PATIENT OUTCOMES

Using standardized patients to evaluate hospital-based intervention outcomes

Li Li,[1]* Chunqing Lin[1] and Jihui Guan[2]

[1]Semel Institute for Neuroscience and Human Behavior, Center for Community Health, University of California at Los Angeles, Los Angeles, CA, USA and [2]Fujian Provincial Center for Disease Control and Prevention, Fuzhou, China

*Corresponding author. UCLA Semel Institute Center for Community Health, 10920 Wilshire Blvd., Suite 350, Los Angeles, CA 90024, USA. E-mail: lililili@ucla.edu

Accepted	7 November 2013
Background	The standardized patient approach has proved to be an effective training tool for medical educators. This article explains the process of employing standardized patients in an HIV stigma reduction intervention in healthcare settings in China.
Methods	The study was conducted in 40 hospitals in two provinces of China. One year after the stigma reduction intervention, standardized patients made unannounced visits to participating hospitals, randomly approached service providers on duty and presented symptoms related to HIV and disclosed HIV-positive test results. After each visit, the standardized patients evaluated their providers' attitudes and behaviours using a structured checklist. Standardized patients also took open-ended observation notes about their experience and the evaluation process.
Results	Seven standardized patients conducted a total of 217 assessments (108 from 20 hospitals in the intervention condition; 109 from 20 hospitals in the control condition). Based on a comparative analysis, the intervention hospitals received a better rating than the control hospitals in terms of general impression and universal precaution compliance as well as a lower score on stigmatizing attitudes and behaviours toward the standardized patients.
Conclusion	Standardized patients are a useful supplement to traditional self-report assessments, particularly for measuring intervention outcomes that are sensitive or prone to social desirability.
Keywords	Standardized patients, stigma, HIV/AIDS, intervention outcome, China

Introduction

In human behavioural research, the accuracy of self-report assessments has been questioned in many studies.[1–3] In particular, reporting bias due to social desirability has been a major challenge in surveys of sensitive behavioural or medical issues.[4–6] As research communities continue to explore alternatives for assessments in intervention trials, the use of standardized patients to moderate reporting bias is worth investigating.

A standardized patient (SP), also known as a 'simulated patient' or 'pseudo patient', is trained to present a particular case or symptom to a healthcare professional and to record the details of the encounter.[7,8] SP examinations have been deemed a useful tool for evaluating medical educators[9–13] and healthcare

providers.[14,15] SPs have been used with medical students, pharmacists and physicians to test a broad range of skills, including history-taking, physical examination and counselling.[16-18] The Rosenhan experiment, conducted in 1973 with eight SPs, is considered an important and influential criticism of psychiatric diagnostic procedures.[19] Although the SP method has the potential to be applied to the efficacy assessment of intervention trials, very few studies, if any, have employed it in this manner. This study describes an initial effort to apply the SP method in a stigma-reduction behavioural intervention trial as part of the efficacy assessment.

From October 2008 to December 2010, a randomized controlled intervention trial was conducted in two provinces of China to address HIV-related stigma and discrimination in healthcare settings. The intervention outcomes based on providers' self-reports at baseline and 6- and 12-month follow-up assessments are published elsewhere.[20] In addition to traditional self-report assessments with service providers, this study uses the SP method to measure stigmatizing attitudes and behaviours among service providers. The main purpose of this article is to describe the procedures for using SPs to assess service providers' stigmatizing behaviours and attitudes. We consider the feasibility of the methodology (i.e. whether SP assessments can be conducted without the trained research staff being detected or imposing a significant burden on researchers or participants). We also assess whether the approach can generate useful data for measuring HIV-related stigma in healthcare settings. Findings from this investigation could advance assessment strategies and provide an alternative data collection method for future intervention trials.

Methods
The randomized intervention trial
The Institutional Review Boards of the University of California, Los Angeles, and the National Center for AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention (NCAIDS/CCDC), approved this study. The randomized controlled trial was conducted in 40 county-level hospitals in Fujian Province and Yunnan Province. The 40 hospitals were randomly selected from a total of 214 county hospitals in the two provinces, with 20 hospitals selected from each province. The hospitals were randomized to either an intervention condition or a control condition. Guided by Diffusion of Innovation theory,[21] popular opinion leaders were identified and trained among the service providers to disseminate stigma reduction messages within their medical community. At the same time, universal precaution supplies (e.g. gloves, surgical masks and sharps containers) were provided to all participating hospitals to enhance the self-protection of providers. Self-administrated paper-

and-pencil surveys conducted at baseline and 6- and 12-month follow-up assessments showed lower prejudicial attitudes and avoidance intent to treat patients living with HIV (PLH) in the intervention group providers compared with the control condition providers. More details about the study design, intervention and survey outcomes are reported elsewhere.[20]

Standardized patients procedure
A total of seven staff members from the two provincial CDCs served as SPs in this study. Among them, there were three men and four women. All of the SPs satisfied the following criteria: (i) have formal medical training; (ii) have basic knowledge of HIV; (iii) have experience in hospital procedures; and (iv) have no familiarity with the hospitals or providers to be evaluated. The selected SPs were separated from the intervention team and the survey team members, and they were not informed of the intervention status of the hospital. All SPs received a 1-day training regarding the purpose and procedures of the study, ethical issues in conducting human research and emergency protocols. The designed scenarios were reviewed during the training, and role plays were used to illustrate how to react in different situations. In order to achieve inter-rater consistency for the assessment, detailed guidelines for evaluation, scoring and taking field notes were provided to the SPs. The guidelines specified which behaviours should be interpreted as stigma/discrimination (e.g. wearing gloves when collecting or handling blood or bodily fluids was considered normal universal precaution compliance, but wearing double gloves when performing a physical examination was considered overprotection).

The SP evaluation was conducted 1 year after the baseline assessment. Four to six patient visits were performed at each participating hospital by two different SPs. The service providers working in the hospitals were informed of the possibility of SP visits at the beginning of the study, but the exact SP visiting times were never announced. When performing the assessment, SPs assumed one of three guises: (i) 'normal urban' (casual clothing, well-kept, modest economic background, Mandarin-speaking); (ii) 'normal rural' (rustic clothing, local dialect speech pattern); or (iii) 'deviant looking' (risqué clothing, visible tattoos, flirtatious behaviour). These guises were built upon findings from our previous study that showed providers tended to judge patients' HIV infection risk based on their appearance.[22] SPs visited all participating hospitals in the provinces, registered as normal patients and randomly approached service providers who were on duty in various departments of the hospital. The SPs followed the same four steps when speaking with a service provider: (i) describe symptoms and complaints that might indicate HIV infection, including fever, skin infection, dental problems and/or the need for surgery. The symptoms were tailored to the department the SP was visiting

(e.g. report a urinary tract infection when seeing a urologist, report a skin rash when seeing a dermatologist etc.); (ii) confess a history of risky sexual behaviour; (iii) disclose HIV-positive status with a fake medical report; and (iv) seek treatment opinions from the provider. For example, when seeing a doctor in the gynaecology department, the SP reported vaginal itching and discharge. When the doctor queried related history, the SP raised the concern that her husband was having an affair and sometimes engaged in commercial sex, which might be the cause of her vaginal symptoms. Then the SP presented a fake HIV-positive report, indicating it was from another hospital. The SP asked the doctor what the testing result meant and sought the doctor's advice whether she could get pregnant under the circumstances. SPs recorded his or her perceptions of the doctors' attitudes and behaviours.

Evaluation
Immediately after the visit, the SPs used a standardized structured evaluation checklist to rate service providers on the following three categories:

(i) General satisfaction. SPs evaluated service providers on three dimensions: general impression, interaction with patients and quality of care. These three aspects were rated on a 10-point scale from 0 (the worst) to 10 (the best). A summary satisfaction score was generated by summing the above three responses.

(ii) Perceived stigmatizing attitude and behavior. SPs coded the level of fear of getting infected, the level of avoidance, other unfair treatment and disrespectfulness of patient privacy on a 5-point scale: 1 (none), 2 (a little), 3 (to some extent), 4 (very much) or 5 (an extreme amount). A summary stigma score was calculated by summing the above four items.

(iii) Universal precaution compliance. In a previous publication, we reported that a lack of universal precaution compliance among service providers in China[23] correlated with perceived occupational risk and avoidance attitudes toward PLH.[22] In this study, SPs appraised each service provider's level of compliance to universal precaution protocols using a 5-item scale. These items included whether gloves were worn when performing intrusive procedures, whether hands were washed between patient contacts, whether a surgical mask was worn when contact with blood and bodily fluids was likely, whether the provider avoided recapping used needles and whether the provider placed used sharps in a puncture-resistant container.[24] Each area was rated on a 3-point scale: 1 (not compliant), 2 (partially compliant) or 3 (strictly compliant). A score for universal precaution was calculated by summing all five items.

In addition to the closed-ended checklist, the SPs also took field observation notes about each visit and recorded their general observations and comments using a free-response format.

Data analysis
SAS statistical software (version 9.2) was used to analyse the data collected from the checklist. For each item on the checklist, random-effect ANOVA was used to compare the differences between the intervention and control groups in order to account for the clustering effect within hospital. Field observation notes were reviewed jointly by the research team, after which the qualitative responses were divided into meaningful analytical segments. The coding process was completed in Chinese, and the results were later translated into English using a translate-back-translate procedure. Some of the SPs' open-ended remarks were quoted. Please note that due to the brief nature of the field notes, we were unable to systematically analyse the qualitative data.

Results
The SPs completed a total of 217 assessments in the 40 participating hospitals, among which 80 (36.9%) were conducted in Fujian Province and 137 (63.1%) in Yunnan Province; about half (N = 108, 49.8%) of the assessments were conducted in the intervention hospitals. The assessment covered providers in the following departments: internal medicine, surgery, obstetrics-gynaecology (OBGYN), dermatology/sexually transmitted disease (STD), dentistry, outpatient/emergency, infectious disease and ophthalmology/otolaryngology. Most assessments were conducted with service providers in internal medicine (N = 43, 19.8%), followed by OBGYN (N = 39, 18.0%), and surgery (N = 35, 16.1%). The majority of the service providers assessed were doctors (N = 162, 74.7%) and relatively fewer were nurses (N = 35, 16.1%). The distribution of departments and professions was comparable between the intervention and control groups (Table 1).

Comparison of standardized patient evaluation between intervention and control groups
The service providers in the intervention group received higher scores in all three aspects of the general evaluation (Table 2): the mean score of general impression was 7.3 for the intervention and 6.5 for the control ($P = 0.001$); the mean score of interaction with patients was 7.2 for the intervention and 6.2 for the control ($P < 0.001$); and the mean for quality of care was 7.2 for the intervention and 6.4 for the control ($P = 0.001$). The summary satisfaction score was higher for the intervention group (21.7) than the control group (19.1; $P < 0.001$).

Table 1 Description of the SP assessment sample

Sample	Total (N = 217) N (%)	Intervention (N = 108) N (%)	Control (N = 109) N (%)	P[a]
Department				0.749
Internal medicine	43 (19.8%)	21 (19.4%)	22 (20.2%)	
Obstetrics-gynaecology (OBGYN)	39 (18.0%)	19 (17.6%)	20 (18.4%)	
Surgery	35 (16.1%)	15 (13.9%)	20 (18.4%)	
Outpatient/emergency	30 (13.8%)	17 (15.7%)	13 (11.9%)	
Dentistry	25 (11.5%)	11 (10.2%)	14 (12.8%)	
Dermatology and STD	21 (9.7%)	9 (8.3%)	12 (11.0%)	
Infectious disease	12 (5.5%)	8 (7.4%)	4 (3.7%)	
Ophthalmology/otolaryngology	8 (3.7%)	5 (4.6%)	3 (2.8%)	
Paediatrics	2 (0.9%)	1 (0.9%)	1 (0.9%)	
Other	2 (0.9%)	2 (1.9%)	0	
Profession				0.940
Doctor-in-chief /associate doctor-in-chief	32 (14.7%)	16 (14.8%)	16 (14.7%)	
Doctor-in-charge	38 (17.5%)	18 (16.7%)	20 (18.3%)	
Doctor-in-residence	92 (42.4%)	47 (43.5%)	45 (41.3%)	
Nurse	35 (16.1%)	18 (16.7%)	17 (15.6%)	
A mix of doctors and nurses	20 (9.2%)	9 (8.3%)	11 (10.1%)	

[a]Chi-square test or Fisher's exact test.

The service providers in the intervention group generally received lower scores in all four items of stigmatizing behaviour and attitude: the mean level of fear of getting infected was 1.6 for the intervention and 2.0 for the control ($P = 0.046$); the level of avoidance was 1.9 for the intervention and 2.3 for the control ($P = 0.011$); the mean score of disrespectfulness was 2.5 for the intervention and 2.9 for the control ($P < 0.001$); and the mean score for other unfair treatment observed or felt was 1.7 and 2.3 for the intervention and the control group, respectively ($P < 0.005$). The summary stigma score was lower for intervention group (6.7) than the control group (8.4; $P = 0.006$).

The score for universal precaution compliance was also higher for the intervention group compared with the control group (11.6 vs 10.5; $P = 0.006$). The summary satisfaction score was correlated with the universal precaution score ($r = 0.335$, $P < 0.001$) and negatively associated with summary stigma score ($r = -0.689$, $P < 0.001$).

Examples of field observation notes

The stigmatizing attitude and/or behaviours felt or observed by SPs were documented in the open-ended observation notes. For example, two doctors who were 'unwilling to perform physical examination on the SP' and 'asked the SP not to touch the tissue box on the table' received a 3 (to some extent) for 'fear of getting infected'. A surgeon was rated 4 (very much) for 'level of avoidance' because he declined the SP's appendectomy request with an excuse that 'their hospital was not equipped to do such an operation'. A dermatologist 'placed the HIV-positive report on the table and discussed the SP's infection status when other patients were present' and thus was rated 4 for 'disrespectfulness of patient privacy'. A gynaecologist suggested the SP not get pregnant because 'the chance of having an HIV-positive baby is 100%' and thus was scored 3 for 'other unfair treatment'.

Feasibility of SP method

A typical SP evaluation took on average 1.5 h to complete (10 min for patient registration; 40 min waiting time; 20 min for the service provider meeting; and 20 min to complete the evaluation form). The cost of SP evaluation included the patient registration fee, SP training cost, transportation and per diem and incidental expenses which were no higher than for a traditional face-to-face interview. The SPs noted that it was relatively easy to approach the providers and follow the pre-designed four-step conversation in most cases. SPs indicated that the fake HIV-positive test report was particularly useful in eliciting responses from providers. Based on the SPs' observations, no providers identified them as 'fake patients' and most SPs were prescribed medication and/or laboratory testing just as other patients. At the time of

Table 2 Comparison of the SP evaluation between intervention and control conditions

Evaluation	Intervention (N = 108) Mean (SD)	Control (N = 109) Mean (SD)	P^a
Overall evaluation			
General impression			
0 (worst) to 10 (best)	7.3 (1.8)	6.5 (1.5)	0.001
Interaction with patients			
0 (worst) to 10 (best)	7.2 (1.8)	6.2 (1.5)	<0.001
Quality of care			
0 (worst) to 10 (best)	7.2 (1.9)	6.4 (1.6)	0.001
Summary satisfaction score	21.7 (5.4)	19.1 (4.2)	<0.001
Stigmatizing attitudes and behaviours			
Level of fear of getting infected			
1 (none) to 5 (extreme amount)	1.6 (1.1)	2.0 (1.2)	0.046
Level of avoidance			
1 (none) to 5 (extreme amount)	1.9 (1.3)	2.3 (1.4)	0.011
Disrespectfulness of patient's privacy			
1 (none) to 5 (extreme amount)	2.5 (0.8)	2.9 (0.8)	<0.001
Unfair treatment observed or felt			
1 (none) to 5 (extreme amount)	1.7 (1.1)	2.3 (1.2)	0.005
Summary stigma score	6.7 (3.7)	8.4 (4.0)	0.006
Universal precaution compliance			
Five items with 1 (not compliant) to 3 (strictly compliant); range: 5–15	11.6 (1.64)	10.5 (1.66)	0.006

[a]Random-effect ANOVA.

this writing, no provider or hospital staff member has reported to the research team about an SP visit. None of the SPs, providers or hospitals described in this study expressed concerns about the extra burden posed by SP evaluation.

The pattern of scoring for each SP was examined. The average stigma score given by all SPs ranged from 6.00 to 11.56, all within one standard deviation (SD) away from the mean stigma score (8.59 ± 3.91). Although there were some differences in individual SP average scores, they should not be major concerns for the validity of the comparison results since the SPs were evenly assigned to the intervention and control hospitals ($P = 0.9989$).

Discussion

In the effort to reduce HIV-related stigma among healthcare providers, it is important to accurately measure the level of stigma and monitor its change over time. Self-reported results are subject to systematic under- or over-reporting biases that are often introduced to the findings.[25] The SP method has been widely used in the medical field to evaluate the services provided by clinicians and

pharmacists,[14,26,27] but this study is the first to use SPs to assess HIV-related stigma and discrimination in healthcare settings in China.

The study proved the feasibility of using the SP method to evaluate levels of stigma and discrimination among service providers. The SP method is inexpensive, easy to carry out without detection and not burdensome for researchers or participants. The SP assessment also yielded useful data regarding HIV-related stigma in healthcare settings. Analysis of the SP checklist used in this study showed less stigmatizing attitudes and behaviours among service providers in the intervention group, a result that validated the service providers' self-reports.[20] Qualitative field observation notes could serve as useful feedback in guiding providers to reduce HIV-related stigmatizing attitudes and behaviours in their medical practice.

Advantages of the SP method include the fact that SPs subjectively reported the service providers' responses in the simulation, and the unannounced visits reflected the service providers' usual practice. The accuracy and validity of using SPs has been reported in previous studies.[12] Moreover, some scholars have pointed out that this evaluation technique is more cost effective than a traditional client survey.[28,29] On the other hand, a few limitations of

the SP approach must be noted. First, it was difficult to enlist a representative sample of service providers. In our study, service providers who worked in outpatient or emergency departments were oversampled, and the providers in the inpatient wards and/or operating rooms were difficult to approach. Second, the optimal time to perform an SP assessment is not easy to determine, given the variability in service providers' working schedules. Third, intra- and inter-rater inconsistency could be a concern.[30] Finally, SPs may not be able to simulate all relevant symptoms and emotional states.

The ethical concerns of SP assessment warrant specific comment. The nature of SP assessment precluded us from obtaining providers' prior consent to being observed at certain times. Also, an SP may divert a provider's attention away from care for real patients in order to attend to a 'fake' patient. Walker and George[29] have commented that SP assessments could potentially undermine the honesty and trust between a provider and patient. To ensure ethical conduct, all SPs in this study underwent training in research ethics and obtained an online training certificate. They also signed a confidentiality statement prior to participation. The providers were informed of possible SP observation at the beginning of the study and were given the opportunity to opt out of participation.

There were also limitations in the design of the study that deserve attention in future applications of the SP method. First, no baseline SP assessment was available to demonstrate the change of stigma over time. Second, the type of SP guise used during assessment was not recorded on the evaluation form, so it is not possible to evaluate whether the random allocation of the types was achieved across hospitals. In future applications, it will be crucial to document the type of guise used by SPs. Third, SPs in this study had formal medical training that potentially influenced their judgment of service providers' evaluations. Future studies could test the validity of using SPs from different backgrounds. Finally, although we conducted training and provided instructions for SPs in evaluation and scoring, some SPs had higher average scores than others, which suggests a lack of inter-rater consistency. In future SP studies, more intensive training in coding is needed. One way to achieve this would be to have the SP team code several mock scenarios together and discuss any observed discrepancies in coding.

In conclusion, this study attempts to raise awareness of the SP method as a useful tool to measure stigma in medical settings and to provide valuable supplemental information to self-reported outcome assessments in stigma-reduction intervention trials.

KEY MESSAGES

- This paper describes the procedures of using the standardized patient approach to assess service providers' stigmatizing behaviours and attitudes.

- The study demonstrated that the standardized patient approach is a feasible and useful tool to measure discriminating behaviours and attitudes in medical settings.

- The standardized patient approach could provide an alternative data collection method for future intervention trials and feedback for service improvement.

References

1 Guest G, Bunce A, Johnson L, Akumatey B, Adeokun L. Fear, hope and social desirability bias among women at high risk for HIV in West Africa. *J Fam Plann Reprod Health Care* 2005;**31:**285–87.

2 Pequegnat W, Fishbein M, Celentano D *et al*. NIMH/APPC Workgroup on Behavioural and Biological Outcomes in HIV/STD Prevention Studies: A position statement. *Sex Transm Dis* 2000;**27:**127–32.

3 Vinski MT, Watter S. Priming honesty reduces subjective bias in self-report measures of mind wandering. *Conscious Cogn* 2012;**21:**451–55.

4 Ganster DC, Hennessey HW, Luthans F. Social desirability response effects: Three alternative models. *Acad Manage J* 1983;**26:**321–31.

5 Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff NP. Common method biases in behavioural research: a critical review of the literature and recommended remedies. *J Appl Psychol* 2003;**88:**879–903.

6 Tourangeau R, Yan T. Sensitive questions in surveys. *Psychol Bull* 2007;**133:**859–83.

7 Beullens J, Rethans JJ, Goedhuys J, Buntinx F. The use of standardized patients in research in general practice. *Fam Pract* 1997;**14:**58–62.

8 Wallace P. Following the threads of an innovation: The history of standardized patients in medical education. *Caduceus* 1997;**13:**5–28.

9 Epstein RM, Levenkron JC, Frarey L, Thompson J, Anderson K, Franks P. Improving physicians' HIV risk-assessment skills using announced and unannounced standardized patients. *J Gen Intern Med* 2001;**16:**176–80.

10 Gibbons SW, Adamo G, Padden D *et al*. Clinical evaluation in advanced practice nursing education: Using standardized patients in health assessment. *J Nurs Educ* 2002; **41:**215–21.

11 Petrusa ER. Clinical performance assessments. In: *International Handbook of Research in Medical Education*. Dordrecht, The Netherlands: Kluwer Academic, 2002.

12 Petrusa ER. Status of standardized patient assessment: taking standardized patient-based examinations to the next level. *Teach Learn Med* 2004;**16:**98–110.

13 Bramstedt KA, Moolla A, Rehfield PL. Use of standardized patients to teach medical students about living organ donation. *Prog Transplant* 2012;**22:**86–90.

14 Steinman KJ, Kelleher K, Dembe AE, Wickizer TM, Hemming T. The use of a 'mystery shopper' methodology to evaluate children's access to psychiatric services. *J Behav Health Serv Res* 2012;**39:**305–13.

15 Moriarty H, McLeod D, Dowell A. Mystery shopping in health service evaluation. *Br J Gen Pract* 2003;**53:**942–46.

16 Barrows HS. An overview of the use of standardized patients for teaching and evaluating clinical skills. *Acad Med* 1993;**68:**443–51.

17 Hawkins R, Gross R, Beuttel S, Holmboe E, Gliva-McConvey G, Haley H. Use of standardized patients for teaching and evaluating the genitourinary examination skills of internal medicine residents. *Teach Learn Med* 1998;**10:**65–68.

18 Collins JP, Harden RM. The use of real patients, simulated patients and simulators in clinical examinations. AMEE Medical Education Guide 13. *Med Teach* 1998;**20:** 508–21.

19 Rosenhan DL. On being sane in insane places. *Science* 1973;**179:** 250–58.

20 Li L, Wu Z, Liang L-J *et al*. Reducing HIV-related stigma in health care settings: a randomized controlled trial in China. *Am J Public Health* 2013;**103:**286–92.

21 Rogers EM. *Diffusion of Innovations*. 3rd edn. New York, NY: Free Press, 1983.

22 Wu S, Li L, Wu Z *et al*. Universal precautions in the era of HIV/AIDS: perception of health service providers in Yunnan, China. *AIDS Behav* 2008;**12:**806–14.

23 Li L, Lin C, Wu Z, Guan J, Jia M, Yan Z. HIV-related avoidance and universal precaution in medical settings: opportunities to intervene. *Health Serv Res* 2011;**46:**617–31.

24 Center for Disease Control and Prevention. *2007 Guideline for Isolation Precautions: Preventing Transmission of Infectious Agents in Healthcare Settings*. http://www.cdc.gov/hicpac/2007IP/2007isolationPrecautions.html (22 October 2012, date last accessed).

25 Harrington KF, DiClemente RJ, Wingood GM *et al*. Validity of self-reported sexually transmitted diseases among African-American female adolescents participating in an HIV/STD prevention intervention trial. *Sex Transm Dis* 2001;**28:**468–71.

26 Watson MN, Granas A. A systematic review of the use of simulated patients and pharmacy practice research. *Int J Pharm Pract* 2006;**14:**83–93.

27 Mesquita AR, Lyra DP Jr, Brito GC, Balisa-Rocha BJ, Aguiar PM, de Almeida Neto AC. Developing communication skills in pharmacy: a systematic review of the use of simulated patient methods. *Patient Educ Couns* 2010;**78:** 143–48.

28 Finn A, Kayande U. Unmasking a phantom: A psychometric assessment of mystery shopping. *J Retailing* 1999; **75:**195–217.

29 Walker T, George S. Mystery shopping in psychiatric service evaluation—ethical issues. *Psychiatrist* 2010;**34:** 121–22.

30 Woodward CA, Gliva-McConvey G. The effect of simulating on standardized patients. *Acad Med* 1995;**70:**418–20.

Commentary: Key issues to consider for reviewing and designing simulated patient studies

Kenneth J Steinman

Adjunct Assistant Professor, Division of Health Behavior and Health Promotion, The Ohio State University College of Public Health, 359-A Cunz Hall, 1841 Neil Avenue, Columbus, OH 43210, USA. E-mail: kjsteinman@gmail.com

Li and colleagues' worthwhile paper[1] is among a growing number of studies that use a simulated patient (SP) methodology. Many readers may welcome this approach, yet remain unfamiliar with certain specifics of research design that can affect the validity and reliability of such studies. This commentary provides a brief overview of the SP methodology and a few key issues for researchers