

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

Development and validation of smartphone-based virtual patients to assess the quality of primary health care in rural areas: a study protocol

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2017-020943
Article Type:	Protocol
Date Submitted by the Author:	08-Dec-2017
Complete List of Authors:	<p>Liao, Jing; Sun Yat-sen University, Department of Epidemiology and Biostatistics, School of Public Health; Sun Yat-sen University, Sun Yat-sen Global Health Institute, Institute of State Governance</p> <p>Chen, Yao-Long; Lanzhou University, Evidence-Based Medicine Center, School of Basic Medical Sciences</p> <p>Cai, Yiyuan; Sun Yat-sen University, Department of Epidemiology and Biostatistics, School of Public Health; Guizhou Medical University, School of Public Health</p> <p>Zhang, Nan; Inner Mongolia Medical University, Department of Health Management, School of Health Management</p> <p>Sylvia, Sean ; University of North Carolina at Chapel Hill, Department of Health Policy and Management, Gillings School of Global Public Health</p> <p>Wang, Hong; Bill & Melinda Gates Foundation, Health Economics, Financing & Systems</p> <p>Wasserheit, Judith; University of Washington, Departments of Global Health, Medicine, and Epidemiology, Schools of Medicine and Public Health</p> <p>Gong, Wenjie; Central South University, School of Public Health; University of Rochester Medical Center, Department of Psychi</p> <p>Zhou, Zhongliang; Xi'an Jiaotong University,</p> <p>Pan, Jay; Sichuan University, West China School of Public Health</p> <p>Wang, Xioahui; Lanzhou University, School of Public Health</p> <p>Tang, Chengxiang; Guangzhou University, School of Public Administration</p> <p>Zhou, Wei; Xiangya Hospital Central South University,</p> <p>Xu, Dong; Sun Yat-sen University , Sun Yat-sen Global Health Institute, Institute of State Governance; Sun Yat-sen University, Department of Epidemiology and Biostatistics, School of Public Health</p>
Keywords:	Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, PRIMARY CARE, quality assessment tool

SCHOLARONE™
Manuscripts

Development and validation of smartphone-based virtual patients to assess the quality of primary health care in rural areas: a study protocol

Jing Liao^{1,2}, Yaolong Chen³, Yiyuan Cai⁴, Nan Zhan⁵, Sean Sylvia⁶, Hong Wang⁷, Judith N Wasserheit⁸, Wenjie Gong⁹, Zhongliang Zhou¹⁰, Jay Pan¹¹, Xiaohui Wang¹², Chenxiang Tang¹³, Wei Zhou¹⁴, Dong Xu^{1,2}*

¹ Department of Epidemiology and Biostatistics, School of Public Health, Sun Yat-sen University, No.74 Zhongshan 2nd Road, Guangzhou, P.R. China, 510080

² Sun Yat-sen Global Health Institute, Institute of State Governance, Sun Yat-sen University, No.135 Xingang West Road, P.R. China, 510275

³ Evidence Based Medicine Center, School of Basic Medical Sciences, Lanzhou University, No. 199 Donggang West Rd, Lanzhou City, Gansu Province, 730000 China

⁴ School of Public Health, Guizhou Medical University, UniverCity of Guan New Area, Guizhou, China, 550025

⁵ Department of Health Management, School of Health Management, Inner Mongolia Medical University, Jinshan Development Zone, Hohhot, Inner Mongolia, P.R. China, 010110

⁶ 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

⁷ Health Economics, Financing & Systems, PO Box 23350, Seattle WA, 98102, USA

⁸ Departments of Global Health, Medicine, and Epidemiology, Schools of Medicine and Public Health, University of Washington, 1705 NE Pacific Street, Box 357965 Seattle, Washington, 98195-7965, United States

⁹ Xiangya school of public health, Central South University, Xiangya school of public health, Central South University, China

¹⁰ School of Public Policy and Administration, Xi'an Jiaotong University, No. 28 Xianning West Road, Xi'an, Shaanxi, 710049, China.

¹¹ West China School of Public Health, Sichuan University, N.o. 17, Ren Min Nan Road, Chengdu, China

¹² School of Public Health, Lanzhou University, Lanzhou, Gansu Province, P. R. China, 730000

¹³ School of Public Administration, Guangzhou University, Guangzhou, Guangdong, 510320, China.

¹⁴ Hospital Administration Institute, Xinagya Hospital, Central South University, No. 87 Xiangya Road, Changsha, Hunan., China

Corresponding author*: Dong Xu, xudong5@mail.sysu.edu.cn, Department of Epidemiology and Biostatistics, School of Public Health, Sun Yat-sen University, No.74 Zhongshan 2nd Road, Guangzhou, P.R. China, 510080

ABSTRACT

Introduction: Valid, low cost and easy-to-implement tools for evaluating care quality are not readily available. While the Unannounced Standardized Patient (USP) is recommended as the gold standard for assessing quality, its use is restricted by the limited range of medical conditions that can be studied and a high implementation cost. Clinical vignettes provide a low-cost alternative; however, their lack of realism have brought their validity into question. Computerized virtual patients (VPs) can create high-fidelity, visualized and interactive simulations of doctor-patient encounters closely replicating clinical complexity that can be easily implemented via smartphone, at low marginal cost. Our study aims to develop and validate smartphone-based VP against USP as a quality assessment tool that can be used for research purposes and routinely applied to evaluate the quality of primary health care provided by primary health centers (PHCs) in rural areas.

Methods and analysis: The study will be implemented in outpatient settings of rural PHCs in seven Chinese provinces, and physicians practicing at township health centers and village clinics will be our study population. The development of VPs will involve three steps: (1) identifying 10 VP cases that can best represent rural PHCs' work, (2) designing each case by a case-specific development team, and (3) developing corresponding quality scoring criteria. After being externally reviewed for content validation, these VP cases will be implemented on a smartphone-based platform and will be tested for feasibility and face validity. This smartphone-based VP tool will then be validated for its criterion validity against USP and its reliability (i.e., internal

1
2
3 consistency and stability), with 1260 VP/USP-clinician encounters across the seven
4
5
6 study provinces for all 10 VP cases.
7

8 **Ethics and dissemination:** Sun Yat-sen University: No. 2017-007. Study findings will
9
10 be published and the tools developed will be freely available to low- and
11
12
13 middle-income countries for research purposes.
14

15 **Strengths and limitations of this study**

- 16 ● Developing and validating smartphone-based VP as a quality assessment tool
17 for research and routine use in rural primary health care centers.
- 18 ● Following an evidence-based approach to develop VP cases and scoring
19 criteria.
- 20 ● Systematically validating the VP assessment tool via a cross-national
21 multicenter study
- 22 ● The extent to which the VP assessment can reflect practitioners' real clinical
23 practice needs to be verified.
24
25
26
27

28 **Key words:** Quality in health care, primary care, care quality assessment tool
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

INTRODUCTION

Universal health coverage (UHC) is a paramount goal of health system development for countries at all income levels.¹ The achievement of UHC is not possible without primary health care services,¹ which ensure integrated care close to the population they serve and link to health-related sustainable development goals.² However, service coverage alone cannot improve health outcomes if the quality of care is poor. Despite the fact that much attention has been shifted to enhance health care services, a dearth of scientific evidence exists on the quality of primary health care in resource-poor settings, particularly of low- and middle- income countries (LMICs).³⁻⁵

This scarcity of evidence may partially result from the limited availability of valid, low cost, and easy-to-implement quality assessment tools. As defined by Donabedian's framework for health care quality, quality can be evaluated by the *structure* of care (e.g., staff, equipment), the *process* of care delivery (e.g., doctor-patient interactions), and health *outcomes* (e.g., death or complications).⁶ Process measures have been increasingly adopted, considering the benefits of frequent and timely evaluation, and the usefulness in improving practice.^{7,8} The 'gold standard' of assessing process is the unannounced standardized patient (USP), namely a trained actor who simulates the symptoms, signs, and emotions of a real patient in a standardized fashion and presents him- or herself unannounced to clinics to assess care quality.⁹ USP can reduce recall bias better than patient exit interviews, minimize the Hawthorne effect that inevitably occurs in direct observation, and allow for comparisons between users as case- and patient-mix are controlled.^{3,7,10} Nonetheless,

1
2
3 the USP can only portray a limited number of conditions without obvious
4
5 physiological symptoms and risk of invasive examinations. Also, training and
6
7 implementation of USP can require substantial personnel and resources, making USP
8
9 impractical for large-scale and routine quality assessment.^{11 12}
10
11
12

13 On the other hand, clinical vignettes or case simulations have been widely used
14
15 as a low-cost and convenient alternative to assess care quality.^{7 13} Vignettes have been
16
17 implemented in a paper-and-pencil form,⁷ presented by an enumerator,⁵ and
18
19 streamlined by a computer.¹³ The validity of the vignette in assessing the quality of
20
21 patient care has gotten mixed reports. Some studies showed that vignettes only reflect
22
23 clinicians' competency (know-how) rather than their actual behaviors and can cause
24
25 overestimation of clinical performance.^{7 14} By contrast, other studies found that
26
27 vignette-based results, particularly those streamlined by computer, are quite close to
28
29 the USP-based assessment.^{13 15} The enumerator-administered vignette is similar to the
30
31 announced standardized patient and thus is expensive and difficult to implement.⁵ A
32
33 computerized vignette can be interactive and can better represent the realistic
34
35 complexity of a clinical encounter.¹³ As a further improvement on computerized
36
37 vignettes,¹³ smartphone virtual patients (VPs) create high-fidelity, visualized, and
38
39 interactive simulations that replicate clinical complexity and can be easily
40
41 implemented at only marginal cost.¹⁶ Although VPs cannot remove the Hawthorne
42
43 effect, their advanced features may reduce the measurement gap between competency
44
45 and real practice.^{8 13} While VPs have been used in medical education to train and test
46
47 clinical skills such as clinical reasoning, making diagnoses, and therapeutic
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 decisions,¹⁷ the relative validity of their use as a measure for the quality of care has
4
5 yet to be studied. Strengths and limitations of the abovementioned three methods are
6
7 compared in Table 1.
8
9

10
11 In the present study, we propose to adapt smartphone-based VP for medical
12
13 education as a quality-of-care assessment tool, given its advantages in 1)
14
15 **standardization** (VPs are highly standardized, ensuring consistent assessments across
16
17 users), 2) **flexibility** (Assessments can be delivered by smartphones for multiple users
18
19 at anytime, anywhere, providing the data connectivity is available), 3) **scalability**
20
21 (VPs can be modified to demonstrate and assess almost any clinical conditions with
22
23 low marginal cost), and 4) **training** (VPs can further be used as a training tool to
24
25 improve health care quality and thus to address the ‘so what’ question after quality
26
27 assessment). These characteristics may especially benefit quality assessment and
28
29 improvement in primary care of rural communities which are geographically scattered
30
31 and difficult to reach and manage.
32
33
34
35
36
37

38
39 Therefore, our study aims to develop and validate **smartphone-based VPs**
40
41 against USPs as a quality assessment tool that can not only be used for research
42
43 purposes but also routinely applied to evaluate the quality of primary health care
44
45 provided by primary health centers (PHCs) in rural areas. To maximize its validity,¹³
46
47 we will properly construct high-fidelity VP cases to reflect clinical complexity in rural
48
49 PHC contexts with real-time patient-doctor interactions and temporal constraints, as
50
51 well as use evidence-based quality care scoring criteria; additionally, we will make
52
53 the VP-based test anonymous to minimize the Hawthorne effect. The initial phase of
54
55
56
57
58
59
60

1
2
3 the study will mainly focus on rural China, while the ultimate goal is to develop and
4
5
6 validate tools that can have a broad application in other LMICs.
7

8 **METHODS AND ANALYSIS**

9

10 **Study setting**

11
12 The validation study will be implemented in the outpatient setting of rural PHCs (i.e.,
13
14 township health centers and village clinics) in seven Chinese provinces (Guizhou,
15
16 Sichuan, Gansu, Inner Mongolia, Shaanxi, Hunan, and Guangdong). We are selecting
17
18 these provinces not only to reflect the five strata of low-to-high life-expectancies and
19
20 various burden-of-disease patterns in China,¹⁸ but also to contrast geographic regions
21
22 with ethnic diversities, from southwest mountainous regions, to the northern plateau,
23
24 middle inland region, and southeast coastal areas (**Error! Reference source not**
25
26 **found.**). Our study targets township health centers and village clinics because they
27
28 provide most primary health care in rural China.¹⁹ At township health centers, primary
29
30 health care is delivered by a workforce including licensed/unlicensed physicians,
31
32 licensed/unlicensed assistant physicians, and registered nurses; while at village clinics,
33
34 services are mainly delivered by one full- or part-time ‘village doctor’ who is a
35
36 clinician with rudimentary medical training.¹⁹⁻²¹ The outpatient setting is chosen due
37
38 to the few inpatient cases in township health centers and village clinics.
39
40
41
42
43
44
45
46

47 **VP case development**

48 *VP case selection*

49
50 We intend to select 10 cases that can best represent the work of rural PHCs. The
51
52 selection of the VP cases will be based on the following criteria: 1) high frequency of
53
54
55
56
57

1
2
3 clinical encounters in the primary care settings in rural areas, and/or 2) association with
4 significant disease burden; 3) representation of the major areas of work of PHCs in
5 rural China overall (e.g., public health service delivery, chronic disease management,
6 infectious disease control, health education, and patient-centered care); and 4)
7 suitability with USP methodology (e.g., no obvious physiological signs, low risk for
8 invasive tests) for the sake of criterion validation in the current study. A case selection
9 committee will be comprised of stakeholders, including physicians, public health
10 practitioners, policy-makers, and members of the research team. Based on the literature
11 review, the research team will prepare a list of the 30 most frequently seen conditions in
12 township health centers and village clinics reported by either community dwellers²² or
13 the rural PHC clinicians (Appendix1) for the committee to rate and select.

30 *VP case design*

31
32 The 10 selected VP cases will then be constructed individually by 10 case-specific
33 development teams (Figure 2). These teams consist of one *condition expert* in the
34 relevant specialties of a tertiary teaching hospital who will be responsible for drafting
35 the VP case; an *evidence-synthesis group* involving epidemiologists and
36 evidence-based researchers who will search and synthesize evidence of the selected
37 condition for the condition expert to work on; a *clinical consensus group* which
38 consists of several condition-related clinical experts who will review the
39 corresponding case from a scientific perspective; an overall all-condition shared
40 *context-expert panel*, which includes clinicians and health managers from community
41 health centers, township health centers, and village clinics, who will review the
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 contextual appropriateness of the cases in the rural PHC setting; and a *case*
4
5
6 *coordinator* who will coordinate development of each case.
7

8 Each VP case will be structured into five domains—medical history, physical
9
10 examination, laboratory and imaging studies, diagnosis, and management and
11
12 treatment plan—to simulate real-life clinical scenarios.^{10 17} The structured VP cases
13
14 will have the ability to evaluate the examinee's performance by each domain and to
15
16 aggregate performance scores across conditions. Besides these five condition-related
17
18 domains, another practice contextual adjustor will be built into each case to consider
19
20 medical resource constraints in rural practices (e.g., availability of basic medical
21
22 equipment and medicines).
23
24
25
26

27 28 *Scoring criteria*

29 Corresponding to each VP case, care quality scoring criteria will be developed. These
30
31 criteria include *process quality*, the *accuracy of diagnosis*, and the *appropriateness of*
32
33 *the treatment and management plan*.^{3 4} Process quality will be evaluated in reference
34
35 to a clinical process checklist (to be detailed later) including all necessary questions
36
37 that should be asked and physical examinations that should be performed by clinicians,
38
39 alongside redundant or even potentially harmful practices. Diagnoses will be rated as
40
41 correct, partially correct, or incorrect based on predetermined standards. The
42
43 treatment & management plan will be considered appropriate if the clinician
44
45 prescribes any of the correct medications or refers the patient to an upper level
46
47 physician depending on the VP cases.
48
49
50
51
52
53

54 In addition, *cost of care* and *time-spent per encounter* will also be recorded.
55
56

1
2
3 Patient costs will cover medication fees and clinic fees charged per case. In order to
4
5
6 recode clinicians' reaction time to each domain and to impose temporal constraints as
7
8
9 in real clinical practices, the entire clinician-VP interaction process will be timed. This
10
11 will include time spent on taking history, conducting physical examinations,
12
13 prescribing drugs and treatments, and any interruptions)

14
15
16 As for the development of the aforementioned checklist for the predetermined
17
18 standards of correctness regarding appropriateness of the treatment and management
19
20 plan, a systematic evidence-based approach will be adopted (Appendix 2). Briefly, the
21
22 evidence-synthesis group will systematically search and extract condition-specific
23
24 checklist items and standards from clinical guidelines, reputable textbooks, and
25
26 systematic reviews, etc., in that order, whereas the quality of the evidence will then be
27
28 rated by the Appraisal of Guidelines for Research & Evaluation II (AGREE II)²³ or
29
30 the revised Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2)
31
32 according to its type.²⁴ Afterwards, the clinical consensus group and the
33
34 contextual-expert panel will review and revise initial standards using a Delphi
35
36 process.²⁵

37 38 39 40 41 42 **VP case external review**

43
44
45 To validate the content of VP cases, an independent expert panel of physicians,
46
47 general practitioners, and rural PHC clinicians who otherwise are not involved in the
48
49 study, will be convened to review the cases for content accuracy and appropriateness.
50
51
52 The content validation involves qualitative and quantitative phases. In the qualitative
53
54 phase, the expert panel will be required to evaluate the cases with respect to the
55
56
57
58
59
60

1
2
3 following: overarching assessment goal, representativeness of the goal and test items
4
5
6 to each domain, the logical relationship of the content tested, and the appropriate
7
8 wording, grammar, understandability, and relatedness to the rural PHC context. The
9
10 panel will also mention their suggestions, if any, next to each item. Modified VP cases
11
12 will then be given back to the expert panel for quantitative evaluation with respect to
13
14 their adjustments for simplicity and clarity, as well as necessity and relevance to the
15
16 assessment, using a four-point Likert scale ranging from 1 (the lowest) to 4 (the
17
18 highest). The content validity index (CVI) will be computed for each domain and for
19
20 the entirety of VP cases.²⁶
21
22
23
24

25 **Technical implementation**

26
27 Revised VP cases will be implemented on *CureFUN*, an existing smartphone- based
28
29 training platform using VPs with special customizations and set-ups to suit the
30
31 assessment purpose. A live demonstration of a simplified VP can be accessed from
32
33 http://www.curefun.com/zhiqu_front/www/experience/experience.html#/caseList and
34
35
36 is also illustrated in Appendix 3. The smartphone-based VP assessment tool will not
37
38 only present interactive clinical scenarios, but also automatically record each
39
40 examinee's diagnosis pathway and grade it against the scoring criteria (Figure 33).
41
42
43
44

45 **Feasibility study**

46
47 Before a full-scale validation study, a feasibility study with 30% of the validation
48
49 study sample (see study sample section) will be conducted to test the VP assessment
50
51 tool's usability, accessibility, and stability, particularly in remote village clinics with
52
53 weak phone connectivity. Selected clinicians will be instructed to individually attempt
54
55
56
57

1
2
3 two random VP cases within a given time, using their own smartphone devices from
4
5 their workplace. Clinicians without a smartphone will be given a temporary one
6
7 installed with the customized *CureFUN* applications. Clinicians' willingness to
8
9 participate and adherence to the VP-based tests (e.g., percentage completing VP cases,
10
11 score of the assessment, and number of attempts made at each case per person) will be
12
13 automatically recorded. Upon completion of the cases, participants will be asked to
14
15 fill in a five-point Likert-scale questionnaire regarding their subjective attitude toward
16
17 the simulator VP experience (with 1 being the most negative response and 5 being the
18
19 most positive), regarding ease of use, experienced assessment process and outcome,
20
21 realism, device competence, accessibility, and other general comments. These results
22
23 will be used to determine VP cases' face validity, whereby a satisfying score equals
24
25 frequency (%) multiplied by positive evaluations; and scores no less than 1.5 are
26
27 considered acceptable.²⁷

28 29 30 31 32 33 34 35 **Validation of VP as a quality assessment tool**

36 37 38 *Study design*

39
40 The prospective validation study is a nationwide multicenter study with two main
41
42 functions. This study will 1) assess the criterion validity of the VP-tool in assessing
43
44 the quality of primary health care, by analyzing its measurement concordance against
45
46 the standard USP measure, and 2) test the reliability of the VP tool, by examining its
47
48 internal consistency and the stability of repeated VP assessments on the same
49
50 examinees.
51
52
53
54
55
56
57
58
59
60

Study sample

From each of our seven sample provinces, two counties will be selected with sufficient variations in socio-economic conditions, demographics, and disease burdens between them while also approximating the provincial condition in general. Within each county, the government registry of all township health centers and village clinics will serve as our sampling frame, which will include 1) licensed practicing physicians, 2) clinicians who have not been licensed but are providing clinical services under the supervision of licensed physicians at township health centers, as well as 3) full- or part-time village clinicians. Temporarily-visiting clinicians (often senior clinicians sent by higher level medical institutions to support the development of township health centers), nurses, and allied health workers without prescription privileges will be excluded.

The sample size calculation is based on individual VP/USP-clinician encounter, and ensures sufficient power to detect variations at individual case level per county. For village clinics, one VP/USP case will be examined per time to minimize the detection of USPs. Assuming a 5% type I error and 80% power, to determine whether a moderate concordance correlation coefficient²⁸ of 0.90²⁹ between VP and USP differs from zero, seven paired VP/USP-clinician encounters will be required for each of the 10 cases per county. As a stratified sampling strategy will be deployed that first samples townships and then villages from each township, the design effect should be taken into consideration when calculating the sample size. Assuming an intra-class correlation of 0.05 and 6 village clinics per township, then 9 paired VP/USP-clinician

1
2
3 encounters is needed. These nine paired VP/USP-clinician encounters will be assigned
4
5
6 to 3 township health centers and 6 village clinics using probability proportional to size
7
8 (PPS) method. These 9 paired VP/USP-clinician encounters will be assigned to
9
10 township health centers and village clinics based on the ratio of the total number of
11
12 clinicians at township health centers over the total number of village clinicians for
13
14 each county. There are 1260 VP/USP-clinician encounters across our seven study
15
16 provinces for all 10 VP cases. Figure 4 shows the sampling process and study flow for
17
18 one VP case using Guizhou Province (Danzhai County) as a demonstration.
19
20
21

22 23 *Criterion validity*

24
25 Criterion validity³⁰ of the VP to assess quality of care will be evaluated mainly by its
26
27 measurement concordance against the USP measure as the recognized gold standard³¹
28
29 for assessing quality of care in real practice. The USPs will be developed in a related
30
31 study, sharing the development teams for VP and a similar development process. The
32
33 method of fielding USPs in rural China will follow approaches similar to those of the
34
35 previous USP study in rural China.³ Identical quality scoring criteria, described above,
36
37 will be applied to scores. Each selected clinician will first see a USP (to avoid the
38
39 practice effect due to the USP's unannounced feature), and then complete a
40
41 smartphone-based VP assessment of the same condition. The clinician to be assessed
42
43 will be randomly selected onsite by the USP from any on-duty clinicians on the day of
44
45 the USP visit to the sampled township health center and village clinics. This situation
46
47 would especially apply to township health centers, as most village clinics have only
48
49 one clinician (note: Chinese patients normally see their primary care clinicians as a
50
51
52
53
54
55
56
57
58
59
60

1
2
3 walk-in patient and appointments are seldom needed). To record USP-clinician
4 interactions, USPs will complete checklists immediately after their visit, as well as
5
6 retain the medication prescription and the fee charge slips by the clinician. A week
7
8 after the USP clinic visit, clinicians will be assigned a smartphone-based VP
9
10 assessment. The VP-clinician interactions, drugs dispensed, and fees charged will all
11
12 be recorded automatically by the online assessment system.
13
14
15
16
17

18 The concordance of the two assessments between USPs and VPs will then be
19
20 analyzed by Lin's concordance correlation coefficient (r_c)²⁸ for continuous *process*
21
22 *quality scores, fees charged (yuan), and time spent (min)*, and the Kappa statistic³² for
23
24 dichotomous *diagnoses and treatment & management* measures. r_c evaluates how
25
26 close pairs of observation fell on a 45° line (the perfect concordance line) through the
27
28 origin in addition to their correlation. Kappa measures agreement in assessment beyond
29
30 what is expected by chance alone. In addition, for continuous measures, a
31
32 Bland-Altman plot will also be used to visualize the concordance.^{33 34} For dichotomous
33
34 measures, we will analyze their sensitivity (i.e., strength to detect correct diagnosis,
35
36 treatment plan, etc.) and specificity (i.e., strength to detect incorrect diagnosis,
37
38 treatment plan, etc.) using USP as the reference.
39
40
41
42
43

44 *Reliability*

45
46 To establish *test-retest reliability*, clinicians previously being assessed by VPs will be
47
48 instructed to retake the same VP tests four weeks after their last assessment. The
49
50 second VP test is set one month later than the first to reduce the practice effect,³⁵
51
52 assuming the clinician's general medical knowledge remains constant.³⁰ The
53
54
55
56
57
58
59
60

1
2
3
4 concordance of the two repeated tests indicate the stability of the VP assessment
5
6 tool.³⁰ Similar concordance measures (i.e., r_c for continuous and Kappa for
7
8 dichotomous measures) as described above will be used. The *internal consistency*, the
9
10 intercorrelation of scores for process quality indicators, will be computed by
11
12 Cronbach's alpha coefficients (α),³⁶ with $\alpha > 0.7$ representing acceptable reliability.³⁷
13
14 Table 2 summarizes the validity, reliability, and feasibility measures that will be
15
16 examined in our study.
17
18
19

20 **ETHICS AND DISSEMINATION**

21
22
23 The study has been approved by the Institutional Review Board of the School of
24
25 Public Health (IRB), Sun Yat-sen University (No. 2017-007). Informed consent will
26
27 be obtained from all participating clinicians of VP tests. However, to reduce
28
29 participation bias due to self-selection,³⁸ our IRB has approved the implementation of
30
31 USP without prior informed consent from the individual participants, on the condition
32
33 that involved clinicians will be fully de-identified and all analyses will only be
34
35 conducted at the population level.³⁸ Study data will be securely stored and only
36
37 de-identified information will be used for analysis. We will seek peer-reviewed
38
39 publications for study findings and produce reports to inform health authorities. The
40
41 tools and technology developed in this study will be freely available to other LMICs
42
43 for research purposes.
44
45
46
47
48
49

50 **DISCUSSION**

51
52 To the best of our knowledge, this is the first study validating VP as a quality
53
54 assessment tool in rural primary health care centers. This study follows an
55
56
57
58
59
60

1
2
3
4 evidence-based approach to develop VP cases and scoring criteria, implements them
5
6 on a widely accessible platform (i.e., a smartphone), and systematically validates the
7
8 VP assessment tool via a cross-national multicenter study representing rural PHCs
9
10 over a wide range of geographic areas with distinct life expectancies and economic
11
12 development levels. The VP assessment tool's accessibility, flexibility and scalability
13
14 give it good potential to be easily adapted to other LMICs.
15
16

17
18 Nevertheless, it is to be noted that given its simulated nature, the VP-test
19
20 theoretically may never completely bridge the 'competency-practice' gap. The
21
22 validation study is thus essential to quantify the concordance/discordance between
23
24 VP- and USP-based quality assessments. Our study will generate firsthand empirical
25
26 evidence contributing to the understanding of the 'know-do gap',^{5,39} and further shed
27
28 light on circumstances that cannot be tested by USPs.
29
30
31

32
33 A limitation of the study, however, is that, in order to test the validity of VP
34
35 against USP as the reference standard, we restrict the selection of VP cases to those
36
37 that can be simulated by USP as well. Thus, the present study may not fully exhibit
38
39 the potential of the VP in assessing quality of care. Further, the two
40
41 purposely-selected counties for each province may not represent the provincial
42
43 conditions entirely, although we will make every effort to consider provincial
44
45 representation when selecting counties. Third, while the validation study is
46
47 exclusively conducted on PHCs in rural China, the extent to which the VP assessment
48
49 tool can be transported to other LMICs remains to be evaluated. Even so, the selected
50
51 provinces from the 'five-Chinas' may improve the generalizability of our study
52
53
54
55
56
57
58
59
60

1
2
3
4 considering the comparable life expectancies of LMICs and these provinces.
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

Author's contributions

All authors contributed to the conceptualization and design of the study. DX, JL, and YYC conceived the initial study design, analytical methods, and composition of the team. JL was responsible for the study concept, initial draft, and revisions. DX was responsible for the study concept and revising the draft. YLC and XHW were responsible for the development of the scoring criteria. SS, HW, JNW, ZLZ, ZN, WJG, JP, CXT and WZ provided critical review and revision to the study design. All authors read and approved the final revision.

Funding

The study is supported by the China Medical Board through its Health Policy and Systems Sciences Open Competition grant 'Quality in primary health care: using unannounced standardized patients' (grant No.: CMB16-260, XD, PI).

Acknowledgement

The study will be led by the Sun Yat-sen Global Health Institute of Sun Yat-sen University with a consortium of researchers from seven other Chinese universities, including Central South University, Guangzhou University, Guizhou Medical University, Inner Mongolia Medical University, Lanzhou University, Sichuan University, and Xi'an Jiaotong University. We thank all the students from these universities who have contributed to our project, especially Wenjun He who produced Figure 1, Jianjian Wang who assisted with the evidence collection of the Scoring Criteria, and Ash Harris who contributed to the VP computerization.

Competing interests

The development of the VP assessment tool is a joint project of the Sun Yat-sen University Global Health Institute (SGHI) (representing the seven universities in China), and *CureFUN*. However, the VP cases will be independently developed and the validation studies will be rigorously conducted by the research team from SGHI and the seven universities, whereas *CureFun* will technically implement the cases on smartphones and have no influence over the study design and analysis.

REFERENCES

1. Kieny MP, Evans DB. Universal health coverage. *East Mediterr Health J* 2013;**19**(4):305-6
2. Transforming our world: The 2030 agenda for sustainable development. : United Nations- Sustainable Development Knowledge Platform, 2015.
3. 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
4. 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
5. Das J, Hammer J, Leonard K. The quality of medical advice in low-income countries. *The Journal of Economic Perspectives* 2008;**22**(2):93-114
6. Donabedian A. Evaluating the quality of medical care. *The Milbank Quarterly* 2005;**83**(4):691-729
7. Peabody JW, Luck J, Glassman P, et al. Comparison of vignettes, standardized patients, and chart abstraction: A prospective validation study of 3 methods for measuring quality. *JAMA* 2000;**283**(13):1715-22
8. Shah R, Edgar D, Evans BJ. Measuring clinical practice. *Ophthalmic and Physiological Optics* 2007;**27**(2):113-25
9. 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 Improvement* 2000;**26**(11):644-53
10. Glassman PA, Luck J, O'Gara EM, et al. Using standardized patients to measure quality: Evidence from the literature and a prospective study. *The Joint Commission journal on quality improvement* 2000;**26**(11):644-53
11. Collins J, Harden R. The use of real patients, simulated patients and simulators in clinical examinations Association for Medical Education in Europe (AMEE) Guide, 2004.
12. Triola M, Feldman H, Kalet AL, et al. A randomized trial of teaching clinical skills using virtual and live standardized patients. *Journal General Internal Medicine* 2006;**21**(5):424-9
13. Peabody JW, Luck J, Glassman P, et al. Measuring the quality of physician practice by using clinical vignettes: A prospective validation study. *Annals of Internal Medicine* 2004;**141**(10):771-80
14. Shah R, Edgar DF, Evans BJ. A comparison of standardised patients, record abstraction and clinical vignettes for the purpose of measuring clinical practice. *Ophthalmic and Physiological Optics* 2010;**30**(3):209-24
15. Dresselhaus TR, Peabody JW, Luck J, et al. An evaluation of vignettes for predicting variation in the quality of preventive care. *Journal of general internal medicine* 2004;**19**(10):1013-18
16. Ellaway R, Candler C, Greene P, et al. An architectural model for ubiquitous virtual patients. Baltimore: MedBiquitous, 2006.

17. Cook DA, Triola MM. Virtual patients: A critical literature review and proposed next steps. *Med Educ* 2009;**43**(4):303-11 doi: 10.1111/j.1365-2923.2008.03286.x[published Online First: Epub Date]].
18. 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
19. Babiarz KS, Miller G, Yi H, et al. China's new cooperative medical scheme improved finances of township health centers but not the number of patients served. *Health Affairs* 2012;**31**(5):1065-74
20. Qian D, Pong RW, Yin A, et al. Determinants of health care demand in poor, rural china: The case of gansu province. *Health Policy and Planning* 2009;**24**(5):324-34
21. Yip WC, Wang H, Liu Y. Determinants of patient choice of medical provider: A case study in rural china. *Health Policy and Planning* 1998;**13**(3):311-22
22. An analysis report of national health services survey in china. Beijing: Center for Health Statistics and Information, NHFPC, 2013.
23. Brouwers MC, Kho ME, Browman GP, et al. Agree ii: Advancing guideline development, reporting and evaluation in health care. *Canadian Medical Association Journal* 2010;**182**(18):E839-E42
24. Whiting PF, Rutjes AW, Westwood ME, et al. Quadas-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 2011;**155**(8):529-36
25. Hsu CC, Sandford BA. The delphi technique: Making sense of consensus. *Practical Assessment Research & Evaluation* 2007;**26**(10):289–304
26. Zamanzadeh V, Ghahramanian A, Rassouli M, et al. Design and implementation content validity study: Development of an instrument for measuring patient-centered communication. *Journal of Caring Sciences* 2015;**4**(2):165
27. Lacasse Y, Godbout C, Series F. Health-related quality of life in obstructive sleep apnoea. *European Respiratory Journal* 2002;**19**(3):499-503
28. Lawrence I, Lin K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989:255-68
29. McBride G. A proposal for strength-of-agreement criteria for lin's concordance correlation coefficient. NIWA Client Report: HAM2005-062. Hamilton: National Institute of Water & Atmospheric Research, Ltd (NZ), 2005.
30. Kimberlin CL, Winetrstein AG. Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacy* 2008;**65**(23)
31. 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
32. Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 1960;**20**(1):37-46
33. Kwicien R, Koppsehneider A, Blettner M. Concordance analysis: Part 16 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt International* 2011;**108**(30):515
34. Bland JM, Altman DG. Statistical methods for assessing agreement between two

- 1
2
3 methods of clinical measurement. *Lancet* 1986;**327**(8476):307-10
4 35. McCabe D, Langer KG, Borod JC, et al. Practice effects. In: Kreutzer JS, DeLuca
5 J, Caplan B, eds. *Encyclopedia of clinical neuropsychology*. New York: Springer
6 2011:1988-89.
7
8 36. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*
9 1951;**16**(3):297-334
10 37. DeVellis RF. *Scale development: Theory and applications*. Los Angeles: Sage
11 2012:109-10.
12 38. Rhodes KV, Miller FG. Simulated patient studies: An ethical analysis. *Milbank*
13 *Quarterly* 2012;**90**(4):706-24
14
15 39. Mohanan M, Vera-Hernández M, Das V, et al. The know-do gap in quality of
16 health care for childhood diarrhea and pneumonia in rural india. *JAMA Pediatrics*
17 2015;**169**(4):349-57
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Figure 1. Seven sample provinces in China referencing countries of equivalent life
4 expectancy in parentheses.
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 2. VP case development team role and responsibilities.

For peer review only

1
2
3
4 Figure 3. Main components of smartphone-based VP program.
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 4. Sampling and study process for one VP case in Danzhai County, Guizhou Province (THC: township health center; VC: village clinic)

For peer review only

Table 1. The strengths and limitations of methods to assess care quality.

Process Measure of Quality	Strengths	Limitations	Assessment Level
Unannounced Standardized Patient (USP)	Standardized, controlling for case-mix and patient-mix; Unannounced, no Hawthorne effect (if not be detected); Reduced recall bias	Expensive Limited medical conditions First-visit bias Selection bias if informed consent is required	Action (Do) Gold standard
Clinical Vignettes	Cost-effective; Suitable for large scale and cross-system comparison; Covering all illnesses; Accounting for case-mix and patient-mix.	Hawthorne effect Selection bias	Competence (Know how) Over-estimating care quality (best answers) or similar
Virtual Patients (VP)	Interactive Real-time response & automatic record Highly standardized Scalability & low marginal cost Efficient delivery: anytime, anywhere Suitable for large scale study	Hawthorne effect Selection bias High cost in initial development	Performance (Show how) ?

Table 2. Main validation domains of the study.

Domain	Indicator	Data collection		Statistical analysis
		Phase	Method	
Content validity	Content validity index (CVI)	VP case review	Evaluations by an expert panel after reviewing VP cases, measured by a 4-point Likert scale (1=lowest, 4=highest).	CVI for VP case and for specific VP domain will be computed, where $CVI = \frac{\text{number of raters giving a rating of 3 or 4}}{\text{total number of raters}}$.
Feasibility	Willingness to participate; Adherence rate	Feasibility study	The subsample of clinicians' interactions with the 2 VP cases will be recorded by the online assessment	Willingness to participate = clinicians taking the VP tests/clinician being selected % Adherence rate = clinicians completed 2 VP cases/clinicians taking the VP tests %
Face validity	Satisfying score		Clinicians' subjective attitude toward the VP test experience measured by a 5-point Likert scale (1=most negative, 5=most positive).	Satisfying score for VP case and for specific aspects (e.g., usability, accessibility, etc.) will be computed, where satisfying score = frequency (%) * positive evaluations (3 to 5), and scores ≥ 1.5 are considered acceptable.
Criterion validity	Concordance correlation coefficient (r_c); Kappa statistic	Validation study	The same clinician receives a USP visit and a VP test for a matching condition. The USP-clinician interaction is evaluated by the USP using the checklist, including fees and time per visit; while VP-clinician interaction is graded by the system.	The concordance of VP-test scores against USP-test score (gold standard) or two-repeated VP-tests will be examined by r_c for continuous process quality scores, fees charged (yuan), and time spent (min); and Kappa for dichotomous diagnoses and treatment & management measures.
Test-retest Reliability			Repeat VP-tests on the same clinician in a month	
Internal consistency	Cronbach's alpha coefficient (α)		VP-test scores on a single occasion	Intercorrelation of scores for process quality indicators with $\alpha > 0.7$ is acceptable.

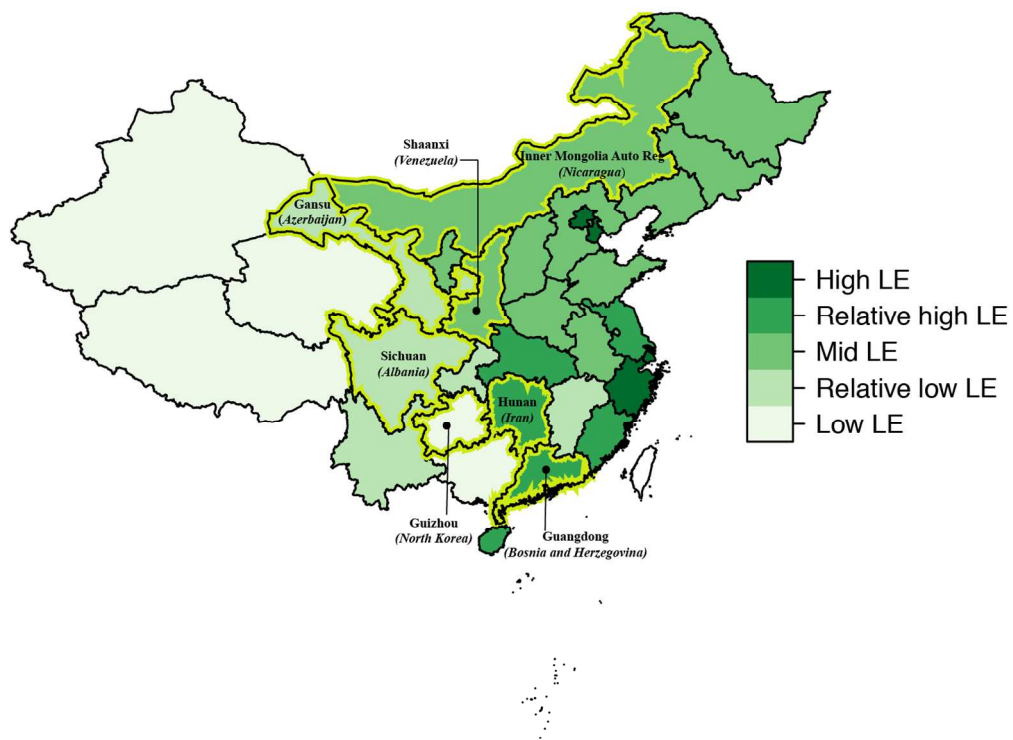


Figure 1. Seven sample provinces in China referencing countries of equivalent life expectancy in parentheses.

133x98mm (300 x 300 DPI)

www only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

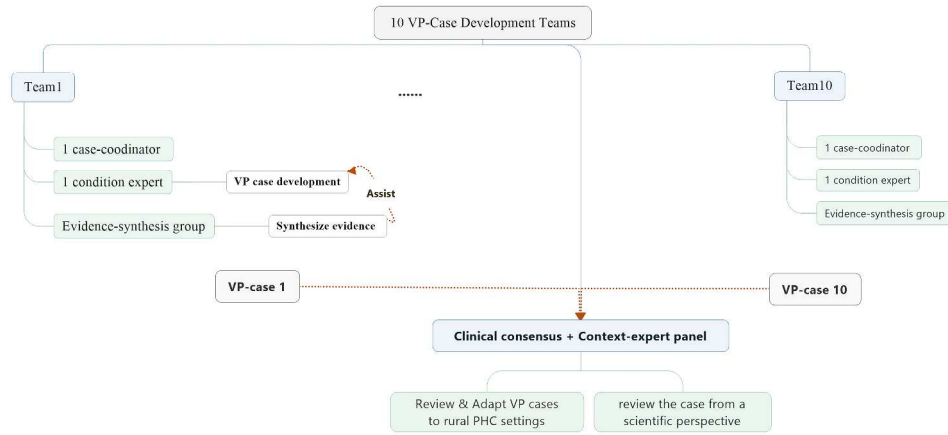


Figure 2. VP case development team role and responsibilities.

654x317mm (300 x 300 DPI)

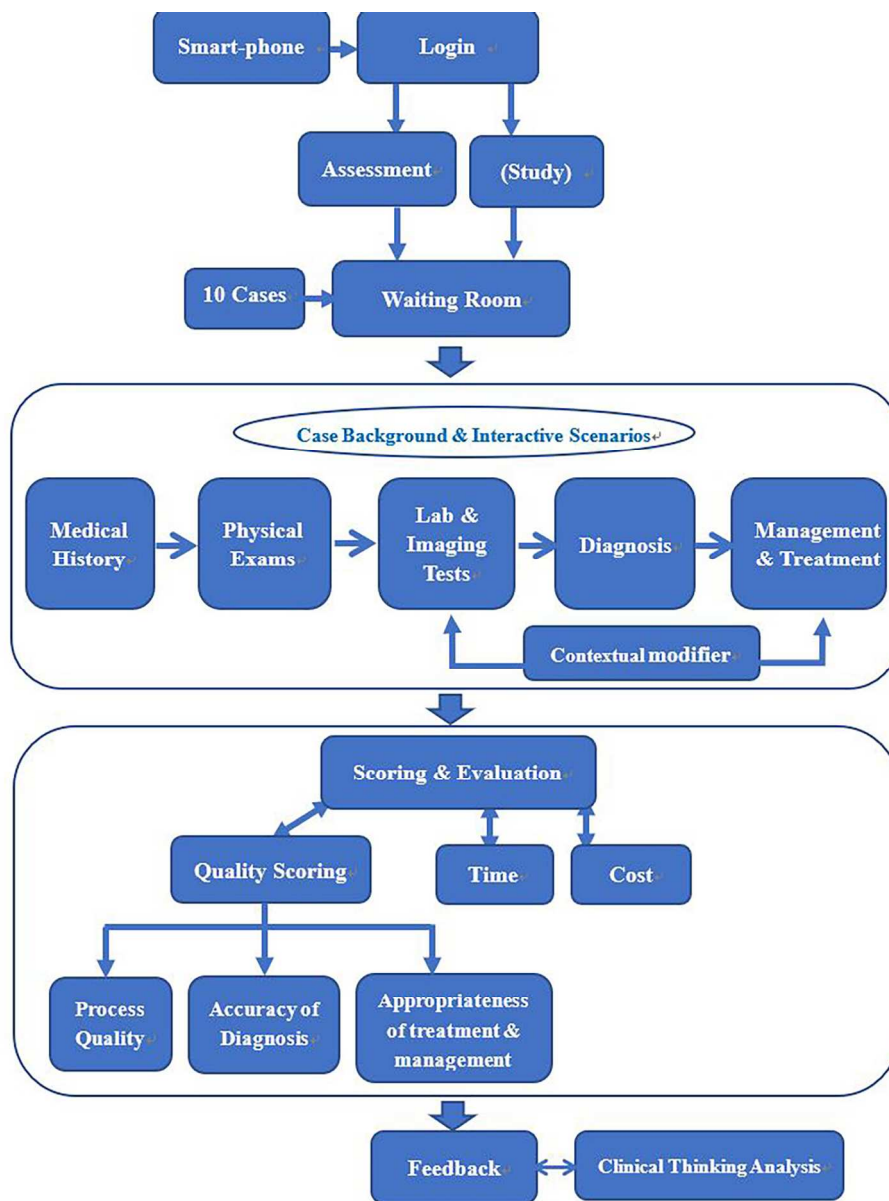
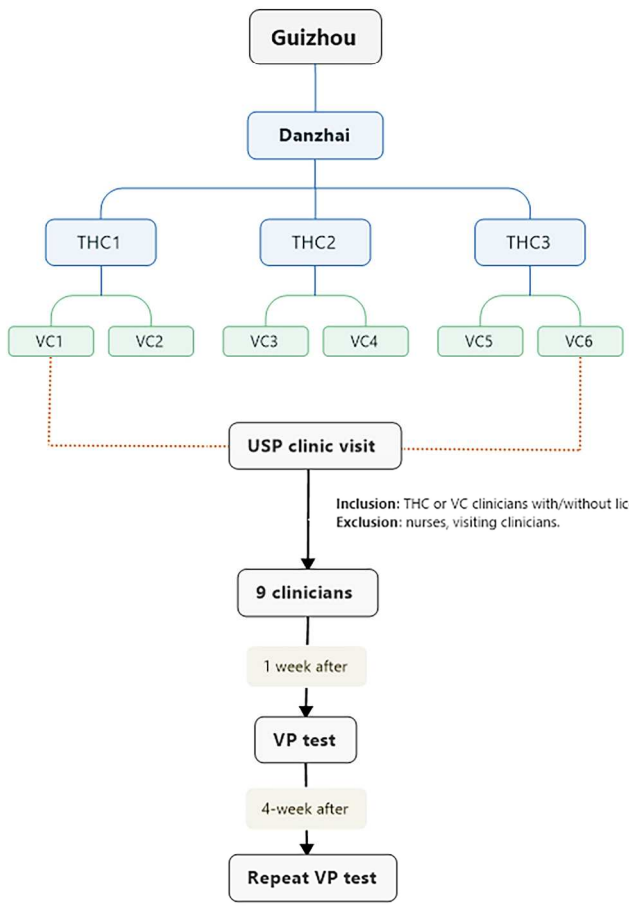


Figure 3. Main components of smartphone-based VP program. (VP: virtual patient)

158x203mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



Sampling and study process for one VP case in Danzhai County, Guizhou Province (THC: township health center; VC: village clinic)

275x269mm (300 x 300 DPI)



Appendix 1. Top 30 conditions of high-frequency clinical encounters in primary health care settings in rural China.

Clinical condition	Two-week consultation constituent ratio ¹		Township health center ²		Village clinics ²	
	RANK	%	RANK	%	RANK	%
Cold	1	28	1	13.60	1	19.50
Hypertension	2	21.8	4	7.90	6	9.80
Diabetes mellitus	4	3.9	7	4.60	8	4.60
Chronic tracheitis	8	2	2	9.50	4	10.50
Acute tracheitis			3	9.00	3	10.70
Gastritis	3	5.5	5	7.50	5	10.30
Diarrhea			6	5.30	2	11.70
Urinary tract infection			8	4.50	9	2.90
Osteoarthritis	7	2.30	17	2.40	18	0.50
Low back pain	5	3.10	16	2.50	15	0.80
Psoatic strain			14	2.60	14	0.80
Peptic ulcer			11	2.90	11	1.70
General trauma			10	3.10	13	1.10
Sciatica			19	1.80	22	0.30
Child dyspepsia			9	3.20	7	7.00
Pelvic inflammatory disease			12	2.70	16	0.60
Vaginitis			13	2.70	17	0.50
Dysmenorrhoea			18	2.30	19	0.50
Cholecystitis			15	2.60	12	1.60
Toothache	10	1.30	22	1.30	10	2.70
Menopausal syndrome			21	1.40	27	0.10
Cholelithiasis			20	1.60	20	0.40
Idiopathic headache	6	2.50	25	0.60	25	0.20
Hemorrhoids			23	1.10	21	0.40
Asthma			28	0.60	26	0.20
Chronic dermatitis			29	0.20	29	0.10
Tympanitis			24	0.70	24	0.20
Conjunctivitis			27	0.60	28	0.10
Sinusitis			26	0.60	23	0.30
Ischemic heart disease	9	1.5				

¹ Self-reported two-week consultation constituent ratio by community dwellers, information from the 2013 National Health Service Survey in China.

² Clinicians reported common clinical conditions in primary health care centers by centers' type.

Appendix 2. Methods for checklist and standards development

To evaluate the quality of care in primary health care institutions, key diagnosis and treatment points of common and frequently-occurring diseases will be developed. The *WHO Handbook for Guideline Development* and evidence-based evaluation principles will be adopted. The main procedures are comprised in the following six steps that will be implemented.

1. Expert group recruitment: Convene a multidisciplinary group consisting of experts in public health, evidence-based medicine/document retrieval, as well as clinical physicians.
2. Data retrieval and literature evaluation: Employ a 5S model to retrieve and incorporate clinical practice guidelines, textbooks, systematic reviews, meta-analysis, and important literature reviews. Retrieve literature from Wanfang, Medlive, MEDLINE, National Guideline Clearinghouse (NGC), National Institute for Health and Care Excellence (NICE), World Health Organization (WHO), DynaMed, and UpToDate. Evaluate the literature using AGREE II, AMSTAR, and QUADAS-2 for the included clinical practice guidelines, systematic reviews, and diagnostic tests, respectively.
3. Preliminary items pool development: Extract essential diagnostic and treatment procedures from the high-quality literature attained.
4. Clinical expert consensus: Apply a 2- to 3-round Delphi method to achieve consensus for diagnosis and treatment. The importance, necessity, and feasibility of the items should be considered in the process of Delphi, and additional medical information must be supplemented in terms of the clinical practice. Furthermore, all items should be classified as: necessary (3 points), selective (2 points), irrelevant (1 points), and erroneous (0 points).
5. Pilot and revise: Conduct a pilot test among 2~3 primary health care settings using the preliminary items. Revise the items and finalize the key diagnosis and treatment point evaluation items.

- 1
 - 2
 - 3
 - 4
 - 5
 - 6
 - 7
 - 8
 - 9
 - 10
 - 11
 - 12
 - 13
 - 14
 - 15
 - 16
 - 17
 - 18
 - 19
 - 20
 - 21
 - 22
 - 23
 - 24
 - 25
 - 26
 - 27
 - 28
 - 29
 - 30
 - 31
 - 32
 - 33
 - 34
 - 35
 - 36
 - 37
 - 38
 - 39
 - 40
 - 41
 - 42
 - 43
 - 44
 - 45
 - 46
 - 47
 - 48
 - 49
 - 50
 - 51
 - 52
 - 53
 - 54
 - 55
 - 56
 - 57
 - 58
 - 59
 - 60
6. Script development: Develop the script of the target disease based on key diagnosis and treatment point evaluation items before conducting the quality of service evaluation in primary health care institutions.

For peer review only

Appendix 3. Demonstrations of *Cure-Fun* smartphone-based platform current configurations of interview, physical exam and lab texts, and treatment.



BMJ Open

Using smartphone-based virtual patients to assess the quality of primary health care in rural China: protocol for a prospective multicenter study

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2017-020943.R1
Article Type:	Protocol
Date Submitted by the Author:	03-Mar-2018
Complete List of Authors:	Liao, Jing; Sun Yat-sen University, Department of Epidemiology and Biostatistics, School of Public Health; Sun Yat-sen University, Sun Yat-sen Global Health Institute, Institute of State Governance Chen, Yao-Long; Lanzhou University, Evidence-Based Medicine Center, School of Basic Medical Sciences Cai, Yiyuan; Sun Yat-sen University, Department of Epidemiology and Biostatistics, School of Public Health; Guizhou Medical University, School of Public Health Zhang, Nan; Inner Mongolia Medical University, Department of Health Management, School of Health Management 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 Wang, Hong; Bill & Melinda Gates Foundation, Health Economics, Financing & Systems Wasserheit, Judith; University of Washington, Departments of Global Health, Medicine, and Epidemiology, Schools of Medicine and Public Health Gong, Wenjie; Central South University, School of Public Health; University of Rochester Medical Center, Department of Psychi Zhou, Zhongliang; Xi'an Jiaotong University, Pan, Jay; Sichuan University, West China School of Public Health Wang, Xioahui; Lanzhou University, School of Public Health Tang, Chengxiang; Guangzhou University, School of Public Administration Zhou, Wei; Xiangya Hospital Central South University, Xu, Dong; Sun Yat-sen University , Sun Yat-sen Global Health Institute, Institute of State Governance; Sun Yat-sen University, Department of Epidemiology and Biostatistics, School of Public Health
Primary Subject Heading:	Public health
Secondary Subject Heading:	Public health
Keywords:	Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, PRIMARY CARE, care quality assessment tool

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

SCHOLARONE™
Manuscripts

For peer review only

Using smartphone-based virtual patients to assess the quality of primary health care in rural China: protocol for a prospective multicenter study

Jing Liao^{1,2}, Yaolong Chen³, Yiyuan Cai⁴, Nan Zhan⁵, Sean Sylvia⁶, Kara Hanson⁷, Hong Wang⁸, Judith N Wasserheit⁹, Wenjie Gong¹⁰, Zhongliang Zhou¹¹, Jay Pan¹², Xiaohui Wang¹³, Chenxiang Tang¹⁴, Wei Zhou¹⁵, Dong Xu^{1,2*}

¹ Department of Epidemiology and Biostatistics, School of Public Health, Sun Yat-sen University, No.74 Zhongshan 2nd Road, Guangzhou, P.R. China, 510080

² Sun Yat-sen Global Health Institute, Institute of State Governance, Sun Yat-sen University, No.135 Xingang West Road, P.R. China, 510275

³ Evidence Based Medicine Center, School of Basic Medical Sciences, Lanzhou University, No. 199 Donggang West Rd, Lanzhou City, Gansu Province, 730000 China

⁴ School of Public Health, Guizhou Medical University, UniverCity of Guan New Area, Guizhou, China, 550025

⁵ Department of Health Management, School of Health Management, Inner Mongolia Medical University, Jinshan Development Zone, Hohhot, Inner Mongolia, P.R. China, 010110

⁶ 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

⁷ Department of Global Health and Development, Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine, WC1H 9SH, London, United Kingdom

⁸ Health Economics, Financing & Systems, PO Box 23350, Seattle WA, 98102, USA

⁹ Departments of Global Health, Medicine, and Epidemiology, Schools of Medicine and Public Health, University of Washington, 1705 NE Pacific Street, Box 357965 Seattle, Washington, 98195-7965, United States

¹⁰ Xiangya school of public health, Central South University, Xiangya school of public health, Central South University, China

¹¹ School of Public Policy and Administration, Xi'an Jiaotong University, No. 28 Xianning West Road, Xi'an, Shaanxi, 710049, China.

¹² West China School of Public Health, Sichuan University, No. 17, Ren Min Nan Road, Chengdu, China

¹³ School of Public Health, Lanzhou University, Lanzhou, Gansu Province, P. R. China, 730000

¹⁴ School of Public Administration, Guangzhou University, Guangzhou, Guangdong, 510320, China.

¹⁵ Hospital Administration Institute, Xinagya Hospital, Central South University, No. 87 Xiangya Road, Changsha, Hunan., China

Corresponding author*: Dong Xu, xudong5@mail.sysu.edu.cn, Department of Epidemiology and Biostatistics, School of Public Health, Sun Yat-sen University, No.74 Zhongshan 2nd Road, Guangzhou, P.R. China, 510080

ABSTRACT

Introduction: Valid and low-cost quality assessment tools are not readily available. The Unannounced Standardized Patient (USP), the gold standard for assessing quality, is restricted by a high implementation cost; while clinical vignettes, as a low-cost alternative, have been questioned by their validity. Computerized virtual patients (VPs) create high-fidelity and interactive simulations of doctor-patient encounters which can be easily implemented via smartphone at low marginal cost. Our study thus aims to develop and validate smartphone-based VP as a quality assessment tool for primary care, compared to USP.

Methods and analysis: The study will be implemented in outpatient settings of primary health centers (PHCs) in rural areas of seven Chinese provinces, and physicians practicing at township health centers and village clinics will be our study population. The development of VPs will involve three steps: (1) identifying 10 VP cases that can best represent rural PHCs' work, (2) designing each case by a case-specific development team, and (3) developing corresponding quality scoring criteria. After being externally reviewed for content validation, these VP cases will be implemented on a smartphone-based platform and will be tested for feasibility and face validity. This smartphone-based VP tool will then be validated for its criterion validity against USP and its reliability (i.e., internal consistency and stability), with 1260 VP/USP-clinician encounters across the seven study provinces for all 10 VP cases.

Ethics and dissemination: Sun Yat-sen University: No. 2017-007. Study findings will be published and the tools developed will be freely available to low- and

1
2
3 middle-income countries for research purposes.
4
5

6 Strengths and limitations of this study
7

- 8 ● Developing and validating smartphone-based VP as a quality assessment tool
9 for research and routine use in rural primary health care centers.
- 10 ● Following an evidence-based approach to develop VP cases and scoring
11 criteria.
- 12 ● Systematically validating the VP assessment tool via a cross-national
13 multicenter study
- 14 ● The extent to which the VP assessment can reflect practitioners' real clinical
15 practice needs to be verified.
16
17

18
19 **Key words:** Quality in health care, primary care, care quality assessment tool
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

INTRODUCTION

Universal health coverage (UHC) is a paramount goal of health system development for countries at all income levels.¹ The achievement of UHC is not possible without primary health care services,¹ which ensure integrated care close to the population they serve and link to health-related sustainable development goals.² However, service coverage alone cannot improve health outcomes if the quality of care is poor. Despite the fact that great emphasis has been made to enhance health care services, a dearth of scientific evidence exists on the quality of primary health care in resource-poor settings, particularly of low- and middle- income countries (LMICs).³⁻⁶

This scarcity of evidence may partially result from the limited availability of valid, low cost, and easy-to-implement quality assessment tools⁷. As defined by Donabedian's framework for health care quality, quality can be evaluated by the *structure* of care (e.g., staff, equipment), the *process* of care delivery (e.g., doctor-patient interactions), and health *outcomes* (e.g., death or complications).⁸ Process measures have been increasingly adopted, considering the benefits of frequent and timely evaluation, and the usefulness in improving practice.^{9 10} The 'gold standard' of assessing process is the unannounced standardized patient (USP), namely a trained actor who simulates the symptoms, signs, and emotions of a real patient in a standardized fashion and presents him- or herself unannounced to clinics to assess care quality.¹¹ USP can reduce recall bias better than patient exit interviews, minimize the Hawthorne effect that inevitably occurs in direct observation, and allow for comparisons between users as case- and patient-mix are controlled.^{3 9 11} Nonetheless,

1
2
3 the USP can only portray a limited number of conditions without obvious
4 physiological symptoms and risk of invasive examinations. Also, training and
5 implementation of USP can require substantial personnel and resources, making USP
6 impractical for large-scale and routine quality assessment.^{12 13}
7
8
9
10
11
12

13 On the other hand, clinical vignettes or case simulations have been widely used
14 as a low-cost and convenient alternative to assess care quality.^{9 14} Vignettes have been
15 implemented in a paper-and-pencil form,⁹ presented by an enumerator,⁵ and
16 streamlined by a computer.¹⁴ The validity of the vignette in assessing the quality of
17 patient care has gotten mixed reports. Some studies showed that vignettes only reflect
18 clinicians' competency (know-how) rather than their actual behaviors and can cause
19 overestimation of clinical performance.^{9 15} By contrast, other studies found that
20 vignette-based results, particularly those streamlined by computer, are quite close to
21 the USP-based assessment.^{14 16} The enumerator-administered vignette is similar to the
22 announced standardized patient and thus is expensive and difficult to implement.⁵ A
23 computerized vignette can be interactive and can better represent the realistic
24 complexity of a clinical encounter.¹⁴ As a further improvement on computerized
25 vignettes,¹⁴ smartphone virtual patients (VPs) create high-fidelity, visualized, and
26 interactive simulations that replicate clinical complexity and can be easily
27 implemented at only marginal cost.¹⁷ Although VPs cannot remove the Hawthorne
28 effect, their advanced features may reduce the measurement gap between competency
29 and real practice.^{10 14} While VPs have been used in medical education to train and test
30 clinical skills such as clinical reasoning, making diagnoses, and therapeutic
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 decisions,¹⁸ the relative validity of their use as a measure for the quality of care has
4 yet to be studied. Strengths and limitations of the abovementioned three methods are
5
6 compared in Table 1.
7
8
9

10
11 In the present study, we propose to adapt smartphone-based VP for medical
12 education as a quality-of-care assessment tool, given its advantages in 1)
13 **standardization** (VPs are highly standardized, ensuring consistent assessments across
14 users), 2) **flexibility** (Assessments can be delivered by smartphones for multiple users
15 at anytime, anywhere, providing the data connectivity is available), 3) **scalability**
16 (VPs can be modified to demonstrate and assess almost any clinical conditions with
17 low marginal cost), and 4) **training** (VPs can further be used as a training tool to
18 improve health care quality and thus to address the ‘so what’ question after quality
19 assessment). These characteristics may especially benefit quality assessment and
20 improvement in primary care of rural communities which are geographically scattered
21 and difficult to reach and manage.
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37

38 Therefore, our study aims to develop and validate **smartphone-based VPs**
39 against USPs as a quality assessment tool that can not only be used for research
40 purposes but also routinely applied to evaluate the quality of primary health care
41 provided by primary health centers (PHCs) in rural areas. To maximize its validity,¹⁴
42 we will properly construct high-fidelity VP cases to reflect clinical complexity in rural
43 PHC contexts with real-time patient-doctor interactions and temporal constraints, as
44 well as use evidence-based quality care scoring criteria; additionally, we will make
45 the VP-based test anonymous to minimize the Hawthorne effect. The initial phase of
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 the study will mainly focus on rural China, while the ultimate goal is to develop and
4
5
6 validate tools that can have a broad application in other LMICs.
7

8 METHODS AND ANALYSIS 9

10 **Study setting**

11
12 The validation study will be implemented in the outpatient setting of rural PHCs (i.e.,
13
14 township health centers and village clinics) in seven Chinese provinces (Guizhou,
15
16 Sichuan, Gansu, Inner Mongolia, Shaanxi, Hunan, and Guangdong). We are selecting
17
18 these provinces not only to reflect the five strata of low-to-high life-expectancies and
19
20 various burden-of-disease patterns in China,¹⁹ but also to contrast geographic regions
21
22 with ethnic diversities, from southwest mountainous regions, to the northern plateau,
23
24 middle inland region, and southeast coastal areas (**Error! Reference source not**
25
26 **found.**). Our study targets township health centers and village clinics because they
27
28 provide most primary health care in rural China.^{20 21} At township health centers,
29
30 primary health care is delivered by a workforce including licensed/unlicensed
31
32 physicians, licensed/unlicensed assistant physicians, and registered nurses; while at
33
34 village clinics, services are mainly delivered by one full- or part-time ‘village doctor’
35
36 who is a clinician with rudimentary medical training.^{20 22 23} The outpatient setting is
37
38 chosen due to the few inpatient cases in township health centers and village clinics.
39
40
41
42
43
44
45
46
47 The study recruitment is expected to start from April 2018.
48
49

50 **VP case development**

51 *VP case selection*

52
53
54 We intend to select 10 cases that can best represent the work of rural PHCs. The
55
56
57

1
2
3 selection of the VP cases will be based on the following criteria: 1) high frequency of
4 clinical encounters in the primary care settings in rural areas, and/or 2) association with
5 significant disease burden; 3) representation of the major areas of work of PHCs in
6 rural China overall (e.g., public health service delivery, chronic disease management,
7 infectious disease control, health education, and patient-centered care); and 4)
8 suitability with USP methodology (e.g., no obvious physiological signs, low risk for
9 invasive tests) for the sake of criterion validation in the current study. A case selection
10 committee will be comprised of stakeholders, including physicians, public health
11 practitioners, policy-makers, and members of the research team. Based on the literature
12 review, the research team will prepare a list of the 30 most frequently seen conditions in
13 township health centers and village clinics reported by either community dwellers²⁴ or
14 the rural PHC clinicians (Appendix1) for the committee to rate and select.
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

32 *VP case design*

33
34 The 10 selected VP cases will then be constructed individually by 10 case-specific
35 development teams (Figure 2). These teams consist of one *condition expert* in the
36 relevant specialties of a tertiary teaching hospital who will be responsible for drafting
37 the VP case; an *evidence-synthesis group* involving epidemiologists and
38 evidence-based researchers who will search and synthesize evidence of the selected
39 condition for the condition expert to work on; a *clinical consensus group* which
40 consists of several condition-related clinical experts who will review the
41 corresponding case from a scientific perspective; an overall all-condition shared
42 *context-expert panel*, which includes clinicians and health managers from community
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 health centers, township health centers, and village clinics, who will review the
4 contextual appropriateness of the cases in the rural PHC setting; and a *case*
5
6
7
8 *coordinator* who will coordinate development of each case.
9

10 Each VP case will be structured into five domains—medical history, physical
11 examination, laboratory and imaging studies, diagnosis, and management and
12 treatment plan—to simulate real-life clinical scenarios.^{11 18} The structured VP cases
13 will have the ability to evaluate the examinee’s performance by each domain and to
14 aggregate performance scores across conditions. Besides these five condition-related
15 domains, another practice contextual adjustor will be built into each case to consider
16 medical resource constraints in rural practices (e.g., availability of basic medical
17 equipment and medicines).
18
19
20
21
22
23
24
25
26
27
28
29

30 *Scoring criteria*

31
32 Corresponding to each VP case, care quality scoring criteria will be developed. These
33 criteria include *process quality*, the *accuracy of diagnosis*, and the *appropriateness of*
34 *the treatment and management plan*.^{3 4} Process quality will be evaluated in reference
35 to a clinical process checklist (to be detailed later) including all necessary questions
36 that should be asked and physical examinations that should be performed by clinicians,
37 alongside redundant or even potentially harmful practices. Diagnoses will be rated as
38 correct, partially correct, or incorrect based on predetermined standards. The
39 treatment & management plan will be considered appropriate if the clinician
40 prescribes any of the correct medications or refers the patient to an upper level
41 physician depending on the VP cases.
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 In addition, *cost of care* and *time-spent per encounter* will also be recorded.
4
5
6 Patient costs will cover medication fees and clinic fees charged per case. In order to
7
8 recode clinicians' reaction time to each domain and to impose temporal constraints as
9
10 in real clinical practices, the entire clinician-VP interaction process will be timed. This
11
12 will include time spent on taking history, conducting physical examinations,
13
14 prescribing drugs and treatments, and any interruptions)
15
16

17
18 As for the development of the aforementioned checklist for the predetermined
19
20 standards of correctness regarding appropriateness of the treatment and management
21
22 plan, a systematic evidence-based approach will be adopted (Appendix 2). Briefly, the
23
24 evidence-synthesis group will systematically search and extract condition-specific
25
26 checklist items and standards from clinical guidelines, reputable textbooks, and
27
28 systematic reviews, etc., in that order, whereas the quality of the evidence will then be
29
30 rated by the Appraisal of Guidelines for Research & Evaluation II (AGREE II)²⁵ or
31
32 the revised Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2)
33
34 according to its type.²⁶ Afterwards, the clinical consensus group and the
35
36 contextual-expert panel will review and revise initial standards using a Delphi
37
38 process.²⁷
39
40
41
42
43
44

45 **VP case external review**

46
47 To validate the content of VP cases, an independent expert panel of physicians,
48
49 general practitioners, and rural PHC clinicians who otherwise are not involved in the
50
51 study, will be convened to review the cases for content accuracy and appropriateness.
52
53
54 The content validation involves qualitative and quantitative phases. In the qualitative
55
56
57
58
59
60

1
2
3
4 phase, the expert panel will be required to evaluate the cases with respect to the
5
6 following: overarching assessment goal, representativeness of the goal and test items
7
8 to each domain, the logical relationship of the content tested, and the appropriate
9
10 wording, grammar, understandability, and relatedness to the rural PHC context. The
11
12 panel will also mention their suggestions, if any, next to each item. Modified VP cases
13
14 will then be given back to the expert panel for quantitative evaluation with respect to
15
16 their adjustments for simplicity and clarity, as well as necessity and relevance to the
17
18 assessment, using a four-point Likert scale ranging from 1 (the lowest) to 4 (the
19
20 highest). The content validity index (CVI) will be computed for each domain and for
21
22 the entirety of VP cases.²⁸

23 24 25 26 27 28 **Technical implementation**

29
30 Revised VP cases will be implemented on *CureFUN*, an existing smartphone- based
31
32 training platform using VPs with special customizations and set-ups to suit the
33
34 assessment purpose. A live demonstration of a simplified VP can be accessed from
35
36 http://www.curefun.com/zhiqu_front/www/experience/experience.html#/caseList and
37
38 is also illustrated in Appendix 3. The smartphone-based VP assessment tool will not
39
40 only present interactive clinical scenarios, but also automatically record each
41
42 examinee's diagnosis pathway and grade it against the scoring criteria (Figure 33).
43
44
45
46

47 48 **Feasibility study**

49
50 Before a full-scale validation study, a feasibility study with 30% of the validation
51
52 study sample (see study sample section) will be conducted to test the VP assessment
53
54 tool's usability, accessibility, and stability, particularly in remote village clinics with
55
56
57

1
2
3 weak phone connectivity. Selected clinicians will be instructed to individually attempt
4
5 two random VP cases within a given time, using their own smartphone devices from
6
7 their workplace. Clinicians without a smartphone will be given a temporary one
8
9 installed with the customized *CureFUN* applications. Clinicians' willingness to
10
11 participate and adherence to the VP-based tests (e.g., percentage completing VP cases,
12
13 score of the assessment, and number of attempts made at each case per person) will be
14
15 automatically recorded. Upon completion of the cases, participants will be asked to
16
17 fill in a five-point Likert-scale questionnaire regarding their subjective attitude toward
18
19 the simulator VP experience (with 1 being the most negative response and 5 being the
20
21 most positive), regarding ease of use, experienced assessment process and outcome,
22
23 realism, device competence, accessibility, and other general comments. These results
24
25 will be used to determine VP cases' face validity, whereby a satisfying score equals
26
27 frequency (%) multiplied by positive evaluations; and scores no less than 1.5 are
28
29 considered acceptable.²⁹

37 **Validation of VP as a quality assessment tool**

38 *Study design*

39
40 The prospective validation study is a nationwide multicenter study with two main
41
42 functions. This study will 1) assess the criterion validity of the VP-tool in assessing
43
44 the quality of primary health care, by analyzing its measurement concordance against
45
46 the quality of primary health care, by analyzing its measurement concordance against
47
48 the standard USP measure, and 2) test the reliability of the VP tool, by examining its
49
50 internal consistency and the stability of repeated VP assessments on the same
51
52 examinees.
53
54
55
56

Study sample

From each of our seven sample provinces, two counties will be selected with sufficient variations in socio-economic conditions, demographics, and disease burdens between them while also approximating the provincial condition in general. Within each county, the government registry of all township health centers and village clinics will serve as our sampling frame, which will include 1) licensed practicing physicians, 2) clinicians who have not been licensed but are providing clinical services under the supervision of licensed physicians at township health centers, as well as 3) full- or part-time village clinicians. Temporarily-visiting clinicians (often senior clinicians sent by higher level medical institutions to support the development of township health centers), nurses, and allied health workers without prescription privileges will be excluded.

The sample size calculation is based on individual VP/USP-clinician encounter, and ensures sufficient power to detect variations at individual case level per county. For village clinics, one VP/USP case will be examined per time to minimize the detection of USPs. Assuming a 5% type I error and 80% power, to determine whether a moderate concordance correlation coefficient³⁰ of 0.90³¹ between VP and USP differs from zero, seven paired VP/USP-clinician encounters will be required for each of the 10 cases per county. As a stratified sampling strategy will be deployed that first samples townships and then villages from each township, the design effect should be taken into consideration when calculating the sample size. Assuming an intra-class correlation of 0.05 and 6 village clinics per township, then 9 paired VP/USP-clinician

1
2
3 encounters is needed. These nine paired VP/USP-clinician encounters will be assigned
4
5
6 to 3 township health centers and 6 village clinics using probability proportional to size
7
8 (PPS) method. These 9 paired VP/USP-clinician encounters will be assigned to
9
10 township health centers and village clinics based on the ratio of the total number of
11
12 clinicians at township health centers over the total number of village clinicians for
13
14 each county. There are 1260 VP/USP-clinician encounters across our seven study
15
16 provinces for all 10 VP cases. Figure 4 shows the sampling process and study flow for
17
18 one VP case using Guizhou Province (Danzhai County) as a demonstration.
19
20
21

22 *Criterion validity*

23
24
25 Criterion validity³² of the VP to assess quality of care will be evaluated mainly by its
26
27 measurement concordance against the USP measure as the recognized gold standard³³
28
29 for assessing quality of care in real practice. The USPs will be developed in a related
30
31 study, sharing the development teams for VP and a similar development process. The
32
33 method of fielding USPs in rural China will follow approaches similar to those of the
34
35 previous USP study in rural China.³ Identical quality scoring criteria, described above,
36
37 will be applied to scores. Each selected clinician will first see a USP (to avoid the
38
39 practice effect due to the USP's unannounced feature), and then complete a
40
41 smartphone-based VP assessment of the same condition. The clinician to be assessed
42
43 will be randomly selected onsite by the USP from any on-duty clinicians on the day of
44
45 the USP visit to the sampled township health center and village clinics. This situation
46
47 would especially apply to township health centers, as most village clinics have only
48
49 one clinician (note: Chinese patients normally see their primary care clinicians as a
50
51
52
53
54
55
56
57
58
59
60

1
2
3 walk-in patient and appointments are seldom needed). To record USP-clinician
4 interactions, USPs will complete checklists immediately after their visit, as well as
5
6 retain the medication prescription and the fee charge slips by the clinician. A week
7
8 after the USP clinic visit, clinicians will be assigned a smartphone-based VP
9
10 assessment, which will consist of a demonstration VP case to allow the clinician
11
12 getting familiar with the operation system, and the examination VP case of the same
13
14 USP condition. The VP-clinician interactions, drugs dispensed, and fees charged will
15
16 all be recorded automatically by the online assessment system.
17
18
19
20
21
22

23 The concordance of the two assessments between USPs and VPs will then be
24
25 analyzed by Lin's concordance correlation coefficient (r_c)³⁰ for continuous *process*
26
27 *quality* scores, *fees charged* (yuan), and *time spent* (min), and the Kappa statistic³⁴ for
28
29 dichotomous *diagnoses* and *treatment & management* measures. r_c evaluates how
30
31 close pairs of observation fell on a 45° line (the perfect concordance line) through the
32
33 origin in addition to their correlation. Kappa measures agreement in assessment beyond
34
35 what is expected by chance alone. In addition, for continuous measures, a
36
37 Bland-Altman plot will also be used to visualize the concordance.^{35 36} For dichotomous
38
39 measures, we will analyze their sensitivity (i.e., strength to detect correct diagnosis,
40
41 treatment plan, etc.) and specificity (i.e., strength to detect incorrect diagnosis,
42
43 treatment plan, etc.) using USP as the reference.
44
45
46
47
48
49

50 *Reliability*

51
52 To establish *test-retest reliability*, clinicians previously being assessed by VPs will be
53
54 instructed to retake the same VP tests four weeks after their last assessment. The
55
56
57
58
59
60

1
2
3 second VP test is set one month later than the first to reduce the practice effect,³⁷
4
5 assuming the clinician's general medical knowledge remains constant.³² The
6
7 concordance of the two repeated tests indicate the stability of the VP assessment
8
9 tool.³² Similar concordance measures (i.e., r_c for continuous and Kappa for
10
11 dichotomous measures) as described above will be used. The *internal consistency*, the
12
13 intercorrelation of scores for process quality indicators, will be computed by
14
15 Cronbach's alpha coefficients (α),³⁸ with $\alpha > 0.7$ representing acceptable reliability.³⁹
16
17 Table 2 summarizes the validity, reliability, and feasibility measures that will be
18
19 examined in our study.
20
21
22
23
24

25 **Patient and public involvement**

26
27 We will seek feedback from clinicians and patient representatives in the feasibility
28
29 study and use their feedback to refine the VP cases. Our USPs are also lay people
30
31 trained to portray patients and assess care quality based on their interactions with
32
33 clinicians. Our scoring criteria thus are also patient centered. Furthermore, all
34
35 participants will be acknowledged for their involvement in the study and will be
36
37 provided with a final summary report of the study outcomes, as well as have free
38
39 access to the VP training website. All published results will be publicly available.
40
41
42
43
44

45 **ETHICS AND DISSEMINATION**

46
47 The study has been approved by the Institutional Review Board of the School of
48
49 Public Health (IRB), Sun Yat-sen University (No. 2017-007). Informed consent will
50
51 be obtained from all participating clinicians of VP tests. However, to reduce
52
53 participation bias due to self-selection,⁴⁰ our IRB has approved the implementation of
54
55
56
57
58
59
60

1
2
3
4 USP without prior informed consent from the individual participants, on the condition
5
6 that involved clinicians will be fully de-identified and all analyses will only be
7
8 conducted at the population level.⁴⁰ Study data will be securely stored and only
9
10 de-identified information will be used for analysis. We will seek peer-reviewed
11
12 publications for study findings and produce reports to inform health authorities. The
13
14 tools and technology developed in this study will be freely available to other LMICs
15
16 for research purposes.
17
18
19

20 DISCUSSION

21
22
23 To the best of our knowledge, this is the first study validating VP as a quality
24
25 assessment tool in rural primary health care centers. This study follows an
26
27 evidence-based approach to develop VP cases and scoring criteria, implements them
28
29 on a widely accessible platform (i.e., a smartphone), and systematically validates the
30
31 VP assessment tool via a cross-national multicenter study representing rural PHCs
32
33 over a wide range of geographic areas with distinct life expectancies and economic
34
35 development levels. The VP assessment tool's accessibility, flexibility and scalability
36
37 give it good potential to be easily adapted to other LMICs.
38
39
40
41

42
43 VP has mainly been used in medical education to train and test critical thinking¹⁸
44
45^{41 42}, and only till recently few studies start to extend its usage into practice setting to
46
47 change health provider behavior and improve care quality.^{43 44} As a further extension,
48
49 our study proposes to validate VP as a quality assessment tool via widely accessible
50
51 smartphones. Nevertheless, it is to be noted that given its simulated nature, the
52
53 VP-test theoretically may never completely bridge the 'competency-practice' gap. The
54
55
56
57
58
59
60

1
2
3 validation study is thus essential to quantify the concordance/discordance between
4
5
6 VP- and USP-based quality assessments. Our study will generate firsthand empirical
7
8 evidence contributing to the understanding of the ‘know-do gap’,^{5 45} and further shed
9
10 light on circumstances that cannot be tested by USPs.
11
12

13 A limitation of the study, however, is that, in order to test the validity of VP
14
15 against USP as the reference standard, we restrict the selection of VP cases to those
16
17 that can be simulated by USP as well. This conservative first step will nevertheless
18
19 allow us to examine the extent to which VP can reflect care quality, and follow-up
20
21 study will then explore the full potential of the VP in assessing quality of care. Further,
22
23 the two purposely-selected counties for each province may not represent the
24
25 provincial conditions entirely, although we will make every effort to consider
26
27 provincial representation when selecting counties. Third, while the validation study is
28
29 exclusively conducted on PHCs in rural China, the extent to which the VP assessment
30
31 tool can be transported to other LMICs remains to be evaluated. Even so, the selected
32
33 provinces from the ‘five-Chinas’ may improve the generalizability of our study
34
35 considering the comparable life expectancies of LMICs and these provinces.
36
37
38
39
40
41

42 Author’s contributions

43
44 All authors contributed to the conceptualization and design of the study. DX, JL, and
45
46 YYC conceived the initial study design, analytical methods, and composition of the
47
48 team. JL was responsible for the study concept, initial draft, and revisions. DX was
49
50 responsible for the study concept and revising the draft. YLC and XHW were
51
52 responsible for the development of the scoring criteria. SS, KH, HW, JNW, ZLZ, ZN,
53
54
55
56
57
58
59
60

1
2
3 WJG, JP, CXT and WZ provided critical review and revision to the study design. All
4
5
6 authors read and approved the final revision.
7

8 9 Funding

10
11 The study is supported by the China Medical Board through its Health Policy and
12
13 Systems Sciences Open Competition grant ‘Quality in primary health care: using
14
15 unannounced standardized patients’ (grant No.: CMB16-260, XD, PI).
16
17

18 Acknowledgement

19
20 The study will be led by the Sun Yat-sen Global Health Institute of Sun Yat-sen
21
22 University with a consortium of researchers from seven other Chinese universities,
23
24 including Central South University, Guangzhou University, Guizhou Medical
25
26 University, Inner Mongolia Medical University, Lanzhou University, Sichuan
27
28 University, and Xi’an Jiaotong University. We thank all the students from these
29
30 universities who have contributed to our project, especially Wenjun He who produced
31
32 Figure 1, Jianjian Wang who assisted with the evidence collection of the Scoring
33
34 Criteria, and Ash Harris who contributed to the VP computerization.
35
36
37
38
39

40 Competing interests

41
42 The development of the VP assessment tool is a joint project of the Sun Yat-sen
43
44 University Global Health Institute (SGHI) (representing the seven universities in
45
46 China), and *CureFUN*. However, the VP cases will be independently developed and
47
48 the validation studies will be rigorously conducted by the research team from SGHI
49
50 and the seven universities, whereas *CureFun* will technically implement the cases on
51
52 smartphones and have no influence over the study design and analysis.
53
54
55
56
57

REFERENCES

1. Kieny MP, Evans DB. Universal health coverage. *East Mediterr Health J* 2013;19(4):305-6.
2. Transforming our world: the 2030 Agenda for Sustainable Development. : United Nations- Sustainable Development Knowledge Platform, 2015.
3. 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.
4. 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.
5. Das J, Hammer J, Leonard K. The quality of medical advice in low-income countries. *The Journal of Economic Perspectives* 2008;22(2):93-114.
6. 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.
7. 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.
8. Donabedian A. Evaluating the quality of medical care. *The Milbank Quarterly* 2005;83(4):691-729.
9. Peabody JW, Luck J, Glassman P, et al. Comparison of vignettes,

1
2
3
4 standardized patients, and chart abstraction: a prospective validation study of
5
6 3 methods for measuring quality. *JAMA* 2000;283(13):1715-22.

7
8 10. Shah R, Edgar D, Evans BJ. Measuring clinical practice. *Ophthalmic and*
9
10
11 *Physiological Optics* 2007;27(2):113-25.

12
13 11. Glassman PA, Luck J, O'Gara EM, et al. Using standardized patients to
14
15 measure quality: evidence from the literature and a prospective study. *The*
16
17 *Joint Commission journal on quality improvement* 2000;26(11):644-53.

18
19 12. Collins J, Harden R. The Use of Real Patients, Simulated Patients and
20
21 Simulators in Clinical Examinations Association for Medical Education in
22
23 Europe (AMEE) Guide, 2004.

24
25 13. Triola M, Feldman H, Kalet AL, et al. A randomized trial of teaching clinical
26
27 skills using virtual and live standardized patients. *Journal General Internal*
28
29 *Medicine* 2006;21(5):424-9.

30
31 14. Peabody JW, Luck J, Glassman P, et al. Measuring the quality of physician
32
33 practice by using clinical vignettes: a prospective validation study. *Annals of*
34
35 *Internal Medicine* 2004;141(10):771-80.

36
37 15. Shah R, Edgar DF, Evans BJ. A comparison of standardised patients,
38
39 record abstraction and clinical vignettes for the purpose of measuring clinical
40
41 practice. *Ophthalmic and Physiological Optics* 2010;30(3):209-24.

42
43 16. Dresselhaus TR, Peabody JW, Luck J, et al. An evaluation of vignettes for
44
45 predicting variation in the quality of preventive care. *Journal of general internal*
46
47 *medicine* 2004;19(10):1013-18.

- 1
2
3
4 17. Ellaway R, Candler C, Greene P, et al. An architectural model for
5
6 MedBiquitous virtual patients. Baltimore: MedBiquitous, 2006.
7
- 8
9 18. Cook DA, Triola MM. Virtual patients: a critical literature review and
10
11 proposed next steps. *Med Educ* 2009;43(4):303-11. doi:
12
13 10.1111/j.1365-2923.2008.03286.x [published Online First: 2009/04/02]
14
15
- 16 19. Zhou M, Wang H, Zhu J, et al. Cause-specific mortality for 240 causes in
17
18 China during 1990–2013: a systematic subnational analysis for the Global
19
20 Burden of Disease Study 2013. *The Lancet* 2016;387(10015):251-72.
21
22
- 23 20. Babiarz KS, Miller G, Yi H, et al. China's new cooperative medical scheme
24
25 improved finances of township health centers but not the number of patients
26
27 served. *Health Affairs* 2012;31(5):1065-74.
28
29
- 30 21. Li X, Lu J, Hu S, et al. The primary health-care system in China. *The*
31
32 *Lancet* 2017;390(10112):2584-94.
33
34
- 35 22. Qian D, Pong RW, Yin A, et al. Determinants of health care demand in
36
37 poor, rural China: the case of Gansu Province. *Health Policy and Planning*
38
39 2009;24(5):324-34.
40
41
- 42 23. Yip WC, Wang H, Liu Y. Determinants of patient choice of medical provider:
43
44 a case study in rural China. *Health Policy and Planning* 1998;13(3):311-22.
45
46
- 47 24. An Analysis Report of National Health Services Survey in China. Beijing:
48
49 Center for Health Statistics and Information, NHFPC, 2013.
50
51
- 52 25. Brouwers MC, Kho ME, Browman GP, et al. AGREE II: advancing
53
54 guideline development, reporting and evaluation in health care. *Canadian*
55
56

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Medical Association Journal 2010;182(18):E839-E42.

26. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 2011;155(8):529-36.

27. Hsu CC, Sandford BA. The Delphi Technique: Making Sense Of Consensus. *Practical Assessment Research & Evaluation* 2007;26(10):289–304.

28. Zamanzadeh V, Ghahramanian A, Rassouli M, et al. Design and implementation content validity study: development of an instrument for measuring patient-centered communication. *Journal of Caring Sciences* 2015;4(2):165.

29. Lacasse Y, Godbout C, Series F. Health-related quality of life in obstructive sleep apnoea. *European Respiratory Journal* 2002;19(3):499-503.

30. Lawrence I, Lin K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989:255-68.

31. McBride G. A proposal for strength-of-agreement criteria for Lin's concordance correlation coefficient. NIWA Client Report: HAM2005-062: Hamilton: National Institute of Water & Atmospheric Research, Ltd (NZ), 2005.

32. Kimberlin CL, Winetrstein AG. Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacy* 2008;65(23)

33. Rethans JJ, Gorter S, Bokken L, et al. Unannounced standardised patients

1
2
3 in real practice: a systematic literature review. *Medical Education*
4
5
6 2007;41(6):537-49.

7
8 34. Cohen J. A coefficient of agreement for nominal scales. *Educational and*
9
10
11 *Psychological Measurement* 1960;20(1):37-46.

12
13 35. Kwicien R, Koppschneider A, Blettner M. Concordance analysis: part 16
14
15 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt*
16
17
18 *International* 2011;108(30):515.

19
20 36. Bland JM, Altman DG. Statistical methods for assessing agreement
21
22 between two methods of clinical measurement. *Lancet*
23
24
25 1986;327(8476):307-10.

26
27 37. McCabe D, Langer KG, Borod JC, et al. Practice Effects. In: Kreutzer JS,
28
29 DeLuca J, Caplan B, eds. *Encyclopedia of Clinical Neuropsychology*. New
30
31
32 York: Springer 2011:1988-89.

33
34 38. Cronbach LJ. Coefficient alpha and the internal structure of tests.
35
36
37 *Psychometrika* 1951;16(3):297-334.

38
39 39. DeVellis RF. *Scale development: Theory and applications*. Los Angeles:
40
41
42 Sage 2012:109-10.

43
44 40. Rhodes KV, Miller FG. *Simulated Patient Studies: An Ethical Analysis*.
45
46
47 *Milbank Quarterly* 2012;90(4):706-24.

48
49 41. Reliability of a virtual patient simulation as an assessment tool.

50
51
52 PHARMACOTHERAPY; 2017. WILEY 111 RIVER ST, HOBOKEN
53
54
55 07030-5774, NJ USA.

1
2
3
4 42. Urrestigundlach M, Tolks D, Kiessling C, et al. Do virtual patients prepare
5
6 medical students for the real world? Development and application of a
7
8 framework to compare a virtual patient collection with population data. *Bmc*
9
10
11 *Medical Education* 2017;17(1):174.

12
13 43. Blok AC, May CN, Sadasivam RS, et al. Virtual Patient Technology:
14
15 Engaging Primary Care in Quality Improvement Innovations. *Jmir Medical*
16
17 *Education* 2017;3(1):e3.

18
19
20 44. Mollica R, Lavelle J, Fors U, et al. Using the Virtual Patient to Improve the
21
22 Primary Care of Traumatized Refugees. *Journal of Medical Education*
23
24 2017;16(1)

25
26
27 45. Mohanan M, Vera-Hernández M, Das V, et al. The know-do gap in quality
28
29 of health care for childhood diarrhea and pneumonia in rural India. *JAMA*
30
31 *Pediatrics* 2015;169(4):349-57.
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 1. Seven sample provinces in China referencing countries of equivalent life expectancy in parentheses. (LE: life expectancy)

For peer review only

Figure 2. VP case development team role and responsibilities. (VP: virtual patient)

For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 3. Main components of smartphone-based VP program. (VP: virtual patient)

For peer review only

1
2
3
4 Figure 4. Sampling and study process for one VP case in Danzhai County, Guizhou Province
5 (THC: township health center; VC: village clinic)
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

Table 1. The strengths and limitations of methods to assess care quality.

Process Measure of Quality	Strengths	Limitations	Assessment Level
Unannounced Standardized Patient (USP)	Standardized, controlling for case-mix and patient-mix; Unannounced, no Hawthorne effect (if not be detected); Reduced recall bias	Expensive Limited medical conditions First-visit bias Selection bias if informed consent is required	Action (Do) Gold standard
Clinical Vignettes	Cost-effective; Suitable for large scale and cross-system comparison; Covering all illnesses; Accounting for case-mix and patient-mix.	Hawthorne effect Selection bias	Competence (Know how) Over-estimating care quality (best answers) or similar
Virtual Patients (VP)	Interactive Real-time response & automatic record Highly standardized Scalability & low marginal cost Efficient delivery: anytime, anywhere Suitable for large scale study	Hawthorne effect Selection bias High cost in initial development	Performance (Show how) ?

Table 2. Main validation domains of the study.

Domain	Indicator	Data collection		Statistical analysis
		Phase	Method	
Content validity	Content validity index (CVI)	VP case review	Evaluations by an expert panel after reviewing VP cases, measured by a 4-point Likert scale (1=lowest, 4=highest).	CVI for VP case and for specific VP domain will be computed, where CVI = number of raters giving a rating of 3 or 4 divided by the total number of raters.
Feasibility	Willingness to participate; Adherence rate	Feasibility study	The subsample of clinicians' interactions with the 2 VP cases will be recorded by the online assessment	Willingness to participate = clinicians taking the VP tests divided by the percentage of clinician selected Adherence rate = clinicians completed 2 VP cases divided by the percentage of clinicians taking VP tests
			Clinicians' subjective attitude toward the VP test experience measured by a 5-point Likert scale (1=most negative, 5=most positive).	Satisfying score for VP case and for specific aspects (e.g., usability, accessibility, etc.) will be computed, where satisfying score = frequency multiply by positive evaluations (3 to 5), and scores ≥ 1.5 are considered acceptable.
Face validity	Satisfying score			
Criterion validity	Concordance correlation coefficient (r_c); Kappa statistic	Validation study	The same clinician receives a USP visit and a VP test for a matching condition. The USP-clinician interaction is evaluated by the USP using the checklist, including fees and time per visit; while VP-clinician interaction is graded by the system.	The concordance of VP-test scores against USP-test score (gold standard) or two-repeated VP-tests will be examined by r_c for continuous process quality scores, fees charged (yuan), and time spent (min); and Kappa for dichotomous diagnoses and treatment & management measures.
			Repeat VP-tests on the same clinician in a month	
Test-retest Reliability				
Internal consistency	Cronbach's alpha coefficient (α)		VP-test scores on a single occasion	Intercorrelation of scores for process quality indicators with alpha > 0.7 is acceptable.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

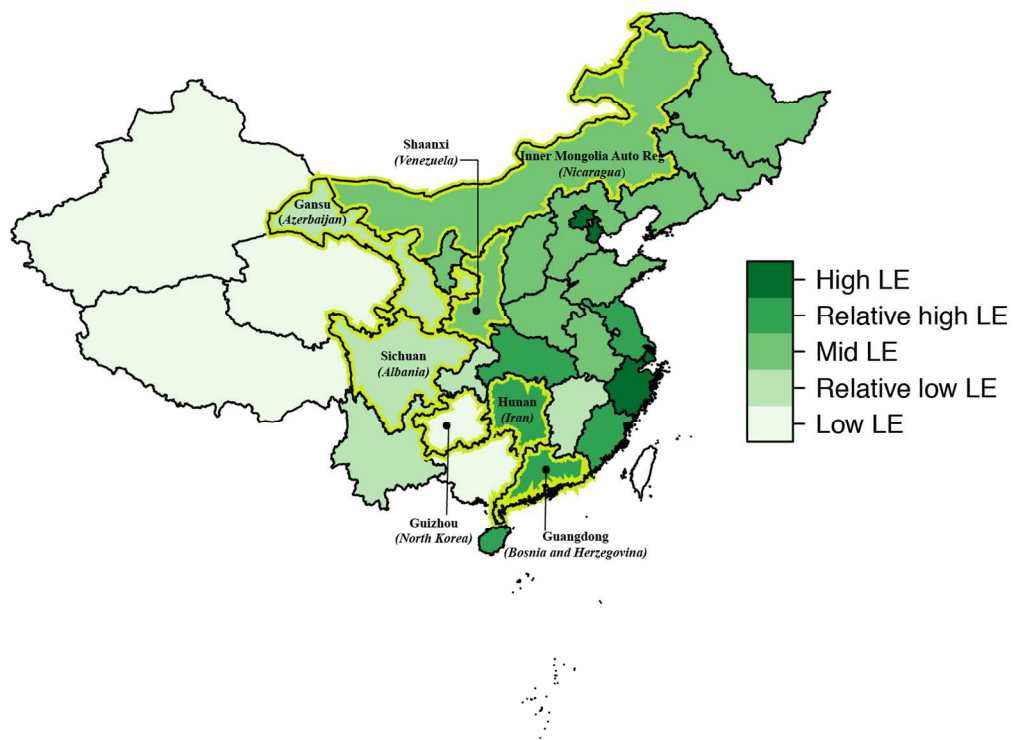


Figure 1. Seven sample provinces in China referencing countries of equivalent life expectancy in parentheses.

133x98mm (300 x 300 DPI)

www only

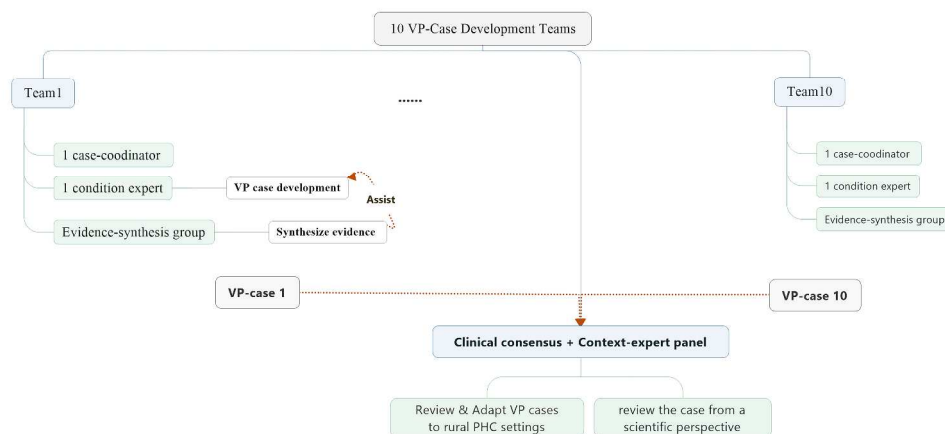
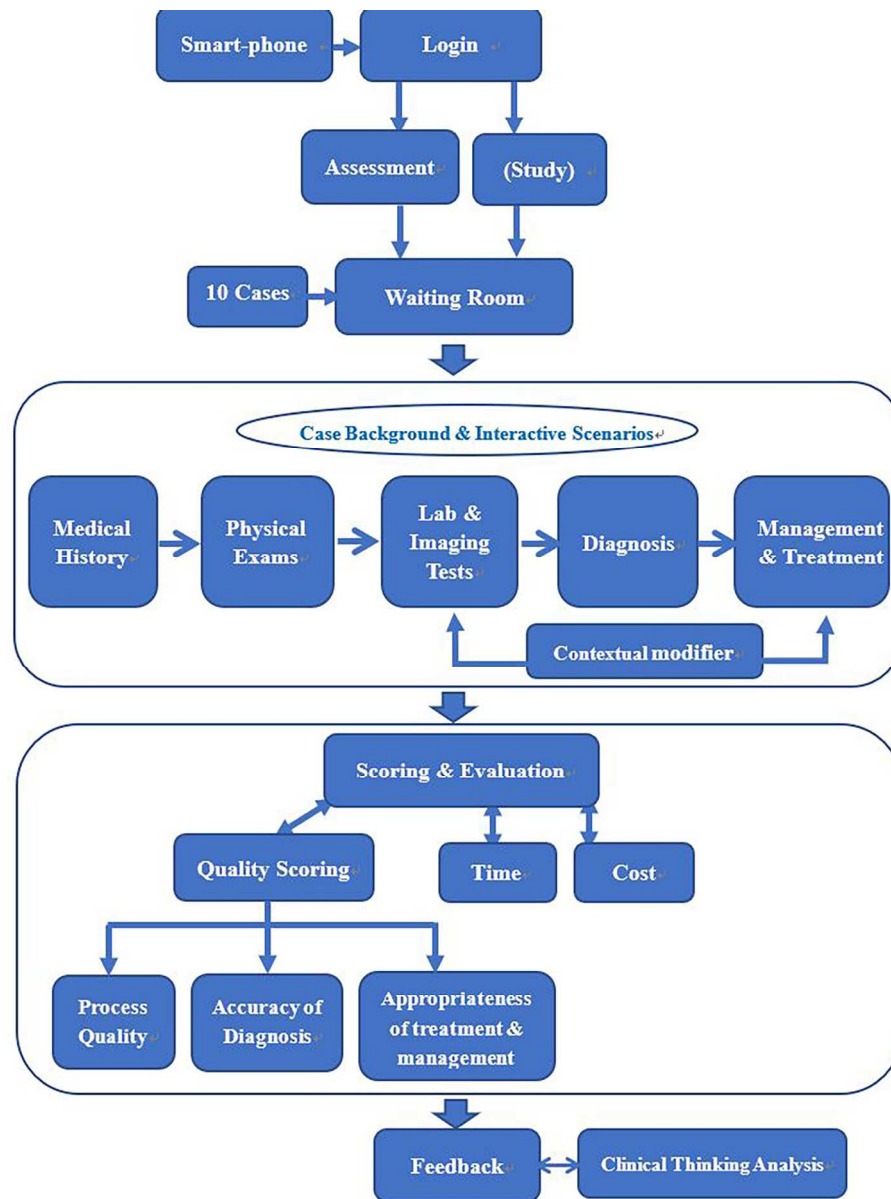


Figure 2. VP case development team role and responsibilities.

654x317mm (300 x 300 DPI)



45 Figure 3. Main components of smartphone-based VP program. (VP: virtual patient)

46 158x203mm (300 x 300 DPI)

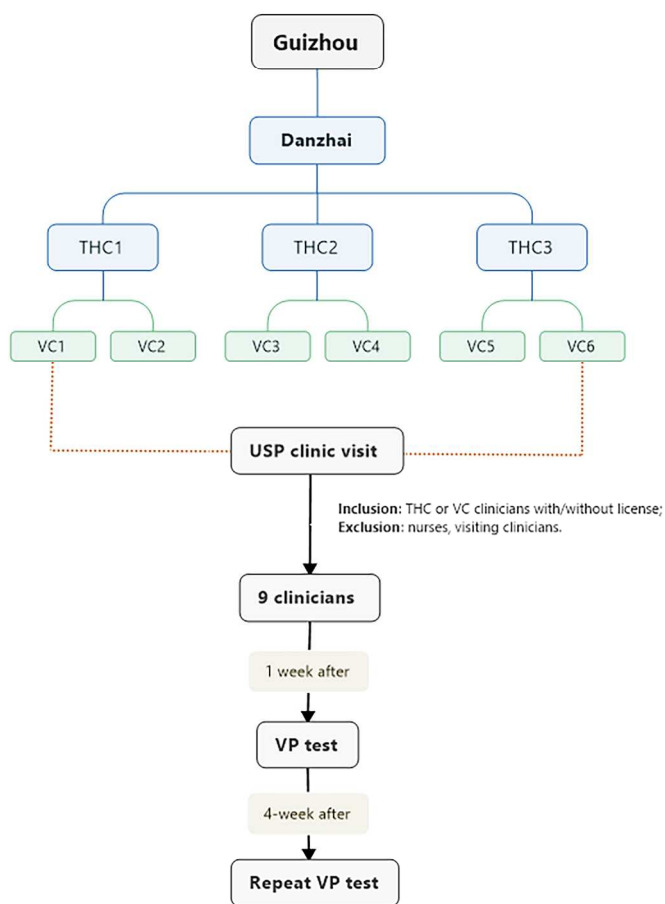


Figure 4. Sampling and study process for one VP case in Danzhai County, Guizhou Province (THC: township health center; VC: village clinic)

275x269mm (300 x 300 DPI)



Appendix 1. Top 30 conditions of high-frequency clinical encounters in primary health care settings in rural China.

Clinical condition	Two-week consultation constituent ratio ¹		Township health center ²		Village clinics ²	
	RANK	%	RANK	%	RANK	%
Cold	1	28	1	13.60	1	19.50
Hypertension	2	21.8	4	7.90	6	9.80
Diabetes mellitus	4	3.9	7	4.60	8	4.60
Chronic tracheitis	8	2	2	9.50	4	10.50
Acute tracheitis			3	9.00	3	10.70
Gastritis	3	5.5	5	7.50	5	10.30
Diarrhea			6	5.30	2	11.70
Urinary tract infection			8	4.50	9	2.90
Osteoarthritis	7	2.30	17	2.40	18	0.50
Low back pain	5	3.10	16	2.50	15	0.80
Psoatic strain			14	2.60	14	0.80
Peptic ulcer			11	2.90	11	1.70
General trauma			10	3.10	13	1.10
Sciatica			19	1.80	22	0.30
Child dyspepsia			9	3.20	7	7.00
Pelvic inflammatory disease			12	2.70	16	0.60
Vaginitis			13	2.70	17	0.50
Dysmenorrhoea			18	2.30	19	0.50
Cholecystitis			15	2.60	12	1.60
Toothache	10	1.30	22	1.30	10	2.70
Menopausal syndrome			21	1.40	27	0.10
Cholelithiasis			20	1.60	20	0.40
Idiopathic headache	6	2.50	25	0.60	25	0.20
Hemorrhoids			23	1.10	21	0.40
Asthma			28	0.60	26	0.20
Chronic dermatitis			29	0.20	29	0.10
Tympanitis			24	0.70	24	0.20
Conjunctivitis			27	0.60	28	0.10
Sinusitis			26	0.60	23	0.30
Ischemic heart disease	9	1.5				

¹ Self-reported two-week consultation constituent ratio by community dwellers, information from the 2013 National Health Service Survey in China.

² Clinicians reported common clinical conditions in primary health care centers by centers' type.

Appendix 2. Methods for checklist and standards development

To evaluate the quality of care in primary health care institutions, key diagnosis and treatment points of common and frequently-occurring diseases will be developed. The *WHO Handbook for Guideline Development* and evidence-based evaluation principles will be adopted. The main procedures are comprised in the following six steps that will be implemented.

1. Expert group recruitment: Convene a multidisciplinary group consisting of experts in public health, evidence-based medicine/document retrieval, as well as clinical physicians.
2. Data retrieval and literature evaluation: Employ a 5S model to retrieve and incorporate clinical practice guidelines, textbooks, systematic reviews, meta-analysis, and important literature reviews. Retrieve literature from Wanfang, Medlive, MEDLINE, National Guideline Clearinghouse (NGC), National Institute for Health and Care Excellence (NICE), World Health Organization (WHO), DynaMed, and UpToDate. Evaluate the literature using AGREE II, AMSTAR, and QUADAS-2 for the included clinical practice guidelines, systematic reviews, and diagnostic tests, respectively.
3. Preliminary items pool development: Extract essential diagnostic and treatment procedures from the high-quality literature attained.
4. Clinical expert consensus: Apply a 2- to 3-round Delphi method to achieve consensus for diagnosis and treatment. The importance, necessity, and feasibility of the items should be considered in the process of Delphi, and additional medical information must be supplemented in terms of the clinical practice. Furthermore, all items should be classified as: necessary (3 points), selective (2 points), irrelevant (1 points), and erroneous (0 points).
5. Pilot and revise: Conduct a pilot test among 2~3 primary health care settings using the preliminary items. Revise the items and finalize the key diagnosis and treatment point evaluation items.

- 1
 - 2
 - 3
 - 4
 - 5
 - 6
 - 7
 - 8
 - 9
 - 10
 - 11
 - 12
 - 13
 - 14
 - 15
 - 16
 - 17
 - 18
 - 19
 - 20
 - 21
 - 22
 - 23
 - 24
 - 25
 - 26
 - 27
 - 28
 - 29
 - 30
 - 31
 - 32
 - 33
 - 34
 - 35
 - 36
 - 37
 - 38
 - 39
 - 40
 - 41
 - 42
 - 43
 - 44
 - 45
 - 46
 - 47
 - 48
 - 49
 - 50
 - 51
 - 52
 - 53
 - 54
 - 55
 - 56
 - 57
 - 58
 - 59
 - 60
6. Script development: Develop the script of the target disease based on key diagnosis and treatment point evaluation items before conducting the quality of service evaluation in primary health care institutions.

For peer review only

Appendix 3. Demonstrations of *Cure-Fun* smartphone-based platform current configurations of interview, physical exam and lab texts, and treatment.



BMJ Open

Using smartphone-based virtual patients to assess the quality of primary health care in rural China: protocol for a prospective multicenter study

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2017-020943.R2
Article Type:	Protocol
Date Submitted by the Author:	16-May-2018
Complete List of Authors:	Liao, Jing; Sun Yat-sen University, Department of Epidemiology and Biostatistics, School of Public Health; Sun Yat-sen University, Sun Yat-sen Global Health Institute, Institute of State Governance Chen, Yao-Long; Lanzhou University, Evidence-Based Medicine Center, School of Basic Medical Sciences Cai, Yiyuan; Sun Yat-sen University, Department of Epidemiology and Biostatistics, School of Public Health; Guizhou Medical University, School of Public Health Zhang, Nan; Inner Mongolia Medical University, Department of Health Management, School of Health Management 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 Wang, Hong; Bill & Melinda Gates Foundation, Health Economics, Financing & Systems Wasserheit, Judith; University of Washington, Departments of Global Health, Medicine, and Epidemiology, Schools of Medicine and Public Health Gong, Wenjie; Central South University, School of Public Health; University of Rochester Medical Center, Department of Psychi Zhou, Zhongliang; Xi'an Jiaotong University, Pan, Jay; Sichuan University, West China School of Public Health Wang, Xioahui; Lanzhou University, School of Public Health Tang, Chengxiang; Guangzhou University, School of Public Administration Zhou, Wei; Xiangya Hospital Central South University, Xu, Dong; Sun Yat-sen University , Sun Yat-sen Global Health Institute, Institute of State Governance; Sun Yat-sen University, Department of Epidemiology and Biostatistics, School of Public Health
Primary Subject Heading:	Public health
Secondary Subject Heading:	Public health
Keywords:	Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, PRIMARY CARE, care quality assessment tool

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

SCHOLARONE™
Manuscripts

For peer review only

Using smartphone-based virtual patients to assess the quality of primary health care in rural China: protocol for a prospective multicenter study

Jing Liao^{1,2}, Yaolong Chen³, Yiyuan Cai⁴, Nan Zhan⁵, Sean Sylvia⁶, Kara Hanson⁷, Hong Wang⁸, Judith N Wasserheit⁹, Wenjie Gong¹⁰, Zhongliang Zhou¹¹, Jay Pan¹², Xiaohui Wang¹³, Chenxiang Tang¹⁴, Wei Zhou¹⁵, Dong Xu^{1,2*}

¹ Department of Epidemiology and Biostatistics, School of Public Health, Sun Yat-sen University, No.74 Zhongshan 2nd Road, Guangzhou, P.R. China, 510080

² Sun Yat-sen Global Health Institute, Institute of State Governance, Sun Yat-sen University, No.135 Xingang West Road, P.R. China, 510275

³ Evidence Based Medicine Center, School of Basic Medical Sciences, Lanzhou University, No. 199 Donggang West Rd, Lanzhou City, Gansu Province, 730000 China

⁴ School of Public Health, Guizhou Medical University, UniverCity of Guan New Area, Guizhou, China, 550025

⁵ Department of Health Management, School of Health Management, Inner Mongolia Medical University, Jinshan Development Zone, Hohhot, Inner Mongolia, P.R. China, 010110

⁶ 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

⁷ Department of Global Health and Development, Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine, WC1H 9SH, London, United Kingdom

⁸ Health Economics, Financing & Systems, PO Box 23350, Seattle WA, 98102, USA

⁹ Departments of Global Health, Medicine, and Epidemiology, Schools of Medicine and Public Health, University of Washington, 1705 NE Pacific Street, Box 357965 Seattle, Washington, 98195-7965, United States

¹⁰ Xiangya school of public health, Central South University, Xiangya school of public health, Central South University, China

¹¹ School of Public Policy and Administration, Xi'an Jiaotong University, No. 28 Xianning West Road, Xi'an, Shaanxi, 710049, China.

¹² West China School of Public Health, Sichuan University, No. 17, Ren Min Nan Road, Chengdu, China

¹³ School of Public Health, Lanzhou University, Lanzhou, Gansu Province, P. R. China, 730000

¹⁴ School of Public Administration, Guangzhou University, Guangzhou, Guangdong, 510320, China.

¹⁵ Hospital Administration Institute, Xinagya Hospital, Central South University, No. 87 Xiangya Road, Changsha, Hunan., China

Corresponding author*: Dong Xu, xudong5@mail.sysu.edu.cn, Department of Epidemiology and Biostatistics, School of Public Health, Sun Yat-sen University, No.74 Zhongshan 2nd Road, Guangzhou, P.R. China, 510080

ABSTRACT

Introduction: Valid and low-cost quality assessment tools examining care quality are not readily available. The Unannounced Standardized Patient (USP), the gold standard for assessing quality, is costly to implement while the validity of clinical vignettes, as a low-cost alternative, has been challenged. Computerized virtual patients (VPs) create high-fidelity and interactive simulations of doctor-patient encounters which can be easily implemented via smartphone at low marginal cost. Our study aims to develop and validate smartphone-based VP as a quality assessment tool for primary care, compared to USP.

Methods and analysis: The study will be implemented in primary health centres (PHCs) in rural areas of seven Chinese provinces, and physicians practicing at township health centers and village clinics will be our study population. The development of VPs involves three steps: (1) identifying 10 VP cases that can best represent rural PHCs' work, (2) designing each case by a case-specific development team, and (3) developing corresponding quality scoring criteria. After being externally reviewed for content validity, these VP cases will be implemented on a smartphone-based platform and will be tested for feasibility and face validity. This smartphone-based VP tool will then be validated for its criterion validity against USP and its reliability (i.e., internal consistency and stability), with 1260 VP/USP-clinician encounters across the seven study provinces for all 10 VP cases.

Ethics and dissemination: Sun Yat-sen University: No. 2017-007. Study findings will be published and tools developed will be freely available to low- and middle-income

1
2
3 countries for research purposes.
4
5

6 Strengths and limitations of this study
7

- 8 ● Developing and validating smartphone-based VP as a quality assessment tool
9 for research and routine use in rural primary health care centers.
- 10 ● Following an evidence-based approach to develop VP cases and scoring
11 criteria.
- 12 ● Systematically validating the VP assessment tool via a cross-national
13 multicenter study
- 14 ● The extent to which the VP assessment will reflect practitioners' real clinical
15 practice needs to be verified.
16
17

18
19 **Key words:** Quality in health care, primary care, care quality assessment tool
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

INTRODUCTION

Universal health coverage (UHC) is a paramount goal of health system development for countries at all income levels.¹ The achievement of UHC is not possible without primary health care services,¹ which ensure integrated care close to the population they serve and link to the health-related sustainable development goals.² However, service coverage alone cannot improve health outcomes if the quality of care is poor. Despite efforts devoted to improving health care services, there is a lack of scientific evidence on the quality of primary health care in resource-poor settings, particularly of low- and middle- income countries (LMICs).³⁻⁶

This scarcity of evidence may partially result from the limited availability of valid, low cost, and easy-to-implement quality assessment tools⁷. As defined by Donabedian's framework, health care quality can be evaluated by the *structure* of care (e.g., staff, equipment), the *process* of care delivery (e.g., doctor-patient interactions), and health *outcomes* (e.g., death or complications).⁸ Increasingly, process measures are being used, because of their advantages in terms of frequent and timely evaluation, and the usefulness in improving practice.^{9 10} The 'gold standard' of assessing process is the unannounced standardized patient (USP), namely a trained actor who simulates the symptoms, signs, and emotions of a real patient in a standardized fashion and presents him- or herself unannounced to clinics to assess care quality.¹¹ USP can reduce recall bias better than patient exit interviews, minimize the Hawthorne effect that inevitably occurs in direct observation, and allow for comparisons between users as case- and patient-mix are controlled.^{3 9 11} Nonetheless, the USP can only portray a

1
2
3 limited number of conditions without obvious physiological symptoms and risk of
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

limited number of conditions without obvious physiological symptoms and risk of
invasive examinations. Also, training and implementation of USP can require
substantial personnel and resources, making USP impractical for large-scale and
routine quality assessment.^{12 13}

As an alternative, clinical vignettes or case simulations have been widely used
as a low-cost and convenient method for assessing care quality.^{9 14} Vignettes have
been implemented in a paper-and-pencil form,⁹ presented by an enumerator,⁵ and
streamlined by a computer.¹⁴ Evidence of the validity of vignettes in assessing the
quality of patient care is mixed. Some studies showed that vignettes reflect clinicians'
competency (know-how) rather than their actual behaviors and can lead to
overestimation of clinical performance.^{9 15} By contrast, other studies found that
vignette-based results, particularly those streamlined by computer, are quite close to
the USP-based assessment.^{14 16} The enumerator-administered vignette is similar to the
announced standardized patient and thus is expensive and difficult to implement.⁵ A
computerized vignette can be interactive and can more realistically represent the
complexity of a clinical encounter.¹⁴ As a further improvement on computerized
vignettes,¹⁴ smartphone virtual patients (VPs) create high-fidelity, visualized, and
interactive simulations that replicate clinical complexity and can be easily
implemented at a low marginal cost.¹⁷ Although VPs cannot remove the Hawthorne
effect, their advanced features may reduce the measurement gap between competency
and actual practice.^{10 14} While VPs have been used in medical education to train and
test clinical skills such as clinical reasoning, diagnosis and therapeutic decisions,¹⁸

1
2
3 their relative validity as a measure of quality of care has yet to be studied. Strengths
4
5 and limitations of the abovementioned three methods are compared in Table 1.
6
7

8
9 In the present study, we propose to adapt smartphone-based VP for medical
10
11 education as a quality-of-care assessment tool, given its advantages in 1)
12
13 **standardization** (VPs are highly standardized, ensuring consistent assessments across
14
15 users), 2) **flexibility** (Assessments can be delivered by smartphones for multiple users
16
17 at any time, anywhere, providing data connectivity is available), 3) **scalability** (VPs
18
19 can be modified to demonstrate and assess almost any clinical conditions with low
20
21 marginal cost), and 4) **training** (VPs can also be used as a training tool to improve
22
23 health care quality and thus to address the ‘so what’ question after quality assessment).
24
25 These characteristics may especially benefit quality assessment and improvement in
26
27 rural primary care settings, where communities are geographically scattered and
28
29 difficult to reach and manage.
30
31
32
33
34

35
36 Therefore, our study aims to develop and validate **smartphone-based VPs**
37
38 against USPs as a quality assessment tool that can be used both for research purposes
39
40 and for routine evaluation of quality of primary health care provided by primary
41
42 health centers (PHCs) in rural areas. To maximize its validity,¹⁴ we will systematically
43
44 construct high-fidelity VP cases to reflect clinical complexity in rural PHC contexts
45
46 with real-time patient-doctor interactions and temporal constraints, and use
47
48 evidence-based quality scoring criteria; additionally, we will make the VP-based test
49
50 anonymous to minimize the Hawthorne effect. The initial phase of the study will
51
52 mainly focus on rural China, while the ultimate goal is to develop and validate tools
53
54
55
56
57

1
2
3 that can have a broad application in other LMICs.
4

5 6 METHODS AND ANALYSIS 7

8 9 **Study setting**

10
11 The validation study will be implemented in the outpatient setting of rural PHCs (i.e.,
12 township health centers and village clinics) in seven Chinese provinces (Guizhou,
13 Sichuan, Gansu, Inner Mongolia, Shaanxi, Hunan, and Guangdong). We are selecting
14 these provinces not only to reflect the five strata of low-to-high life-expectancies and
15 various burden-of-disease patterns in China,¹⁹ but also to contrast geographic regions
16 with diverse ethnic composition, including southwest mountainous regions, the
17 northern plateau, the middle inland region, and southeast coastal areas (**Error!**
18 **Reference source not found.**). Our study targets township health centers and village
19 clinics because they provide the majority of primary health care in rural China.^{20 21} At
20 township health centers, primary health care is delivered by a workforce including
21 licensed/unlicensed physicians, licensed/unlicensed assistant physicians, and
22 registered nurses; while at village clinics, services are mainly delivered by one full- or
23 part-time ‘village doctor’ who is a clinician with rudimentary medical training.^{20 22 23}
24
25 The outpatient setting is chosen due to the small number of inpatient cases in
26 township health centers and village clinics. Study recruitment is expected to start from
27 June 2018.

28 29 **VP case development**

30 31 *VP case selection*

32
33 We intend to select 10 cases that together can represent the work of rural PHCs. The
34
35

1
2
3 selection of the VP cases will be based on the following criteria: 1) high frequency of
4 clinical encounters in the primary care settings in rural areas, and/or 2) association with
5 significant disease burden; 3) representation of the major areas of work of PHCs in
6 rural China overall (e.g., public health service delivery, chronic disease management,
7 infectious disease control, health education, and patient-centered care); and 4)
8 suitability for the USP methodology (e.g., no obvious physiological signs, low risk for
9 invasive tests) for the sake of criterion validation in the current study. A case selection
10 committee will be comprised of a range of stakeholders, including physicians, public
11 health practitioners, policy-makers, and members of the research team. Based on the
12 literature review, the research team will prepare a shortlist of the 30 most frequently
13 seen conditions in township health centers and village clinics reported by either
14 community dwellers²⁴ or rural PHC clinicians (Appendix 1) from which the committee
15 will select.

35 *VP case design*

36
37 The 10 selected VP cases will then be constructed individually by 10 case-specific
38 development teams (Figure 2). These teams consist of one *condition expert* from the
39 relevant specialty of a tertiary teaching hospital who will be responsible for drafting
40 the VP case; an *evidence-synthesis group* involving epidemiologists and
41 evidence-based researchers who will search and synthesize evidence about the
42 selected condition for the condition expert to work on; a *clinical consensus group*
43 which consists of several condition-related clinical experts who will review the
44 corresponding case from a scientific perspective; an overall all-condition shared

1
2
3
4 *context-expert panel*, which includes clinicians and health managers from community
5
6 health centers, township health centers, and village clinics, who will review the
7
8 contextual appropriateness of the cases for the rural PHC setting; and a *case*
9
10 *coordinator* who will coordinate development of each case.
11
12

13 Each VP case will be structured into five domains—medical history, physical
14
15 examination, laboratory and imaging studies, diagnosis, and management and
16
17 treatment plan—to simulate real-life clinical scenarios.^{11 18} The structured VP cases
18
19 will permit the examinee's performance in each domain to be evaluated, and for
20
21 performance scores to be aggregated across conditions. In addition to these five
22
23 condition-related domains, another practice contextual adjustment will be built into
24
25 each case to consider medical resource constraints in rural practices (e.g., availability
26
27 of basic medical equipment and medicines).
28
29
30
31

32 *Scoring criteria*

33
34 Care quality scoring criteria will be developed for each VP case. These criteria
35
36 include *process quality*, the *accuracy of diagnosis*, and the *appropriateness of the*
37
38 *treatment and management plan*.^{3 4} Process quality will be evaluated in reference to a
39
40 clinical process checklist (to be detailed later) including all necessary questions that
41
42 should be asked and physical examinations that should be performed by clinicians,
43
44 together with redundant or even potentially harmful practices. Diagnoses will be rated
45
46 as correct, partially correct, or incorrect based on predetermined standards. The
47
48 treatment and management plan will be considered appropriate if the clinician
49
50 prescribes any of the correct medications or refers the patient to a higher-level
51
52
53
54
55
56
57
58
59
60

1
2
3 physician depending on the VP case.
4

5
6 In addition, *cost of care* and *time-spent per encounter* will also be recorded.
7
8 Patient costs will cover medication fees and clinic fees charged per case. In order to
9
10 link clinician reaction time to each domain and to impose the temporal constraints
11
12 seen in real clinical practices, the entire clinician-VP interaction process will be timed.
13
14 This will include time spent on taking history, conducting physical examinations,
15
16 prescribing drugs and treatments, and any interruptions.
17
18

19
20 A systematic evidence-based approach will be adopted to developing the scoring
21
22 checklist for the treatment and management plan, (Appendix 2). Briefly, the
23
24 evidence-synthesis group will systematically search and extract condition-specific
25
26 checklist items and standards from clinical guidelines, reputable textbooks, and
27
28 systematic reviews, etc., in that order. The quality of the evidence will then be rated
29
30 by the Appraisal of Guidelines for Research & Evaluation II (AGREE II)²⁵ or the
31
32 revised Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) according
33
34 to its type.²⁶ Afterwards, the clinical consensus group and the contextual-expert panel
35
36 will review and revise initial standards using a Delphi process.²⁷
37
38
39
40
41
42

43 **VP case external review**

44
45 To validate the content of VP cases, an independent expert panel of physicians,
46
47 general practitioners, and rural PHC clinicians who otherwise are not involved in the
48
49 study, will be convened to review the cases for content accuracy and appropriateness.
50
51
52 The content validation involves qualitative and quantitative phases. In the qualitative
53
54 phase, the expert panel will be required to evaluate the cases with respect to the
55
56
57
58
59
60

1
2
3 following: overarching assessment goal, representativeness of the goal and test items
4
5
6 to each domain, the logical relationship of the content tested, and the appropriate
7
8 wording, grammar, understandability, and relevance to the rural PHC context. The
9
10 panel will also record their suggestions, if any, next to each item. Modified VP cases
11
12 will then be given back to the expert panel for quantitative evaluation. They will be
13
14 asked to assess the cases for simplicity and clarity, as well as necessity and relevance
15
16 to the assessment, using a four-point Likert scale ranging from 1 (the lowest) to 4 (the
17
18 highest). The content validity index (CVI) will be computed for each domain and for
19
20 the entirety of VP cases.²⁸

25 **Technical implementation**

26
27 Revised VP cases will be implemented on *CureFUN*, an existing smartphone-based
28
29 training platform using VPs with special customizations and set-up to suit the
30
31 assessment purpose. A live demonstration of a simplified VP can be accessed from
32
33 http://www.curefun.com/zhiqu_front/www/experience/experience.html#/caseList and
34
35 is also illustrated in Appendix 3. The smartphone-based VP assessment tool will not
36
37 only present interactive clinical scenarios but will also automatically record each
38
39 examinee's diagnosis pathway and grade it against the scoring criteria (Figure 33).

45 **Feasibility study**

46
47 Before a full-scale validation study, a feasibility study with 30% of the validation
48
49 study sample (see study sample section) will be conducted to test the VP assessment
50
51 tool's usability, accessibility, and stability, particularly in remote village clinics with
52
53 weak phone connectivity. Selected clinicians will be instructed to individually attempt
54
55

1
2
3 two random VP cases within a given time, using their own smartphone devices from
4
5 their workplace. Clinicians without a smartphone will be given a temporary device on
6
7 which the customized *CureFUN* applications will be pre-installed. Clinicians'
8
9 willingness to participate and adherence to the VP-based tests (e.g., percentage
10
11 completing VP cases, score of the assessment, and number of attempts made at each
12
13 case per person) will be automatically recorded. Upon completion of the cases,
14
15 participants will be asked to fill in a five-point Likert-scale questionnaire regarding
16
17 their subjective attitude toward the simulator VP experience (with 1 being the most
18
19 negative response and 5 being the most positive), regarding ease of use, their
20
21 experience of the assessment process and outcome, realism, device competence,
22
23 accessibility, and other general comments. These results will be used to determine the
24
25 face validity of the VP cases, with scores calculated by multiplying frequency (%) by
26
27 positive evaluations (3 to 5); and scores no less than 1.5 are considered acceptable.²⁹
28
29
30
31
32
33
34

35 **Validation of VP as a quality assessment tool**

36 *Study design*

37
38 The prospective validation study is a nationwide multicenter study with two main
39
40 purposes: 1) to assess the criterion validity of the VP-tool in assessing the quality of
41
42 primary health care, by analyzing its measurement concordance against the standard
43
44 USP measure, and 2) to test the reliability of the VP tool, by examining its internal
45
46 consistency and the stability of repeated VP assessments on the same subjects.
47
48
49
50
51

52 *Study sample*

53
54 From each of our seven sample provinces, two counties will be selected with
55
56
57

1
2
3 sufficient variations in socio-economic conditions, demographics, and disease burdens
4
5
6 between them while also approximating the provincial condition in general. Within
7
8 each county, the government registry of all township health centers and village clinics
9
10 will serve as our sampling frame, which will include 1) licensed practicing physicians,
11
12
13 2) clinicians who have not been licensed but are providing clinical services under the
14
15 supervision of licensed physicians at township health centers, as well as 3) full- or
16
17 part-time village clinicians. Clinicians visiting on a temporary basis (often senior
18
19 clinicians sent by higher level medical institutions to support the development of
20
21 township health centers), nurses, and allied health workers without prescription
22
23 privileges will be excluded.
24
25
26

27
28 The sample size calculation is based on individual VP/USP-clinician encounter
29
30 and ensures sufficient power to detect variations at individual case level per county.
31
32 For village clinics, one VP/USP case will be examined at a time to minimize the
33
34 detection of USPs. Assuming a 5% type I error and 80% power, to determine whether
35
36 a moderate concordance correlation coefficient³⁰ of 0.90³¹ between VP and USP
37
38 differs from zero, seven paired VP/USP-clinician encounters will be required for each
39
40 of the 10 cases per county. As a stratified sampling strategy will be deployed that first
41
42 samples townships and then villages from each township, sample size calculations
43
44 need to take into account the design effect. Assuming an intra-class correlation of 0.05
45
46 and 6 village clinics per township, then 9 paired VP/USP-clinician encounters is
47
48 needed. These nine paired VP/USP-clinician encounters will be assigned to 3
49
50 township health centers and 6 village clinics using probability proportional to size
51
52
53
54
55
56
57

1
2
3 (PPS). These 9 paired VP/USP-clinician encounters will be assigned to township
4 health centers and village clinics based on the ratio of the total number of clinicians at
5 township health centers to the total number of village clinicians for each county.
6
7
8
9
10
11 There are 1260 VP/USP-clinician encounters across our seven study provinces for all
12
13 10 VP cases. Figure 4 shows the sampling process and study flow for one VP case
14
15 using Guizhou Province (Danzhai County) as an example.
16

17 *Criterion validity*

18
19
20 Criterion validity³² of the VP to assess quality of care will be evaluated primarily by
21
22 its measurement concordance against the USP measure as the recognized gold
23
24 standard³³ for assessing quality of care in practice. The USPs will be developed in a
25
26 related study, sharing the development teams for VP and a similar development
27
28 process. The method of fielding USPs in rural China will follow a similar approach to
29
30 those of the previous USP study in rural China.³ Identical quality scoring criteria,
31
32 described above, will be applied to scores. Each selected clinician will first see a USP
33
34 (to avoid the practice effect due to the USP's unannounced feature), and then
35
36 complete a smartphone-based VP assessment of the same condition. The clinician to
37
38 be assessed will be randomly selected onsite by the USP from any on-duty clinicians
39
40 on the day of the USP visit to the sampled township health center and village clinics.
41
42
43
44
45 This situation would especially apply to township health centers, as most village
46
47 clinics have only one clinician (note: Chinese patients normally see their primary care
48
49 clinicians as a walk-in patient and appointments are seldom needed). To record
50
51 USP-clinician interactions, USPs will complete checklists immediately after their visit,
52
53
54
55
56
57
58
59
60

1
2
3 and retain their prescription and the fee charge slips provided by the clinician. A week
4
5 after the USP clinic visit, clinicians will be assigned a smartphone-based VP
6
7 assessment, which will consist of an initial demonstration VP case to allow the
8
9 clinician to familiarize themselves with how the system operates and then the test VP
10
11 case of the same USP condition. The VP-clinician interactions, drugs dispensed, and
12
13 fees charged will all be recorded automatically by the online assessment system.
14
15

16
17
18 The concordance of the two USP and VP assessments will then be analyzed by
19
20 Lin's concordance correlation coefficient (r_c)³⁰ for continuous *process quality* scores,
21
22 *fees charged* (yuan), and *time spent* (min), and the Kappa statistic³⁴ for dichotomous
23
24 *diagnoses* and *treatment & management* measures. r_c evaluates how close pairs of
25
26 observation fell on a 45° line (the perfect concordance line) through the origin in
27
28 addition to their correlation. Kappa measures agreement in assessment beyond what is
29
30 expected by chance alone. In addition, for continuous measures, a Bland-Altman plot
31
32 will also be used to visualize the concordance.^{35 36} For dichotomous measures, we will
33
34 analyze their sensitivity (i.e., strength to detect correct diagnosis, treatment plan, etc.)
35
36 and specificity (i.e., strength to detect incorrect diagnosis, treatment plan, etc.) using
37
38 USP as the reference.
39
40
41
42
43

44 45 *Reliability*

46
47 To establish *test-retest reliability*, clinicians previously being assessed by VPs will be
48
49 instructed to retake the same VP tests four weeks after their last assessment. The
50
51 second VP test is set one month later than the first to reduce the practice effect,³⁷
52
53 assuming the clinician's general medical knowledge remains constant.³² The
54
55
56
57

1
2
3 concordance of the two repeated tests indicates the stability of the VP assessment
4
5 tool.³² Similar concordance measures (i.e., r_c for continuous and Kappa for
6
7 dichotomous measures) as described above will be used. *Internal consistency*, the
8
9 intercorrelation of scores for process quality indicators, will be computed by
10
11 Cronbach's alpha coefficients (α),³⁸ with $\alpha > 0.7$ representing acceptable reliability.³⁹
12
13 Table 2 summarizes the validity, reliability, and feasibility measures that will be
14
15 examined in our study.
16
17
18
19

20 **Patient and public involvement**

21
22 We will seek feedback from clinicians and patient representatives in the feasibility
23
24 study and use their feedback to refine the VP cases. Our USPs will be lay people
25
26 trained to portray patients and assess care quality based on their interactions with
27
28 clinicians. Our scoring criteria thus are also patient centered. Furthermore, all
29
30 participants will be acknowledged for their involvement in the study and will be
31
32 provided with a final summary report of the study outcomes and will have free access
33
34 to the VP training website. All published results will be publicly available.
35
36
37
38
39

40 **ETHICS AND DISSEMINATION**

41
42 The study has been approved by the Institutional Review Board of the School of
43
44 Public Health (IRB), Sun Yat-sen University (No. 2017-007). Informed consent will
45
46 be obtained from all clinicians participating in the VP tests. However, to reduce
47
48 participation bias due to self-selection,⁴⁰ our IRB has approved the implementation of
49
50 USP without prior informed consent from the individual participants, on the condition
51
52 that involved clinicians will be fully de-identified and all analyses will only be
53
54
55
56
57
58
59
60

1
2
3 conducted at the population level.⁴⁰ Study data will be securely stored and only
4
5 de-identified information will be used for analysis. We will seek to publish study
6
7 findings in peer-reviewed journals and produce reports to inform health authorities.
8
9
10 The tools and technology developed in this study will be freely available to other
11
12 LMICs for research purposes.
13

14 15 DISCUSSION

16
17 To the best of our knowledge, this is the first study validating VP as a quality
18
19 assessment tool in rural primary health care centers. This study follows an
20
21 evidence-based approach to develop VP cases and scoring criteria, implements them
22
23 on a widely accessible platform (i.e., a smartphone), and systematically validates the
24
25 VP assessment tool via a cross-national multicenter study representing rural PHCs
26
27 over a wide range of geographic areas with distinct life expectancies and economic
28
29 development levels. The VP assessment tool's accessibility, flexibility and scalability
30
31 give it good potential to be easily adapted to other LMICs.
32
33
34
35
36

37
38 VP has mainly been used in medical education to train and test critical thinking¹⁸
39
40^{41 42}, and until recently few studies have applied the method in a practice setting to
41
42 influence health provider behavior and improve care quality.^{43 44} As an extension, we
43
44 propose to validate VP as a quality assessment tool delivered via widely accessible
45
46 smartphones. Nevertheless, it is to be noted that given its simulated nature, the
47
48 VP-test theoretically may never completely bridge the 'know-do' gap. The validation
49
50 study is thus essential to quantify the concordance/discordance between VP- and
51
52 USP-based quality assessments. Our study will generate firsthand empirical evidence
53
54
55
56
57

1
2
3 contributing to the understanding of the ‘know-do gap’,^{5 45} and shed light on
4
5
6 circumstances that cannot be tested by USPs.
7

8
9 A limitation of the study, however, is that, in order to test the validity of VP
10
11 against USP as the reference standard, we restrict the selection of VP cases to those
12
13 that can be simulated by USP. This conservative first step will nevertheless allow us
14
15 to examine the extent to which VP can reflect care quality, and a follow-up study will
16
17 then explore the full potential of the VP in assessing quality of care. Further, the two
18
19 purposely-selected counties for each province may not represent the provincial
20
21 conditions entirely, although we will make every effort to consider provincial
22
23 representation when selecting counties. Third, while the validation study is
24
25 exclusively conducted on PHCs in rural China, the extent to which the VP assessment
26
27 tool can be transported to other LMICs remains to be evaluated. Nonetheless, by
28
29 implementing the study in a diverse set of Chinese provinces may improve the
30
31 generalizability of our study considering the comparable life expectancies of LMICs
32
33 and these provinces.
34
35
36
37
38

39 40 Author’s contributions

41
42 All authors contributed to the conceptualization and design of the study. DX, JL, and
43
44 YYC conceived the initial study design, analytical methods, and composition of the
45
46 team. JL was responsible for the study concept, initial draft, and revisions. DX was
47
48 responsible for the study concept and revising the draft. YLC and XHW were
49
50 responsible for the development of the scoring criteria. SS, KH, HW, JNW, ZLZ, ZN,
51
52 WJG, JP, CXT and WZ provided critical review and revision to the study design. All
53
54
55
56
57
58
59
60

1
2
3 authors read and approved the final revision.
4

5 6 Funding

7
8 The study is supported by the China Medical Board through its Health Policy and
9
10 Systems Sciences Open Competition grant ‘Quality in primary health care: using
11
12 unannounced standardized patients’ (grant No.: CMB16-260, XD, PI).
13

14 15 Acknowledgement

16
17 The study will be led by the Sun Yat-sen Global Health Institute of Sun Yat-sen
18
19 University with a consortium of researchers from seven other Chinese universities,
20
21 including Central South University, Guangzhou University, Guizhou Medical
22
23 University, Inner Mongolia Medical University, Lanzhou University, Sichuan
24
25 University, and Xi’an Jiaotong University. We thank all the students from these
26
27 universities who have contributed to our project, especially Wenjun He who produced
28
29 Figure 1, Jianjian Wang who assisted with the evidence collection of the Scoring
30
31 Criteria, and Ash Harris who contributed to the VP computerization.
32
33
34
35
36

37 38 Competing interests

39
40 The development of the VP assessment tool is a joint project of the Sun Yat-sen
41
42 University Global Health Institute (SGHI) (representing the seven universities in
43
44 China), and *CureFUN*. However, the VP cases will be independently developed and
45
46 the validation studies will be rigorously conducted by the research team from SGHI
47
48 and the seven universities, whereas *CureFun* will technically implement the cases on
49
50 smartphones and have no influence over the study design and analysis.
51
52
53
54
55
56
57

REFERENCES

1. Kieny MP, Evans DB. Universal health coverage. *East Mediterr Health J* 2013;19(4):305-6.
2. Transforming our world: the 2030 Agenda for Sustainable Development. : United Nations- Sustainable Development Knowledge Platform, 2015.
3. 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.
4. 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.
5. Das J, Hammer J, Leonard K. The quality of medical advice in low-income countries. *The Journal of Economic Perspectives* 2008;22(2):93-114.
6. 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.
7. 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.
8. Donabedian A. Evaluating the quality of medical care. *The Milbank Quarterly* 2005;83(4):691-729.
9. Peabody JW, Luck J, Glassman P, et al. Comparison of vignettes,

1
2
3
4 standardized patients, and chart abstraction: a prospective validation study of
5
6 3 methods for measuring quality. *JAMA* 2000;283(13):1715-22.

7
8 10. Shah R, Edgar D, Evans BJ. Measuring clinical practice. *Ophthalmic and*
9
10
11 *Physiological Optics* 2007;27(2):113-25.

12
13 11. Glassman PA, Luck J, O'Gara EM, et al. Using standardized patients to
14
15 measure quality: evidence from the literature and a prospective study. *The*
16
17 *Joint Commission journal on quality improvement* 2000;26(11):644-53.

18
19
20 12. Collins J, Harden R. The Use of Real Patients, Simulated Patients and
21
22
23 Simulators in Clinical Examinations Association for Medical Education in
24
25 Europe (AMEE) Guide, 2004.

26
27
28 13. Triola M, Feldman H, Kalet AL, et al. A randomized trial of teaching clinical
29
30 skills using virtual and live standardized patients. *Journal General Internal*
31
32 *Medicine* 2006;21(5):424-9.

33
34
35 14. Peabody JW, Luck J, Glassman P, et al. Measuring the quality of physician
36
37 practice by using clinical vignettes: a prospective validation study. *Annals of*
38
39 *Internal Medicine* 2004;141(10):771-80.

40
41
42 15. Shah R, Edgar DF, Evans BJ. A comparison of standardised patients,
43
44 record abstraction and clinical vignettes for the purpose of measuring clinical
45
46 practice. *Ophthalmic and Physiological Optics* 2010;30(3):209-24.

47
48
49 16. Dresselhaus TR, Peabody JW, Luck J, et al. An evaluation of vignettes for
50
51 predicting variation in the quality of preventive care. *Journal of general internal*
52
53 *medicine* 2004;19(10):1013-18.

- 1
2
3
4 17. Ellaway R, Candler C, Greene P, et al. An architectural model for
5
6 MedBiquitous virtual patients. Baltimore: MedBiquitous, 2006.
7
- 8
9 18. Cook DA, Triola MM. Virtual patients: a critical literature review and
10
11 proposed next steps. *Med Educ* 2009;43(4):303-11. doi:
12
13 10.1111/j.1365-2923.2008.03286.x [published Online First: 2009/04/02]
14
15
- 16 19. Zhou M, Wang H, Zhu J, et al. Cause-specific mortality for 240 causes in
17
18 China during 1990–2013: a systematic subnational analysis for the Global
19
20 Burden of Disease Study 2013. *The Lancet* 2016;387(10015):251-72.
21
22
- 23 20. Babiarz KS, Miller G, Yi H, et al. China's new cooperative medical scheme
24
25 improved finances of township health centers but not the number of patients
26
27 served. *Health Affairs* 2012;31(5):1065-74.
28
29
- 30 21. Li X, Lu J, Hu S, et al. The primary health-care system in China. *The*
31
32 *Lancet* 2017;390(10112):2584-94.
33
34
- 35 22. Qian D, Pong RW, Yin A, et al. Determinants of health care demand in
36
37 poor, rural China: the case of Gansu Province. *Health Policy and Planning*
38
39 2009;24(5):324-34.
40
41
- 42 23. Yip WC, Wang H, Liu Y. Determinants of patient choice of medical provider:
43
44 a case study in rural China. *Health Policy and Planning* 1998;13(3):311-22.
45
46
- 47 24. An Analysis Report of National Health Services Survey in China. Beijing:
48
49 Center for Health Statistics and Information, NHFPC, 2013.
50
51
- 52 25. Brouwers MC, Kho ME, Browman GP, et al. AGREE II: advancing
53
54 guideline development, reporting and evaluation in health care. *Canadian*
55
56

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Medical Association Journal 2010;182(18):E839-E42.

26. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 2011;155(8):529-36.

27. Hsu CC, Sandford BA. The Delphi Technique: Making Sense Of Consensus. *Practical Assessment Research & Evaluation* 2007;26(10):289-304.

28. Zamanzadeh V, Ghahramanian A, Rassouli M, et al. Design and implementation content validity study: development of an instrument for measuring patient-centered communication. *Journal of Caring Sciences* 2015;4(2):165.

29. Lacasse Y, Godbout C, Series F. Health-related quality of life in obstructive sleep apnoea. *European Respiratory Journal* 2002;19(3):499-503.

30. Lawrence I, Lin K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989:255-68.

31. McBride G. A proposal for strength-of-agreement criteria for Lin's concordance correlation coefficient. NIWA Client Report: HAM2005-062: Hamilton: National Institute of Water & Atmospheric Research, Ltd (NZ), 2005.

32. Kimberlin CL, Winetrstein AG. Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacy* 2008;65(23)

33. Rethans JJ, Gorter S, Bokken L, et al. Unannounced standardised patients

1
2
3 in real practice: a systematic literature review. *Medical Education*
4
5
6 2007;41(6):537-49.

7
8 34. Cohen J. A coefficient of agreement for nominal scales. *Educational and*
9
10
11 *Psychological Measurement* 1960;20(1):37-46.

12
13 35. Kwicien R, Koppshneider A, Blettner M. Concordance analysis: part 16
14
15 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt*
16
17
18 *International* 2011;108(30):515.

19
20 36. Bland JM, Altman DG. Statistical methods for assessing agreement
21
22 between two methods of clinical measurement. *Lancet*
23
24
25 1986;327(8476):307-10.

26
27 37. McCabe D, Langer KG, Borod JC, et al. Practice Effects. In: Kreutzer JS,
28
29 DeLuca J, Caplan B, eds. *Encyclopedia of Clinical Neuropsychology*. New
30
31
32 York: Springer 2011:1988-89.

33
34 38. Cronbach LJ. Coefficient alpha and the internal structure of tests.
35
36
37 *Psychometrika* 1951;16(3):297-334.

38
39 39. DeVellis RF. *Scale development: Theory and applications*. Los Angeles:
40
41
42 Sage 2012:109-10.

43
44 40. Rhodes KV, Miller FG. *Simulated Patient Studies: An Ethical Analysis*.
45
46
47 *Milbank Quarterly* 2012;90(4):706-24.

48
49 41. Reliability of a virtual patient simulation as an assessment tool.

50
51
52 PHARMACOTHERAPY; 2017. WILEY 111 RIVER ST, HOBOKEN
53
54
55 07030-5774, NJ USA.

1
2
3
4 42. Urrestigundlach M, Tolks D, Kiessling C, et al. Do virtual patients prepare
5
6 medical students for the real world? Development and application of a
7
8 framework to compare a virtual patient collection with population data. *Bmc*
9
10
11 *Medical Education* 2017;17(1):174.

12
13 43. Blok AC, May CN, Sadasivam RS, et al. Virtual Patient Technology:
14
15 Engaging Primary Care in Quality Improvement Innovations. *Jmir Medical*
16
17 *Education* 2017;3(1):e3.

18
19
20 44. Mollica R, Lavelle J, Fors U, et al. Using the Virtual Patient to Improve the
21
22 Primary Care of Traumatized Refugees. *Journal of Medical Education*
23
24 2017;16(1)

25
26
27 45. Mohanan M, Vera-Hernández M, Das V, et al. The know-do gap in quality
28
29 of health care for childhood diarrhea and pneumonia in rural India. *JAMA*
30
31 *Pediatrics* 2015;169(4):349-57.
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 1. Seven sample provinces in China referencing countries of equivalent life expectancy in parentheses.

For peer review only

1
2
3 Figure 2. Virtual patient case development team role and responsibilities.
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 3. Main components of smartphone-based virtual patient program.

For peer review only

1
2
3
4 Figure 4. Sampling and study process for one VP case in Danzhai County, Guizhou Province.
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

Table 1. Strengths and limitations of quality assessment measures.

Process Measure of Quality	Strengths	Limitations	Assessment Level
Unannounced Standardized Patient (USP)	Standardized, controlling for case-mix and patient-mix; Unannounced, no Hawthorne effect (if not be detected); Reduced recall bias	Expensive Limited medical conditions First-visit bias Selection bias if informed consent is required	Action (Do) Gold standard
Clinical Vignettes	Cost-effective; Suitable for large scale and cross-system comparison; Covering all illnesses; Accounting for case-mix and patient-mix.	Hawthorne effect Selection bias	Competence (Know how) Over-estimating care quality (best answers) or similar
Virtual Patients (VP)	Interactive Real-time response & automatic record Highly standardized Scalability & low marginal cost Efficient delivery: anytime, anywhere Suitable for large scale study	Hawthorne effect Selection bias High cost in initial development	Performance (Show how) ?

Table 2. Main validation domains of the study.

Domain	Indicator	Data collection		Statistical analysis
		Phase	Method	
Content validity	Content validity index (CVI)	VP case review	Evaluations by an expert panel after reviewing VP cases, measured by a 4-point Likert scale (1=lowest, 4=highest).	CVI for VP case and for specific VP domain will be computed, where CVI = number of raters giving a rating of 3 or 4 divided by the total number of raters.
Feasibility	Willingness to participate; Adherence rate	Feasibility study	The subsample of clinicians' interactions with the 2 VP cases will be recorded by the online assessment	Willingness to participate = clinicians taking the VP tests divided by the percentage of clinician selected Adherence rate = clinicians completed 2 VP cases divided by the percentage of clinicians taking VP tests
			Clinicians' subjective attitude toward the VP test experience measured by a 5-point Likert scale (1=most negative, 5=most positive).	Satisfying score for VP case and for specific aspects (e.g., usability, accessibility, etc.) will be computed, where satisfying score = frequency multiply by positive evaluations (3 to 5), and scores ≥ 1.5 are considered acceptable.
Face validity	Satisfying score			
Criterion validity	Concordance correlation coefficient (r_c); Kappa statistic	Validation study	The same clinician receives a USP visit and a VP test for a matching condition. The USP-clinician interaction is evaluated by the USP using the checklist, including fees and time per visit; while VP-clinician interaction is graded by the system.	The concordance of VP-test scores against USP-test score (gold standard) or two-repeated VP-tests will be examined by r_c for continuous process quality scores, fees charged (yuan), and time spent (min); and Kappa for dichotomous diagnoses and treatment & management measures.
			Repeat VP-tests on the same clinician in a month	
Test-retest Reliability				
Internal consistency	Cronbach's alpha coefficient (α)		VP-test scores on a single occasion	Intercorrelation of scores for process quality indicators with alpha > 0.7 is acceptable.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

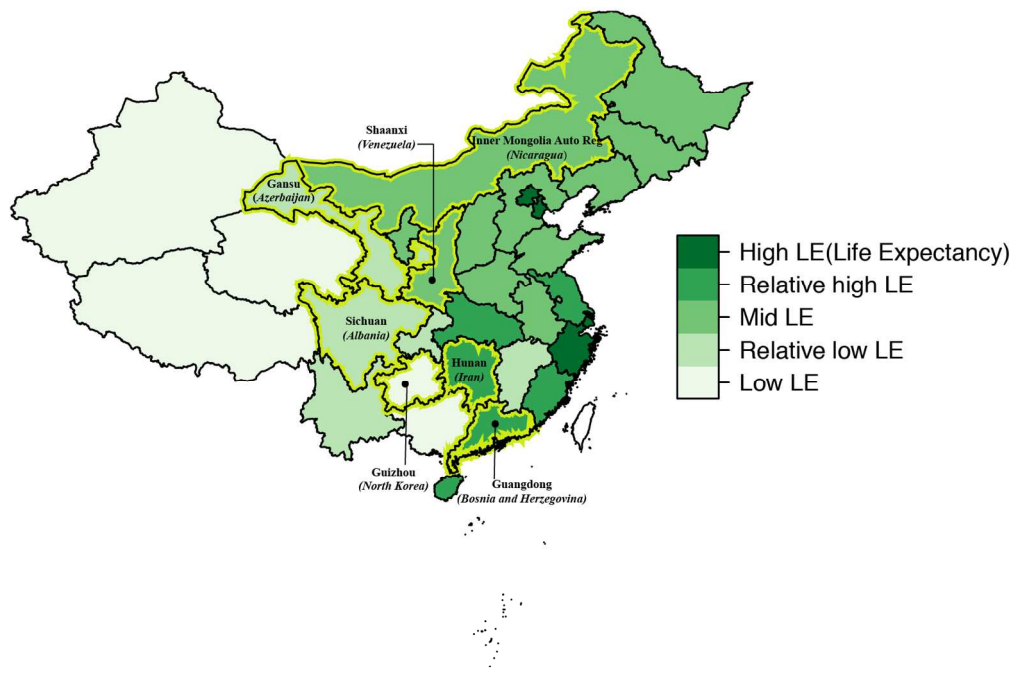


Figure 1. Seven sample provinces in China referencing countries of equivalent life expectancy in parentheses.

147x98mm (300 x 300 DPI)

view only

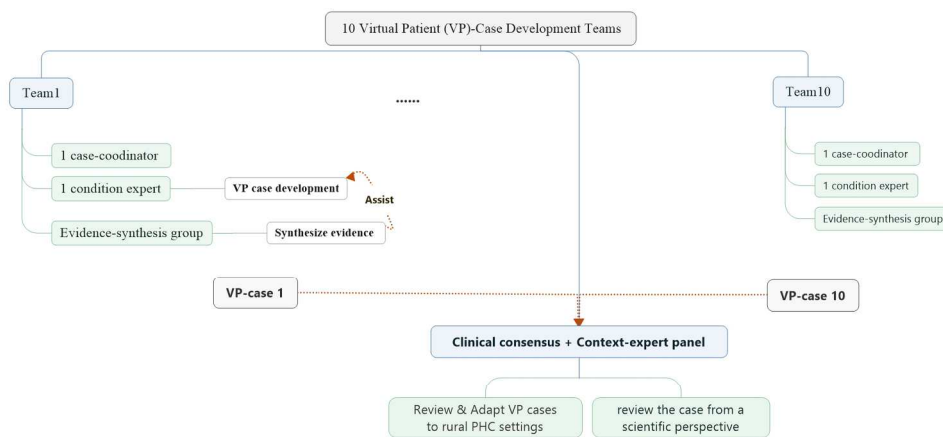
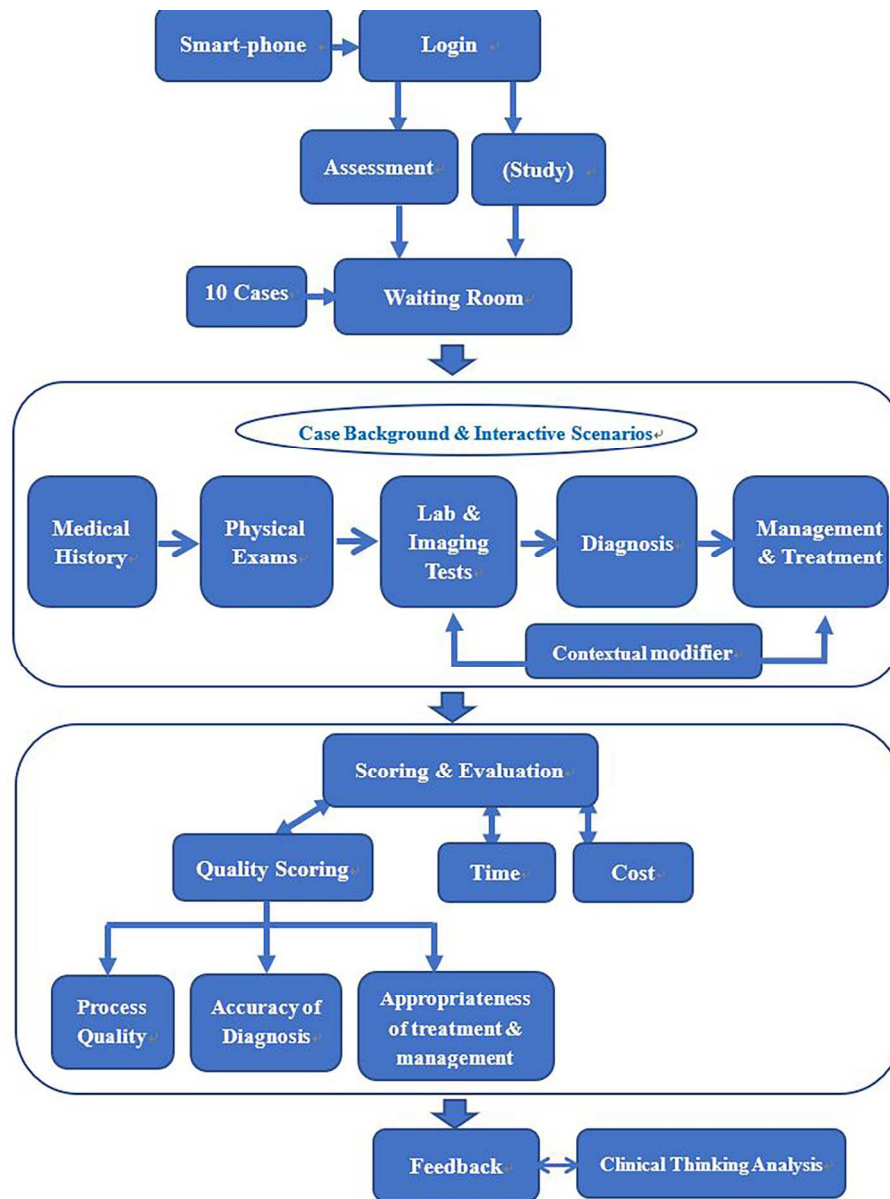


Figure 2. Virtual patient case development team role and responsibilities.

264x198mm (300 x 300 DPI)



45 Figure 3. Main components of smartphone-based virtual patient program.

46 158x203mm (300 x 300 DPI)

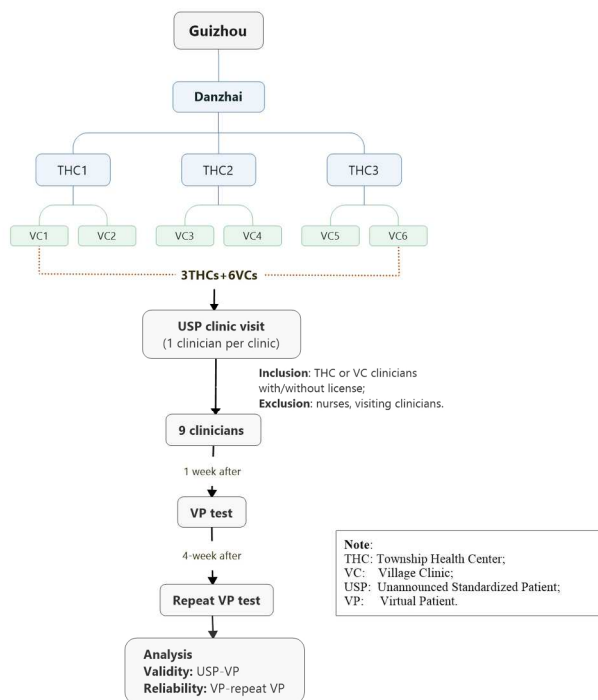


Figure 4. Sampling and study process for one VP case in Danzhai County, Guizhou Province.

264x198mm (300 x 300 DPI)

Appendix 1. Top 30 conditions of high-frequency clinical encounters in primary health care settings in rural China.

Clinical condition	Two-week consultation constituent ratio ¹		Township health center ²		Village clinics ²	
	RANK	%	RANK	%	RANK	%
Cold	1	28	1	13.60	1	19.50
Hypertension	2	21.8	4	7.90	6	9.80
Diabetes mellitus	4	3.9	7	4.60	8	4.60
Chronic tracheitis	8	2	2	9.50	4	10.50
Acute tracheitis			3	9.00	3	10.70
Gastritis	3	5.5	5	7.50	5	10.30
Diarrhea			6	5.30	2	11.70
Urinary tract infection			8	4.50	9	2.90
Osteoarthritis	7	2.30	17	2.40	18	0.50
Low back pain	5	3.10	16	2.50	15	0.80
Psoatic strain			14	2.60	14	0.80
Peptic ulcer			11	2.90	11	1.70
General trauma			10	3.10	13	1.10
Sciatica			19	1.80	22	0.30
Child dyspepsia			9	3.20	7	7.00
Pelvic inflammatory disease			12	2.70	16	0.60
Vaginitis			13	2.70	17	0.50
Dysmenorrhoea			18	2.30	19	0.50
Cholecystitis			15	2.60	12	1.60
Toothache	10	1.30	22	1.30	10	2.70
Menopausal syndrome			21	1.40	27	0.10
Cholelithiasis			20	1.60	20	0.40
Idiopathic headache	6	2.50	25	0.60	25	0.20
Hemorrhoids			23	1.10	21	0.40
Asthma			28	0.60	26	0.20
Chronic dermatitis			29	0.20	29	0.10
Tympanitis			24	0.70	24	0.20
Conjunctivitis			27	0.60	28	0.10
Sinusitis			26	0.60	23	0.30
Ischemic heart disease	9	1.5				

¹ Self-reported two-week consultation constituent ratio by community dwellers, information from the 2013 National Health Service Survey in China.

² Clinicians reported common clinical conditions in primary health care centers by centers' type.

Appendix 2. Methods for checklist and standards development

To evaluate the quality of care in primary health care institutions, key diagnosis and treatment points of common and frequently-occurring diseases will be developed. The *WHO Handbook for Guideline Development* and evidence-based evaluation principles will be adopted. The main procedures are comprised in the following six steps that will be implemented.

1. Expert group recruitment: Convene a multidisciplinary group consisting of experts in public health, evidence-based medicine/document retrieval, as well as clinical physicians.
2. Data retrieval and literature evaluation: Employ a 5S model to retrieve and incorporate clinical practice guidelines, textbooks, systematic reviews, meta-analysis, and important literature reviews. Retrieve literature from Wanfang, Medlive, MEDLINE, National Guideline Clearinghouse (NGC), National Institute for Health and Care Excellence (NICE), World Health Organization (WHO), DynaMed, and UpToDate. Evaluate the literature using AGREE II, AMSTAR, and QUADAS-2 for the included clinical practice guidelines, systematic reviews, and diagnostic tests, respectively.
3. Preliminary items pool development: Extract essential diagnostic and treatment procedures from the high-quality literature attained.
4. Clinical expert consensus: Apply a 2- to 3-round Delphi method to achieve consensus for diagnosis and treatment. The importance, necessity, and feasibility of the items should be considered in the process of Delphi, and additional medical information must be supplemented in terms of the clinical practice. Furthermore, all items should be classified as: necessary (3 points), selective (2 points), irrelevant (1 points), and erroneous (0 points).
5. Pilot and revise: Conduct a pilot test among 2~3 primary health care settings using the preliminary items. Revise the items and finalize the key diagnosis and treatment point evaluation items.

- 1
- 2
- 3 6. Script development: Develop the script of the target disease based on key
- 4 diagnosis and treatment point evaluation items before conducting the quality of
- 5 service evaluation in primary health care institutions.
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20
- 21
- 22
- 23
- 24
- 25
- 26
- 27
- 28
- 29
- 30
- 31
- 32
- 33
- 34
- 35
- 36
- 37
- 38
- 39
- 40
- 41
- 42
- 43
- 44
- 45
- 46
- 47
- 48
- 49
- 50
- 51
- 52
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60

For peer review only

Appendix 3. Demonstrations of *Cure-Fun* smartphone-based platform current configurations of interview, physical exam and lab texts, and treatment.

