PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Using smartphone-based virtual patients to assess the quality of primary health care in rural China: protocol for a prospective multicenter study
AUTHORS	Liao, Jing; Chen, Yao-Long; Cai, Yiyuan; Zhang, Nan; Sylvia, Sean; Hanson, Kara; Wang, Hong; Wasserheit, Judith; Gong, Wenjie; Zhou, Zhongliang; Pan, Jay; Wang, Xiaohui; Tang, Chengxiang; Zhou, Wei; Xu, Dong

VERSION 1 – REVIEW

REVIEWER	Dean Carson Charles Darwin University, Australia
REVIEW RETURNED	22-Dec-2017

GENERAL COMMENTS	A minor editorial question – • I do not understand the sentence commencing “Despite the fact…” page 4 line 16. Attention has been shifted from where to where? How does the first of the sentence link to the second part? Otherwise the paper is extremely well written, well structured, and a good documentation of the study protocol. I do wonder how the (presumed) unfamiliarity of ‘consulting’ a smartphone VP is likely to influence results. A USP replicates a standard clinical encounter, but I suspect that few if any of the clinicians will regularly provide consultations via a smartphone in a way that is similar to VP? There may be a need for some learning time from the clinicians which could be based around a presentation that is not used in the research itself. This would allow an assessment of the clinician’s competency with the technology to be made before assessing the clinical quality.

REVIEWER	Amanda C. Blok, PhD, MSN, PHCNS-BC Assistant Professor of Nursing, Graduate School of Nursing; University of Massachusetts Medical School, Nurse Post-Doctoral Fellow, Center for Healthcare Organization and Implementation Research (CHOIR); United States Department of Veterans Affairs
REVIEW RETURNED	24-Jan-2018

GENERAL COMMENTS	This will be informative work. Criteria 2: Is the Abstract accurate, balanced and complete? The abstract is not balanced, with the introduction needing to be reduced in size (first four sentences) and a discussion section should be added on how this contributes to the science. The aim should be something like, “Our study aims develop and validate smartphone-based VP as a quality assessment tool for primary care,

	compared to the gold standard of USP." Then stating in the discussion more about it's implications, like "VPs could be routinely applied to evaluate the quality of primary health care provided by rural primary health centers", ect. Criteria 3: Is the study design appropriate to answer the research question? Testing the VP will have the same limitations as the USP gold standard, in that only a small range of cases will be tested. I am excited about VP case development teams doing a good amount of work to identify common cases in primary care that a VP can help to test while USPs could only do a subset. However, I think more VP case development teams will be needed (or teams to create more than one case) than only 10, because they will likely identify many common cases in primary care that USPs cannot test. Therefore, you will have a set of, let's say, 40 cases of common primary care complaints (tested with feasibility study), and could test a subset of 10 cases that are compatible with USP testing as well in your criterion validity study. Limiting VP case creation by USP limits would be disappointing, especially because primary care quality will likely shine through with cases that are more essential in primary care quality (identifying pneumonia, infections, etc.). Criteria 7: If statisticas are used are they appropriate and described fully? The text does have a lot of information about statistical analyses. Table 2 has a column on statistical analyses for each of the domains being measured. However, some of them are not clearly written in this table, with short-forms that may not be standard or readily understood (such as using a forward slash, which could indicate division or an alternate measure). Also some of the order is confusing, like "clinicians being selected %", which could mean "percent of clinicians selected". If this is clarified in the table, in a notes section at the bottom of the table or the text, that would be helpful to understand the statistical analyses proposed. Criteria 8: Are the references up-to-date and accurate? The references most use literature from the early 2000s, with no literature from this year or 2017 at all. Virtual patients and mobile health technology, along with quality improvement literature and medical education, is developing rapidly in science. Bringing in the latest from these categories (and including it in the very short discussion section) would be a huge improvement to the paper. Tech journals like JMIR would be extremely helpful. Criteria 13: Is the supplementary reporting complete? Taking Figure 4's flow diagram and adding on all the CONSORT diagram properties to it would make it more informative (specifying allocation, follow-up and analysis to the lower half of the diagram. This Figure, and all of the figures, really need a notes section at the bottom to explain your diagram further. As it is, they all have information that traditionally should be explained in a notes section in the figure's title or throughout the figure. For example, Figure 4's title explains the abbreviations THC and VC in parentheses in the title, which is not typical. This goes for Figures one through four.

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:
Reviewer: 1
Reviewer Name: Dean Carson
Institution and Country: Charles Darwin University, Australia
Please state any competing interests: None declared
Please leave your comments for the authors below
A minor editorial question –
• I do not understand the sentence commencing "Despite the fact…" page 4 line 16. Attention has been shifted from where to where? How does the first of the sentence link to the second part?
Otherwise the paper is extremely well written, well structured, and a good documentation of the study protocol. I do wonder how the (presumed) unfamiliarity of 'consulting' a smartphone VP is likely to influence results. A USP replicates a standard clinical encounter, but I suspect that few if any of the clinicians will regularly provide consultations via a smartphone in a way that is similar to VP? There may be a need for some learning time from the clinicians which could be based around a presentation that is not used in the research itself. This would allow an assessment of the clinician's competency with the technology to be made before assessing the clinical quality.
Response: We thank Dr. Carson for these valuable comments. We apologized for the lack of clarity and have revised that sentence as 'Despite the fact that great emphasis has been made to enhance health care services, …' (Page 4)
We share Dr. Carson's concern that the unfamiliarity of the VP system may create potential barriers. At our feasibility study, we will take users' experience of the VP assessment tool into consideration and make sure the assessment tool is easy to use (Page 12); further, during the actual assessment, a demonstration of the VP operation will be delivered alongside the examination case. We now clarified this point in Methods:"A week after the USP clinic visit, clinicians will be assigned a smartphone-based VP assessment, which will consist of a demonstration VP case to allow the clinician getting familiar with the operation system, and the examination VP case of the same USP condition." (Page 15)
Reviewer: 2
Reviewer Name: Amanda C. Blok, PhD, MSN, PHCNS-BC
Institution and Country: Assistant Professor of Nursing, Graduate School of Nursing; University of Massachusetts Medical School, Nurse Post-Doctoral Fellow, Center for Healthcare Organization and Implementation Research (CHOIR); United States Department of Veterans Affairs
Please state any competing interests: None declared.
Please leave your comments for the authors below
This will be informative work.
Criteria 2: Is the Abstract accurate, balanced and complete?
The abstract is not balanced, with the introduction needing to be reduced in size (first four sentences) and a discussion section should be added on how this contributes to the science. The aim should be something like, "Our study aims develop and validate smartphone-based VP as a quality assessment tool for primary care, compared to the gold standard of USP." Then stating in the discussion more about it's implications, like "VPs could be routinely applied to evaluate the quality of primary health care provided by rural primary health centers", ect.
Response: We thank Professor Amanda for this suggestion, and revised the Introduction of the Abstract accordingly, but left the Discussion to the main text in adherence to the journal's format.
"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 develop and validate smartphone-based VP as a quality assessment tool for primary care,

compared to USP.” (Page 2-Abstract- Introduction)
Criteria 3: Is the study design appropriate to answer the research question?
Testing the VP will have the same limitations as the USP gold standard, in that only a small range of cases will be tested. I am excited about VP case development teams doing a good amount of work to identify common cases in primary care that a VP can help to test while USPs could only do a subset. However, I think more VP case development teams will be needed (or teams to create more than one case) than only 10, because they will likely identify many common cases in primary care that USPs cannot test. Therefore, you will have a set of, let’s say, 40 cases of common primary care complaints (tested with feasibility study), and could test a subset of 10 cases that are compatible with USP testing as well in your criterion validity study. Limiting VP case creation by USP limits would be disappointing, especially because primary care quality will likely shine through with cases that are more essential in primary care quality (identifying pneumonia, infections, etc.).
Response: We totally agree with Prof. Amanda that VP has greater potential than USP considering the unlimited cases that can be portrayed to test care quality. We’d like to clarify that although our research team focuses more on ten cases many more cases are currently under development by other teams using the CureFun platform. Our primary purpose is to using those 10 tracer conditions to validate the VP against USP as the reference standard. We now emphasize this point in the Discussion: ‘A limitation of the study, however, is that, in order to test the validity of VP against USP as the reference standard, we restrict the selection of VP cases to those that can be simulated by USP as well. This conservative first step will nevertheless allow us to examine the extent to which VP can reflect care quality, and follow-up study will then explore the full potential of the VP in assessing quality of care.’ (Page 17)
Criteria 7: If statisticas are used are they appropriate and described fully?
The text does have a lot of information about statistical analyses. Table 2 has a column on statistical analyses for each of the domains being measured. However, some of them are not clearly written in this table, with short-forms that may not be standard or readily understood (such as using a forward slash, which could indicate division or an alternate measure). Also some of the order is confusing, like “clinicians being selected %”, which could mean “percent of clinicians selected”. If this is clarified in the table, in a notes section at the bottom of the table or the text, that would be helpful to understand the statistical analyses proposed.
Response: We thank Prof. Amanda for these great comments and have revised the table as suggested. Table 2 Page 31
Criteria 8: Are the references up-to-date and accurate?
The references most use literature from the early 2000s, with no literature from this year or 2017 at all. Virtual patients and mobile health technology, along with quality improvement literature and medical education, is developing rapidly in science. Bringing in the latest from these categories (and including it in the very short discussion section) would be a huge improvement to the paper. Tech journals like JMIR would be extremely helpful.
Response: We thank Prof. Amanda for the suggestion. We have added several latest references into the text1-3 where appropriate, and particularly in the discussion section as, ‘VP has mainly been used in medical education to train and test critical thinking4-6, and only till recently few studies start to extend its usage into practice setting to change health provider behavior and improve care quality.7 8 As a further extension, our study proposes to validate VP as a quality assessment tool via widely accessible smartphones. Nevertheless, it is to be noted that given its simulated nature, the VP-quality assessment tool theoretically may never completely bridge the ‘competency-practice’ gap. Our validation study is thus essential to quantify the concordance/discordance between VP- and USP-based quality assessments. Our study will generate firsthand empirical evidence contributing to the understanding of the ‘know-do gap’,9, 10 and further shed light on circumstances that cannot be tested by USPs.’ (Page 17)
Criteria 13: Is the supplementary reporting complete?
Taking Figure 4’s flow diagram and adding on all the CONSORT diagram properties to it would make it more informative (specifying allocation, follow-up and analysis to the lower half of the diagram.

This Figure, and all of the figures, really need a notes section at the bottom to explain your diagram further. As it is, they all have information that traditionally should be explained in a notes section in the figure's title or throughout the figure. For example, Figure 4's title explains the abbreviations THC and VC in parentheses in the title, which is not typical. This goes for Figures one through four.
Response: We revised Figure 4 as kindly suggested by Prof. Amanda, but we note that as our validation study is not a randomized control trial, the properties of CONSORT may not be directly applied to our diagram. We now include a note section on the figure title from Figure 1 to 4

References:
1. 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.
2. Li X, Lu J, Hu S, et al. The primary health-care system in China. The Lancet 2017; 390(10112): 2584-94.
3. 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.
4. Cook DA, Triola MM. Virtual patients: a critical literature review and proposed next steps. Med Educ 2009; 43(4): 303-11.
5. D'Angelo R, Smith J, Delic J, Scholtz J. Reliability of a virtual patient simulation as an assessment tool. PHARMACOTHERAPY; 2017: WILEY 111 RIVER ST, HOBOKEN 07030-5774, NJ USA; 2017. p. E162-E3.
6. Urrestigundlach M, Tolks D, Kiessling C, Wagnermenghin M, Härtl A, Hege I. Do virtual patients prepare medical students for the real world? Development and application of a framework to compare a virtual patient collection with population data. Bmc Medical Education 2017; 17(1): 174.
7. Blok AC, May CN, Sadasivam RS, Houston TK. Virtual Patient Technology: Engaging Primary Care in Quality Improvement Innovations. Jmir Medical Education 2017; 3(1): e3.
8. Mollica R, Lavelle J, Fors U, Ekblad S. Using the Virtual Patient to Improve the Primary Care of Traumatized Refugees. Journal of Medical Education 2017; 16(1).
9. 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.
10. Mohanan M, Vera-Hernández M, Das V, et al. The know-do gap in quality of health care for childhood diarrhea and pneumonia in rural India. JAMA Pediatrics 2015; 169(4): 349-57.

VERSION 2 – REVIEW

REVIEWER	Dean Carson Charles Darwin University, Australia
REVIEW RETURNED	06-Mar-2018

GENERAL COMMENTS	The authors have added a sentence about training the practitioners for VP. I had no other concerns with the original draft.

REVIEWER	Amanda Blok Amanda C. Blok, PhD, MSN, PHCNS-BC, Assistant Professor, Graduate School of Nursing and Department of Quantitative Health Sciences, University of Massachusetts Medical School, USA
REVIEW RETURNED	20-Mar-2018

GENERAL COMMENTS	Hello, This is my second review of the manuscript. The authors have made many helpful changes and additions. I think this paper would benefit from thorough editing, as there are grammatical and terminology mistakes that may cause confusion to readers. The authors have provided very good content and it is interesting scientifically, yet changes are needed for readers to fully grasp their

	work. This comment is for the abstract and the entire paper. Additionally, insertion of descriptive words for clarity will help. For example, the first sentence of the abstract reads "Valid and low-cost quality assessment tools are not readily available." For what? I would say "Valid and low-cost HEALTHCARE quality assessment tools EXAMINING PRIMARY CARE PROVIDER CLINICAL PRACTICE are not readily available". This suggestion is for the abstract, as well as the paper. As I mention in the first review, the Tables and Figures are important, as most readers will look at these first or only these. Please do not add notes in parentheses (like abbreviations) in the title of the figures, but include them at the bottom of the figures in their own box. The titles are long and confusing. Please see how other papers at BMJ have done this. Overall, good job on changes, but as I emphasize through editing is needed for this manuscript to be received well and contribute to the literature. Thank you.

VERSION 2 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1

Reviewer Name: Dean Carson

Institution and Country: Charles Darwin University, Australia

Please state any competing interests: None declared

Please leave your comments for the authors below

The authors have added a sentence about training the practitioners for VP. I had no other concerns with the original draft.

Reviewer: 2

Reviewer Name: Amanda Blok

Institution and Country: Amanda C. Blok, PhD, MSN, PHCNS-BC, Assistant Professor, Graduate School of Nursing and Department of Quantitative Health Sciences, University of Massachusetts Medical School, USA

Please state any competing interests: None declared

Please leave your comments for the authors below

Hello, This is my second review of the manuscript. The authors have made many helpful changes and additions. I think this paper would benefit from thorough editing, as there are grammatical and terminology mistakes that may cause confusion to readers. The authors have provided very good content and it is interesting scientifically, yet changes are needed for readers to fully grasp their work. This comment is for the abstract and the entire paper. Additionally, insertion of descriptive words for clarity will help. For example, the first sentence of the abstract reads "Valid and low-cost quality assessment tools are not readily available." For what? I would say "Valid and low-cost HEALTHCARE quality assessment tools EXAMINING PRIMARY CARE PROVIDER CLINICAL PRACTICE are not readily available". This suggestion is for the abstract, as well as the paper.

Response: We thank Prof. Blok for these useful and kind suggestions. We have carefully proofread the whole manuscript, and make sure it is clear without grammatical mistakes (see marked changes throughout the manuscript).

As I mention in the first review, the Tables and Figures are important, as most readers will look at these first or only these. Please do not add notes in parentheses (like abbreviations) in the title of the figures, but include them at the bottom of the figures in their own box. The titles are long and confusing. Please see how other papers at BMJ have done this.

Response: Thanks again, Prof. Blok, we now edited the title as suggested. Page 26-31.

Overall, good job on changes, but as I emphasize through editing is needed for this manuscript to be received well and contribute to the literature. Thank you.