Primary Care Assessment Tool (PCAT) questionnaire in Guangdong, Shanghai, and Hong Kong;[7-9] comprehensiveness of the service provision was used as a proxy for quality through clinician interviewing;[10] and PHC clinicians' adherence to clinical guidelines was assessed with a self-report questionnaire.[11] However, assessment of the quality of PHC has largely remained scanty in China, and the assessment tools are indirect and prone to bias.[12] A number of studies have found the quality of PHC to be low in other low and middle income countries (LMICs)[6 13-18], where robust evidence remains scarce.[19] Commonly-used methods of measuring technical

quality of care include chart abstraction, patient rating of care, and using a clinical vignette to test clinician knowledge. Those methods use indirect information that may not represent real practice. This study instead will use unannounced standardized patients (USPs) to measure the quality of real practice. The Standardized Patient (SP) is a healthy person (or occasionally a real patient) trained to consistently simulate the medical history, physical symptoms, and emotional characteristics of a real patient. The SP, particularly when their visit is unannounced, has several reported advantages: (1) reliability in measurement and cross-provider comparison because the same patient is presented to all providers, (2) elimination of the Hawthorne effect (i.e., that the study itself may change doctors' behavior) due to the nature of disguised and unannounced visit by SPs,[20-22] and (3) reduced recall bias.[23 24]

Despite these advantages, the application of SP in China has been concentrated mainly in the area of medical education.[25] An ongoing systematic review identified four papers only on using the SP for quality assessment in China,[14 26-28], and 44 in other LMICs. Those projects, often based on a small convenience sample, tended to target a limited number of conditions (approximately 70% on family planning services, childhood infectious diseases, sexually transmitted infections, and respiratory tract infections). In this study, we intend to assess the quality of PHC with a probability sample of PHC visits in seven Chinese provinces, using USPs for ten commonly seen conditions in the PHC setting. The project has involved 20 universities across 19 provinces in China as well as researchers from Nepal, US, and UK in a USP Network (https://www.researchgate.net/project/Unannounced-Standardized-Patient-USP-and-Virtual-Patient-VP-to-Measure-Quality-of-Primary-Care). The USP resources will be pooled and shared widely within the network first and then with the general public. This study is the first of a series of studies to be based on quality data collected using USPs. The primary purpose of this study is

to collect and present descriptive data on the quality of China's PHC. We are developing separate protocols for the various hypothesis-driven studies, which will be available elsewhere and from our Network website.[29]

Methods

Survey Design

The purpose of the sample design is to create a representative sample of China's primary health care (PHC) providers so that healthcare quality can be assessed based on USP visits to those providers.

Survey Population/Frame

We considered creating a nationally representative probability sample, but at this stage, we have selected seven provinces to "represent" China due to feasibility considerations. These provinces represent five levels of average life expectancies across China's provinces (Figure 1), which are similar to those of five countries with low to high income levels.[30] We intend to create a probability sample that represents primary health care in these seven provinces. For the survey population, we intend to include (1) licensed physicians and licensed assistant physicians at community/township health centers/stations and urban health stations, (2) certified village doctors (a terminology in China that refers to village clinicians who ha written content ve village-level practice privilege even without a medical license) and village sanitarians (referring to un-certified village doctors who are supposed to work under the supervision of the village doctor) at village clinics; and (3) clinicians with a license notation for general practice, internal medicine, obstetrics/gynecology, and pediatrics at the level I and level II hospitals and the maternal and child care centers. We exclude level 3 hospitals, which provide more specialized

care, and specialty hospitals. Clinician meeting those criteria will constitute the "sampling frame".

Sampling Procedures

The sample will be selected using a multi-stage, clustered sample design covering all eligible clinicians in the seven provinces (Figure 2). In the first stage, stratification will be based on the provinces. Due to the high number of visits in the seven capital cities, we will sample each capital city. Each province is thus divided into two strata consisting of the provincial capital city and other prefecture-level municipalities, leading to 14 strata in total. We will use proportionate allocation (in terms of the number of eligible clinicians) of sample size for each stratum. For each stratum, five rural townships or urban sub-districts (the primary sampling unit/PSU) will be selected using probability proportional to size (PPS). In the second stage, for each PSU, PHC facilities as previously-defined (Secondary Sampling Unit/SSU) will be selected using PPS systematic sampling. Neighboring village clinics will be grouped as an SSU. The number of SSUs for each stratum will vary depending on the size of the stratum – for example, more SSUs will be selected in strata with more PHC clinicians. In the final stage, a fixed number of USP visits will be made to each selected facility or the group of facilities in the case of village clinics. The exact number of visits will be determined once we obtain and examine our sampling frame.

Sample Size Calculation

The sample size was calculated for the primary purpose of the standard descriptive survey analysis of this survey. The sample size (power) calculation for the other related hypothesis related studies will be described in separate study protocols. The primary statistic of interest in this survey is a latent variable measuring clinician's quality, constructed using the 2-parameter logistic item response theory (IRT) model.[31 32] The model was based on a list of

quality checklist items measuring whether doctors asked recommended questions and whether they performed recommended exams (see the *Scoring Method* section below). Survey sample size was calculated based on the desired level of relative precision (coefficient of variation, CV), an estimate for the population element variance for the variable of interest (s^2) from previous study and design effect ($deff$). In this study, our desired level of relative precision (CV) is 0.08. s^2 was estimated to be 4.54, based on Sylvia et al's work on the USP-assessed quality of PHC in three Chinese provinces.[14 27] Design effect is the variance inflation due to cluster sampling. This figure was calculated based on intra-class correlation (ICC) (describing the level of homogeneity of the units in a cluster) and cluster sample size: $deff = 1 + \delta(n - 1)$, where δ is the intra-class correlation (ICC) and n is the average size of the cluster. The ICC of 0.0486 was also estimated from Sylvia et al's work, which was 0.0486. Our estimated average cluster size is 27 clinician-SP encounters per PSU. Accordingly, we calculated the total required sample size to be 1981 clinician-SP encounters. The steps taken to calculate sample size can be found in Web Appendix 1

USP Case Development and Implementation

The development process of a USP case is based on our extensive literature review,[20 33] as well as our own USP experiences in Shaanxi Province, China.[14 27] We are concurrently developing smartphone-based virtual standardized patients (VPs) (details described elsewhere). The two projects will share almost identical case scenarios and quality criteria.

Case Selection

Our purpose is to select ten health problems as tracer conditions for PHC in China. Ideally our selected cases should (1) be highly prevalent in PHC settings, (2) carry challenging features in different aspects of PHC (e.g., some cases focus on curative care while others on

prevention, disease management, culturally-sensitive care,[34] or misuse of low value tests[35-37]), (3) not involve invasive and painful procedures, (4) not require physical signs that cannot be simulated (e.g., jaundice can be simulated with make-up, but heart murmurs cannot.[23]). We created a list of the top 30 conditions commonly seen in PHC in China, combining the results of two national surveys on PHC.[12] A panel of physicians and public health and health system researchers then applied the principles above and selected a dozen PHC problems for USP development (Table 1). Ten final conditions will be selected from this list.

Development Team

We have created an overall development team and 10 case-specific development teams. Each team includes case-specific specialists, general practitioners, and public health and health system researchers (Web Appendix 2). A third overall panel consisting of primary care providers at the village, township and community levels will review all cases for contextual appropriateness in primary care settings. In developing the case, we will follow several principles: (1) limiting case scenarios to those that require definitive clinician action on the first visit to minimize potential "first-visit bias",[38] (2) focusing on the presentation of symptoms for which evidence is well-established for diagnosis and management, (3) deriving some content of the cases from the actual case history of relevant patient files in real practice.[23]

Case Description

The case description describes the relevant clinical roles and psycho-social biographies of the SP.[39] We used a structured description of the cases as follows:

1. Social and demographic profile: (1) socioeconomic information: name, gender, age, ethnicity, education, occupation, family structure (e.g., Married and have two children but live alone), dress style (e.g., dressed in jeans, work boots and a well-worn but neat

sweater), health insurance or other social program participation; (2) personality that may influence interaction with the clinician (e.g., non-proactive and introverted); (3) lifestyle relevant to health (e.g., smoked one pack of cigarette since age 18, like fried pork but also eat much fruit, exercise regularly, watch TV a lot in spare time, play mahjong with friends, visit children every week)

2. Medical history: (1) disease information: severity of the condition (e.g., mild or severe depression), duration of the condition (the first onset? Previously diagnosed/existing (how long)?), comorbidity (any other physical and/or psychological problems?), (2) reason for seeking care for this specific visit (e.g., was feeling down for two months but depression worsened last week), (3) treatment/management already or currently received (e.g., a diabetic "patient" took metoprolol for hypertension but does not monitor his glucose / watch his diet/weight).
3. Physical examination: symptoms the SP will (and will not) portray (e.g., reduced appetite, but not showing agitation), and medical signs the SP has or does not have (e.g., heart murmur).
4. Laboratory and imaging: laboratory and imaging that a clinician may prescribe for the SP. The laboratory and imaging results of the SP may be generated from those of real typical patients.
5. Diagnosis: the correct diagnosis that the clinician should make based on the information presented by the SP.
6. Treatment and management: the decision of the clinician on what medications, procedures, advice, or referral will be given at the end of the consultation.

Script

Corresponding to the six components of the afore-mentioned case description, we will develop a detailed script for the SPs to use in their PHC visit with the clinician. The script ideally should cover all possible questions a clinician may ask as well as the SP's answers during the clinical interaction. Panels of clinicians will be consulted to collect relevant questions that will guide the development of the script. The script will continue to add new questions asked by the clinicians on the SP-clinician interaction. The script will have five sections: (1) an opening: spontaneous information given to the clinician at the start (e.g., Doctor, I have had a headache for two days), (2) the information given only on request, (3) the information for the SP to volunteer even if not asked, (4) the language to insist on a diagnosis if not given, and (5) an ending.[14 20 40]

Quality Checklist

The checklist consists of explicit quality criteria for gathering data on patient history, physical examination, laboratory/imaging, diagnosis and treatment.[14 33] Based on our comprehensive review of 14 articles of literature and evidence-based clinical guideline development methodology,[41] we have established a guiding principle and standard protocol for checklist development. Our process will (1) be evidence-based and augmented by expert opinion,[42] (2) follow a systematic procedure to gather, evaluate and select evidence and criteria, (3) select criteria related to clinician actions that the SP can easily evaluate,[43] (4) keep the number of checklist items under 30 to include high-priority criteria only so that the SP can reliably recall clinician behaviour[43-45]. The details of our checklist development protocol will be described in a separate paper, and key messages are summarized in Web Appendix 2.

Selecting and Training SPs

We will advertise on social media to recruit SPs. The candidate must be in stable health without confounding symptoms; should match the real patients in age, sex, and physical features; are willing to allow the examinations appropriate to their condition; have the intellectual maturity to present the behavior of the actual patient and complete the checklist.[23 46 47] We may consider recruiting real patients with stable conditions to portray the cases not subject to simulation.[23] The training of the SP will aim at portraying the signs, symptoms, and presentations, completing the checklist, and minimizing detection by the provider.[20] The week-long training will have three stages: classroom instruction, a dress rehearsal, and two field tests.[23 47 48] Each case will have three SPs who will be trained according to a standardized training manual that will be developed to guide the training and appraisal of the SPs.

Fielding and Implementing SPs

A disguise plan will be developed for each case to minimize physician detection of the SP status (e.g., convincing excuse for seeking care where they do not usually reside). In the pilot (instrument validation) phase, consent will be sought for audio recording (see below); in these cases, fieldwork will start only three to four weeks after consent is obtained. We will provide each SP with a calamity letter, explaining the project in case of their identity being exposed.

After the facilities are selected, and the number of visits per facility is determined, each of the planned visits will be given a unique identifier (e.g., facility A-1, facility A-2, facility B-1), which will then be randomly ordered to form a random sequence numbered from 1 to 1981 consecutively. One of the ten SP cases will be randomly assigned to each number on this random sequence. The seven SPs per case will be dispatched to the seven provinces concurrently, one SP per province. If multiple clinicians are available in that facility at the time

of a particular SP visit (PHC visits in China do not require appointments), the field coordinator will randomly select a clinician by drawing lots onsite. Each SP is expected to make a total of approximately 30 visits. We plan to complete those SP visits over a three-month time span.

In a separate but related study, a week after the visit of the SP, the same clinician will perform the same consultation but with a standardized virtual patient on a smartphone.[29] We will use this opportunity to administer a detection questionnaire to the clinician, asking whether they suspect they had any visit from an SP over the past week. The detected cases will be treated as missing data in the data analysis.

Variables

Outcome Variables

We will collect a variety of quality of care information and other related explanatory variables. The IOM quality framework (effective, safe, patient-centered, timely, efficient, and equitable) will be used for quality evaluation (Table 2). E**ffectiveness** (avoiding underuse and misuse) and **safety** (avoiding harm), traditional technical goals of quality of care, will be evaluated through the yes/no checklist discussed above (Web Appendix 2). **Patient-centeredness** (respectful of and responsive to individual preferences) will be assessed by the Patient Perception of Patient-centeredness (PPPC) rating scale.[49-51] Using a 4-point Likert scale, PPPC evaluates three dimensions of patient-centeredness: exploring the disease and illness experiences, understanding the whole person, finding common ground.[49] Prior studies have demonstrated the validity of SPs rating clinician communications.[52 53] A separate study will be conducted to test the validity of PPPC scale. **Timeliness** will be assessed by analyzing opening hours, waiting time, and consultation time.[5] **Efficiency** (avoiding waste) will be measured by costs of care of the SP-clinician encounter. **Equity of care** (no variance in quality because of

personal characteristics) will be assessed through a separate but related study in a randomized cross-over trial.

Scoring Method

Technical quality of care will be reflected by a continuous score ranging from 0-1. We will evaluate further whether to classify checklist items in four categories (essential, important, indicated, and non-contributory) with corresponding numeric weights (3, 2, 1, and 0).[54] Two scoring methods will be used: 1) the simple scoring method will use the formula of items performed divided by the total number of items on the checklist for the process scores, whereas 2) the complex method will use an algorism based on item-response-theory (IRT).[31] Using the IRT model approach, we can obtain a latent performance score for each doctor, which has been corrected for measurement error. An ordinal variable will be used for diagnosis and management plans (Table 2), while patient-centeredness will follow the scoring methods of PPPC (possible range of score from 1-4).[51]

Other Variables

We will collect additional information on the predictors, confounders, and effect modifiers to the outcomes in the planned hypothesis testing of the related studies to this survey. The information will include qualifications of the clinician and facility information (environment, amenity, size, location, ownership type, and so forth).

Analytical Methods

USP Validation

USP validation will be based on a convenience sample of clinicians not included in our final survey sample in the project training and pilot phase. Those SP-clinician interactions in the pilot will be audio recorded and transcribed. **Validity** is the extent to which an instrument

measures what it is supposed to measure. We will assess content, face, and criterion validity of the cases. Content validity will be assessed by an expert panel who will use a 4-point Likert scale to evaluate the appropriateness of the written contents of the cases that will include the scenario, scripts and checklists. For the checklist, they will be instructed to check the appropriateness against the published clinical guidelines. The face validity of the SP assessment depends on (1) the SP remaining undetected (detection ratio reported to be 5%-10%[55]), and (2) authentically and consistently portraying the clinical features of the case. We will send the participating clinician in the pilot a "detection form" to report their degrees of suspicion of any SP visit.[46] The authenticity of the SP presentation will be evaluated by checking the transcribed recording to discover whether a key piece of information was divulged by the SP when appropriately prompted, not divulged when prompted, or volunteered when not prompted. Criterion validity will be assessed through the agreement of the SP-completed checklist against that completed by a clinician based on the transcript of the visit (i.e., the clinician rating as the "gold standard").[56-59] Checklist items which depend on visual observation will be excluded. **Reliability** examines the level of consistency of the repeated measurements. The inter-rater reliability of two SPs on the same condition and context will be assessed with two SPs completing the checklist for the same recorded transcript. Test-retest reliability will be analyzed by the concordance of assessment results of the same SP to score his or her own recorded encounter a month later).[57] The agreement will be analyzed with Lin's concordance correlation coefficient (r_c)[60]. r_c indicates how closely pairs of observation fell on a 45° line (the perfect concordance line) through the origin in addition to their correlation.[60-62] Bland-Altman plot will be used to visualize the concordance.[63 64] Table 3 summarizes our methods of validation.

Survey Analysis

We will focus on descriptive analysis to present the quality of PHC in the seven provinces. Hypothesis-driven analyses will be described in separate study protocols. For descriptive analysis, we will first present clinician and facility profiles in tables for all seven provinces and by each province. The clinician profile will include socio-demographic information (age, gender, and ethnicity), professional qualification (general and medical education, licensure, and professional ranks), and service information (volume of visits, number of support personnel). The Facility profile will include information on operation and management (years in operation, ownership types, accreditation, level of hospitals, affiliation with medical universities, revenue, health insurance contracting, payment methods), clinical services (annual number of inpatient and outpatient visits, number of clinical departments), personnel (number of physicians, nurses, and attrition ratio), and equipment. Second, we will tabulate results of overall quality and sub-domains across administrative regions and provider types. Third, we will map out the locations of the facilities along with their quality scores with geospatial analytical tools. Finally, a T-test/Wilcoxon test or Chi-square test will be employed to compare quality differences between public versus private providers, primary care clinics/centers versus hospital outpatient services, care in rural versus urban areas, and across different conditions, clinician educational levels, and payment mechanisms.

Related Studies

This study protocol mainly deals with the descriptive analysis and presentation of the data to be collected by the USPs. Using the USP survey data, we have planned several related studies that will be covered by separate study protocols with details on the background, theoretical framework, and analytical methods. To summarize those related studies, we will

assess (1) the effect of ownership types of the PHC providers (i.e., private versus public) on quality of PHC (study protocol under revision), (2) the know-do gap between the assessment results by a smartphone-based virtual standardized patient and USP (protocol already published),[29] (3) the effect of using smartphone-based virtual patient in improving clinician performance, (4) the effect of types of insurance carried by a patient on quality of care, (5) the impact of gatekeeping by primary care providers on quality of TB care – a mathematical modeling study, and (6) clinician skills in handling low-value or harmful patient requested services, particularly antibiotics and some processed traditional Chinese medicine.

Ethics and Dissemination

This study has received ethical approval from the institutional review board (IRB) of the Sun Yat-sen University School of Public Health with a waiver of informed consent from each participating clinician. USP studies do not necessarily require the consent if they meet certain conditions.[65 66] Our waiver has been granted for the following reasons: (1) our study serves important public good while requiring informed consent may lead to considerable selection bias and greater risk for the detection of the SP; (2) this study does not intend to entrap or reveal identities of any institution or individual and all analyses will be conducted at the broader health system level (after data cleaning all individual identifiers will be destroyed); (3) no audio-visuals will be recorded during the SP-clinician encounter (however, in the pilot stage, we will seek informed consent from participating clinicians as we will use a disguised recording for the validation purposes). The study results will be widely distributed in the form of scientific papers and policy briefs. The data generated from this project and the USP cases and accompanying user manuals will be made available to other researchers upon request after we complete our primary analysis.

Patient and public involvement

We selected the conditions for the USP partly based on results from surveys on common conditions in the context of primary health care as reported by the patient. The USP cases will also be reviewed by a panel that includes patients. The results of the studies will be widely distributed in scientific reports as well as social media to benefit policy-makers, clinicians, and patients.

Discussion

In this study, we will develop, validate and implement methods of assessing the quality of PHC using USPs. Compared to existing studies using USPs,[33] this proposed study has several distinctive features. First, we will establish a large probability random sample so that representative estimates of PHC quality can be achieved in the chosen seven provinces in China. Second, unlike previous studies,[14 27] we include not only village clinics, township health centers, and community health centers but also county hospitals and other level I and level II hospitals in the study. The latter were not officially designated as PHC facilities in China but provided a substantial amount of PHCs. Third, 10 SP cases will be developed through a standardized process using the same template and methodology and will represent common conditions in PHC, while past studies often used two to three conditions.[33] Fourth, an evidence-based systematic method will guide checklist development. In a review, only 12 out of 29 SP articles reported the procedures of checklist development and many checklists were developed by expert consensus only.[54] Fifth, in addition to using the checklist to evaluate technical quality of care as performed in most other USP studies, we will assess patient-centeredness with a global rating scales. Sixth, we have planned a series of related studies to address the quality of PHC in a concerted effort. Most noteworthy, we are developing ten identical conditions as smartphone-

based virtual patients to assess the competency of PHC providers. Seventh, we used the same case for all levels of providers from village doctors to township health centers to county hospitals, but quality checklists for process, diagnosis and treatment will be tailored to fit the expected roles and responsibilities of the different providers. Finally, we have secured the understanding and cooperation of the provincial health authorities.

We note two particular issues. In high-income settings, logistical arrangements for the SP is complex. A significant challenge is to introduce the SP into medical practice.[23 47 48] However, in China and many other LMICs, enrollment with a clinician is not required, and a walk-in visit to clinicians without an appointment is commonplace. However, village doctors usually know their patients well. For these areas, the SPs in other studies pretended to be tourists or friends visiting the families in the village. We will try other pretenses such as a temporary poverty-relief worker who has just arrived in a nearby village. Those poverty-relief workers are common in remote rural areas in China. On a second issue, assessing quality with USP was reported to incur high cost in the developed countries (estimated to be USD 350-400 per visit).[53 67] We expect the cost in China to be considerably less due to the lower labor cost. We will collect detailed cost information to inform the future application of the USP.

The study has several potential limitations. Most important, even though the assessment of SP is considered the gold standard for measuring clinician performance, and in this study we have further expanded the use of SPs to evaluate other elements of quality in the IOM framework such as patient-centeredness, timeliness, and efficiency, we recognize that those quality of care elements are still largely clinician-related, and other important quality aspects such as the quality of laboratory testing cannot be assessed by our SPs. In addition, the USP method has several technical challenges. If healthy people are used to simulate the patient, it is difficult to achieve

complete alignment of patient presentation of signs and symptoms (for instance, it is difficult to fake a sore throat). There are also challenges of obtaining fake laboratory-test results that may be necessary for the diagnosis. Some clinical roles that require the SP to go through invasive investigation may also pose a problem. We will experiment with a real patient in stable conditions to resolve some of those challenges. Next, our judgment of the clinical quality through the first and only visit with the SP may lead to "first-visit bias".[38] The quality of care provided by a clinician who spreads his or her diagnosis and management over several visits may be underestimated. We try to minimize this bias by designing cases that require a definitive decision on the first visit. Last, even though we intend to select ten tracer conditions in the context of PHC, we still need to be cautious in generalizing the findings to the overall quality of PHC.

In conclusion, this proposed study may produce a set of validated tools for the assessment of the quality of PHC using USP and apply it to obtain valuable quality of care information on primary health care in China.

Tables

Table 1 Selected Candidate Conditions

Conditions	Special Focus Areas											
	Chronic Disease Management	Public Health Delivery	Mental Health	Maternal & Child Care	Preventative Care	Referral	Patient-centered care	Older Adults	Low value diagnostic	Antibiotics	Process Traditional Chinese Drug	Injury
Common cold (flu season)					✖					✖	✖	
Hypertension	✖										✖	
T2DM	✖						✖	✖			✖	
Gastritis							✖					
Child diarrhea				✖						✖		
Low back pain (patient requesting low-value test)							✖		✖			
Depression (Maternal care)			✖	✖		✖	✖					
Angina (heavy smoker)					✖	✖	✖				✖	
Headache											✖	
Fall					✖		✖	✖				✖
Asthma												
Tuberculosis		✖			✖	✖						

Table 2 Variables

	Variable name	Type	Coding	Source
1. Effectiveness & Safety				
1.1	% of recommended questions asked	continuous	0-1	SP checklist
1.2	% of recommended exams performed	continuous	0-1	SP checklist
1.3	Diagnosis quality	ordinal	0: incorrect 1:partially correct 2:correct	SP checklist
1.4	Treatment quality	ordinal	0: incorrect 1:partially correct 2:correct	SP checklist
2. Patient-centeredness				
2.1	Patient perception of patient-centeredness	continuous	0-1	PPPC
2.2	Choice of provider	dichotomous	0: no 1: yes	SP checklist
2.3	Ease of navigation in facility	ordinal	0: difficult 1: median 2: easy	SP rating
3. Timeliness				
3.1	Opening hours	continuous	hours	SP checklist
3.2	Wait time	continuous	minutes	SP checklist
3.3	Consultation time	continuous	minutes	SP checklist
4. Efficiency				
4.1	Total cost	continuous	RMB	SP checklist
4.2	Medication cost	continuous	RMB	SP checklist
4.3	Laboratory/imaging cost	continuous	RMB	SP checklist
5. Equity				
5.1	To be analyzed in a separate cross-over trial			

Table 3 Methods of Validation for the USP cases

Domain	Indicator	Data Collection		Statistical Analysis
		Phase	Method	
Content Validity	Content Validity Index (CVI)	USP case review	Expert panel review of SP cases, measured by a 4-point Likert scale (1=lowest, 4=highest)	CVI for SP case and for specific USP, where CVI = number of raters giving a rating of 3 or 4÷total number of raters
Face Validity	Authenticity of SP role-play	Validation Study	Transcripts of the recording of the USP-clinician encounter to be assessed by a member of the project team for accuracy of portraying the clinical case by a 5-point Likert scale (1=100% inaccurate, 5=100% accurate)	Accuracy score = percent of positive evaluations (i.e., evaluation ≥4)
	Detection Ratio		Clinicians receiving an SP visit to complete a "detection form" afterwards to report any suspected USP visits: 0=not suspected; 1=somehow suspected; 2=suspected with certainty)	Detection ratio = number of detected USP visit ÷ total number of USP visits (for case-specific detection ratio and all-case detection ratio, respectively). Detection ration of 10% and less are considered acceptable
Criterion Validity	Lin's Concordance correlation coefficient (rc); Kappa statistic	Validation Study	SP-completed checklist against that by a clinician based on the transcript of the visit (i.e., the clinician rating as the "reference standard")	The concordance of the quality scores based on SP-completed checklist against that based on the reference standard. rc used for continuous process quality scores, and Kappa for dichotomous diagnoses and treatment & management measures
Test-retest Reliability	Lin's Concordance correlation coefficient (rc); Kappa statistic	Validation Study	The same SP to score his own recorded encounter in a month	The concordance to be examined by rc for continuous process quality scores, fees charged (yuan) and time spent (min); and Kappa for dichotomous diagnoses and treatment & management measures
Inter-rater Reliability			Multiple SPs to complete the checklist for the same recorded transcript	

Figures

Figure 1 Seven selected seven sample provinces on the map of China with referencing countries of equivalent life expectancy in brackets

Source: The figure is adapted from the one from the paper by Liao, Jing, et al. "Using smartphone-based virtual patients to assess the quality of primary healthcare in rural China: protocol for a prospective multicentre study." BMJ open 8.7 (2018): e020943. Permission to use has been obtained.

Figure 2 Sampling Procedure

Author contribution

DX conceived the project concept and developed the first protocol draft along with WG. DX, MH, and WH developed the sampling design; and MH, WH and EM wrote the section on samples and performed the sample size calculation. SS provided original data of the previous studies for the sample size estimation and calculated some summary statistics. JL and YC worked on the SP case templates. YC and XW developed the guideline for the development of the quality checklist. KH reviewed the content and edited the manuscript. HH and GC reviewed the statistical plan. SR, JP, HW, ZZ, CT, NZ and WZ reviewed and commented on the design and methods. All co-authors participated in the revision and approved this draft.

Competing interests

There are no competing interests.

Funding

This project is funded through following competitive grants: China Medical Board (grant number: CMB GNL 16-260); National Natural Science Foundation of China (grant number: 81773446); Young Scientists Fund of the National Natural Science Foundation of China (grant number: 81402690).

Data sharing statement

We have not yet started data collection. However, the data generated from this project and the USP cases and accompanying user manuals will be made available to other researchers upon request after we complete our primary analysis.

References

- 1. A/RES/70/1 R. Transforming our world: the 2030 agenda for sustainable development 2015 [Available from: http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E accessed Feburary 17, 2018 2018.
- 2. Hanefeld J, Powell-Jackson T, Balabanova D. Understanding and measuring quality of care: dealing with complexity. *Bulletin of the World Health Organization* 2017;95(5):368.
- 3. Murray CJ, Frenk J. A WHO framework for health system performance assessment: Evidence and Information for Policy, World Health Organization 1999.
- 4. Donabedian A. The quality of care: how can it be assessed? *Archives of pathology & laboratory medicine* 1997;121(11):1145.
- 5. Pongsupap Y, Lerberghe WV. Choosing between public and private or between hospital and primary care: responsiveness, patient - centredness and prescribing patterns in outpatient consultations in Bangkok. *Tropical Medicine & International Health* 2006;11(1):81-89.
- 6. Bitton A, Ratcliffe HL, Veillard JH, et al. Primary health care as a foundation for strengthening health systems in low-and middle-income countries. *Journal of general internal medicine* 2017;32(5):566-71.
- 7. Wei X, Li H, Yang N, et al. Changes in the perceived quality of primary care in Shanghai and Shenzhen, China: a difference-in-difference analysis. *Bulletin of the World Health Organization* 2015;93(6):407-16.
- 8. Zou Y, Zhang X, Hao Y, et al. General practitioners versus other physicians in the quality of primary care: a cross-sectional study in Guangdong Province, China. *BMC family practice* 2015;16(1):134.
- 9. Feng S, Shi L, Zeng J, et al. Comparison of Primary Care Experiences in Village Clinics with Different Ownership Models in Guangdong Province, China. *PloS one* 2017;12(1):e0169241.
- 10. Wong WC, Jiang S, Ong JJ, et al. Bridging the Gaps between patients and primary care in China: a nationwide representative survey. *The Annals of Family Medicine* 2017;15(3):237-45.
- 11. Zeng L, Li Y, Zhang L, et al. Guideline use behaviours and needs of primary care

practitioners in China: a cross-sectional survey. *BMJ open* 2017;7(9):e015379.

- 12. Li X, Lu J, Hu S, et al. The primary health-care system in China. *The Lancet* 2017;390(10112):2584-94.
- 13. Das J, Hammer J. Quality of primary care in low-income countries: facts and economics. *Annu Rev Econ* 2014;6(1):525-53.
- 14. Sylvia S, Shi Y, Xue H, et al. Survey using incognito standardized patients shows poor quality care in China's rural clinics. *Health policy and planning* 2014;30(3):322-33.
- 15. Berendes S, Heywood P, Oliver S, et al. Quality of private and public ambulatory health care in low and middle income countries: systematic review of comparative studies. *PLoS medicine* 2011;8(4):e1000433.
- 16. 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 affairs* 2012;31(12):2774-84.
- 17. Das J, Gertler PJ. Variations in practice quality in five low-income countries: a conceptual overview. *Health affairs* 2007;26(3):w296-w309.
- 18. Das J, Hammer J, Leonard K. The quality of medical advice in low-income countries. *Journal of Economic Perspectives* 2008;22(2):93-114.
- 19. Coarasa J, Das J, Gummerson E, et al. A systematic tale of two differing reviews: evaluating the evidence on public and private sector quality of primary care in low and middle income countries. *Globalization and health* 2017;13(1):24.
- 20. Glassman PA, Luck J, O'Gara EM, et al. Using standardized patients to measure quality: evidence from the literature and a prospective study. *Joint Commission Journal on Quality and Patient Safety* 2000;26(11):644-53.
- 21. Leonard K, Masatu MC. Outpatient process quality evaluation and the Hawthorne Effect. *Social science & medicine* 2006;63(9):2330-40.
- 22. McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. *Journal of clinical epidemiology* 2014;67(3):267-77.
- 23. Woodward CA, McConvey GA, Neufeld V, et al. Measurement of physician performance by standardized patients: refining techniques for undetected entry in physicians' offices. *Medical care* 1985:1019-27.
- 24. Das J, Hammer J. Money for nothing: the dire straits of medical practice in Delhi, India. *Journal of Development Economics* 2007;83(1):1-36.
- 25. 钟玉杰, 王敏, 李勤. 从 10 年文献回顾分析我国标准化病人教学的发展. *中华护理杂志* 2009;44(3):259-61.
- 26. Currie J, Lin W, Zhang W. Patient knowledge and antibiotic abuse: Evidence from an audit study in China. *Journal of health economics* 2011;30(5):933-49.
- 27. 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 Medicine* 2017;14(10):e1002405.
- 28. Li L, Lin C, Guan J. Using standardized patients to evaluate hospital-based intervention outcomes. *International journal of epidemiology* 2013;43(3):897-903.
- 29. Liao J, Chen Y, Cai Y, et al. Using smartphone-based virtual patients to assess the quality of primary healthcare in rural China: protocol for a prospective multicentre study. *BMJ open* 2018;8(7):e020943.

- 30. Zhou M, Wang H, Zhu J, et al. Cause-specific mortality for 240 causes in China during 1990–2013: a systematic subnational analysis for the Global Burden of Disease Study 2013. *The Lancet* 2016;387(10015):251-72.
- 31. Das J, Hammer J. Which doctor? Combining vignettes and item response to measure clinical competence. *Journal of Development Economics* 2005;78(2):348-83.
- 32. Hambleton RK, Swaminathan H, Rogers HJ. Fundamentals of item response theory: Sage 1991.
- 33. Rethans JJ, Gorter S, Bokken L, et al. Unannounced standardised patients in real practice: a systematic literature review. *Medical education* 2007;41(6):537-49.
- 34. Kutob RM, Bormanis J, Crago M, et al. Assessing culturally competent diabetes care with unannounced standardized patients. *Fam Med* 2013;45(6):400-08.
- 35. Fenton JJ, Kravitz RL, Jerant A, et al. Promoting patient-centered counseling to reduce use of low-value diagnostic tests: a randomized clinical trial. *JAMA internal medicine* 2016;176(2):191-97.
- 36. May L, Franks P, Jerant A, et al. Watchful waiting strategy may reduce low-value diagnostic testing. *The Journal of the American Board of Family Medicine* 2016;29(6):710-17.
- 37. ORDERING OF LABS AND TESTS: VARIATION AND CORRELATES OF VALUE-BASED CARE IN AN UNANNOUNCED STANDARDIZED PATIENT VISIT. JOURNAL OF GENERAL INTERNAL MEDICINE; 2016. SPRINGER 233 SPRING ST, NEW YORK, NY 10013 USA.
- 38. Tamblyn RM, Abrahamowicz M, Berkson L, et al. First-visit bias in the measurement of clinical competence with standardized patients. *Academic Medicine* 1992;67(10):S22-4.
- 39. Shepherd HL, Barratt A, Trevena LJ, et al. Three questions that patients can ask to improve the quality of information physicians give about treatment options: a cross-over trial. *Patient education and counseling* 2011;84(3):379-85.
- 40. Peabody JW, Luck J, Jain S, et al. Assessing the accuracy of administrative data in health information systems. *Medical care* 2004;42(11):1066-72.
- 41. Organization WH. WHO handbook for guideline development: World Health Organization 2014.
- 42. Campbell S, Braspenning J, Hutchinson A, et al. Research methods used in developing and applying quality indicators in primary care. *Qual Saf Health Care* 2002;11(4):358-64.
- 43. De Champlain AF, Margolis MJ, King A, et al. Standardized patients' accuracy in recording examinees' behaviors using checklists. *Academic Medicine* 1997;72(10):S85-7.
- 44. Vu NV, Steward DE, Marcy M. An assessment of the consistency and accuracy of standardized patients' simulations. *Academic Medicine* 1987;62(12):1000-2.
- 45. Vu NV, Marcy M, Colliver J, et al. Standardized (simulated) patients' accuracy in recording clinical performance check-list items. *Medical Education* 1992;26(2):99-104.
- 46. Maiburg BH, Rethans JJE, Van Erk IM, et al. Fielding incognito standardised patients as 'known'patients in a controlled trial in general practice. *Medical education* 2004;38(12):1229-35.
- 47. L. Gorter J-JR, Albert JJA Scherpbier, Sjef van der Linden, Marijke HM van Santen-Hoeufft, Désirée MFM van der Heijde, Harry HML Houben, Cees PM van der Vleuten, Simone. How to introduce incognito standardized patients into outpatient clinics of specialists in rheumatology. *Medical teacher* 2001;23(2):138-44.

48. Siminoff LA, Rogers HL, Waller AC, et al. The advantages and challenges of unannounced standardized patient methodology to assess healthcare communication. *Patient education and counseling* 2011;82(3):318-24.

49. Oates J, Weston WW, Jordan J. The impact of patient-centered care on outcomes. *Fam Pract* 2000;49(9):796-804.

50. Hudon C, Fortin M, Haggerty JL, et al. Measuring patients' perceptions of patient-centered care: a systematic review of tools for family medicine. *The Annals of Family Medicine* 2011;9(2):155-64.

51. Brown J, Stewart M, Tessier S. Assessing communication between patients and doctors: a manual for scoring patient-centred communication. *London: Thames Valley Family Practice Research Unit* 1995

52. Ozuah PO, Reznik M. Can standardised patients reliably assess communication skills in asthma cases? *Medical education* 2007;41(11):1104-05.

53. Zabar S, Ark T, Gillespie C, et al. Can unannounced standardized patients assess professionalism and communication skills in the emergency department? *Academic Emergency Medicine* 2009;16(9):915-18.

54. Gorter S, Rethans J-J, Scherpbier A, et al. Developing Case-specific Checklists for Standardized - patient—Based Assessments in Internal Medicine: A Review of the Literature. *Academic Medicine* 2000;75(11):1130-37.

55. Franz CE, Epstein R, Miller KN, et al. Caught in the act? Prevalence, predictors, and consequences of physician detection of unannounced standardized patients. *Health services research* 2006;41(6):2290-302.

56. Swartz MH, Colliver JA, Bardes CL, et al. Validating the standardized-patient assessment administered to medical students in the New York City Consortium. *Academic medicine: journal of the Association of American Medical Colleges* 1997;72(7):619-26.

57. Rethans J, Drop R, Sturmans F, et al. A method for introducing standardized (simulated) patients into general practice consultations. *Br J Gen Pract* 1991;41(344):94-96.

58. Luck J, Peabody JW. Using standardised patients to measure physicians' practice: validation study using audio recordings. *Bmj* 2002;325(7366):679.

59. Shirazi M, Sadeghi M, Emami A, et al. Training and validation of standardized patients for unannounced assessment of physicians' management of depression. *Academic Psychiatry* 2011;35(6):382-87.

60. Lin L. A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometric, 45, 255-268, 1989.

61. Steichen TJ, Cox NJ. A note on the concordance correlation coefficient. *Stata J* 2002;2(2):183-89.

62. Lawrence I, Lin K. Assay validation using the concordance correlation coefficient. *Biometrics* 1992:599-604.

63. Kwiecien R, Kopp-Schneider A, Blettner M. Concordance analysis: part 16 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt International* 2011;108(30):515.

64. Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet* 1986;327(8476):307-10.

65. Rhodes K. Taking the mystery out of "mystery shopper" studies. *New England Journal of Medicine* 2011;365(6):484-86.

66. Rhodes KV, Miller FG. Simulated patient studies: an ethical analysis. *The Milbank Quarterly* 2012;90(4):706-24.

67. Weiner SJ, Schwartz A. Directly observed care: can unannounced standardized patients address a gap in performance measurement? *Journal of general internal medicine* 2014;29(8):1183-87.

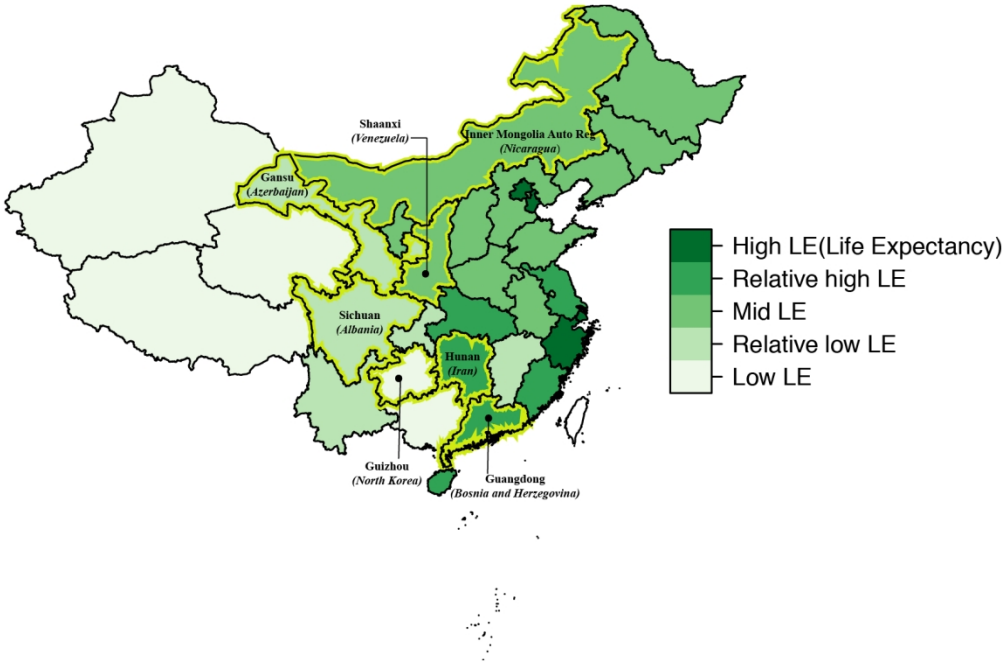

Figure 1 Selected seven sample provinces on the map of China with referencing countries of equivalent life expectancy in the bracket

147x98mm (300 x 300 DPI)

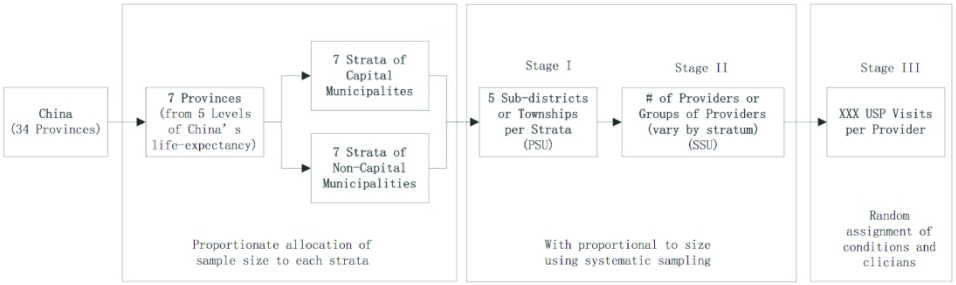

Figure 2 Sampling Procedure

397x122mm (300 x 300 DPI)

Web Appendix

Web Appendix 1 Sample Size Calculation

Compute the sampling variance of the mean: $var(\bar{y})$, based on desired coefficient of variation - 0.08.

$$var(\bar{y}) = se(\bar{y})^2 = (cv * \bar{y})^2 = (0.08 * (-0.9))^2 = 0.0052$$

Estimate number of completed interviews in need for a simple random sample(SRS):n_{srs}

$$n_{srs} = \frac{s^2}{var(\bar{y})} = \frac{4.54}{0.0052} = 875$$

Estimate design effect:

$$d_{eff} = 1 + \delta(n-1) = 1 + 0.0486 * (27-1) = 2.26$$

Multiply n_{srs} by the design effect to account for a complex survey design:

$$n_{complex} = n_{srs} * d_{eff} = 875 * 2.26 \approx 1981$$

Web Appendix 2 Evidence-based process of developing quality criteria for the SP cases

In partnership with the Lanzhou University Evidence Medicine Center, we have developed a working paper on the results of our review of the literature in quality checklist development and also our recommended protocol of developing

those checklists. We provide an abstract of that working paper below and will make available the full paper once it is fully developed.

Objective To explore the procedures and methods for determining the quality checklist for the most common conditions in the context of primary health care, particularly to be used for quality inspection by unannounced standardized patients.

Methods We conducted a systematic search of literature in the subject matter, while adopting the WHO handbook for guideline development. **Results** A total of 14 related articles were included and the methodological aspects were evaluated. Based on this review, we propose five key steps in the checklist development: (1) Forming a multidisciplinary team; (2) Reviewing, evaluating and selecting relevant literature based on evidence-based medicine quality of evidence principles; (3) Extracting essential quality information to form a pool of quality items; (4) using expert consensus to select candidate quality checklist items from the pool; (5) pre-testing to determine the final items. **Discussion** We recommend a checklist development method based on evidence-based method augmented by expert opinions through a multidisciplinary group discussion. The selection of the items on the checklist will

consider their importance and feasibility. Our proposed methods can be mainly applied to common conditions seen in the primary care settings and may not be applied to more complex conditions.