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 paper1 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.
to consider when reading published studies or designing their own.

Retailers have long used ‘mystery shoppers’ who anonymously patronize stores in order to describe customers’ experiences and evaluate employees’ performance. Such ratings are valuable because they provide real-world feedback quickly and inexpensively. Increasingly, researchers are adapting this approach to healthcare settings to assess practices as diverse as pharmacy, psychiatry, paediatrics and sexual health services. Examples range from brief telephone encounters for trying to schedule an appointment, to more extensive interactions with clinicians as in the paper by Li and colleagues. Yet what they all have in common is that researchers surreptitiously pose as patients and systematically record specific aspects of the patient-provider encounter.

Whereas some scholars have described the ethical issues related to SP studies, very few resources review methodological rigour. Part of the problem is the lack of consistent medical subject headings used for studies employing an SP methodology. Searching terms like ‘simulated patient’ or ‘standardized patient’ overwhelmingly yields results on training medical and other health professional students, rather than evaluating existing health services. Other studies use terms like ‘audit’ or ‘mystery shopper’ to describe the approach. Absent such guidance, readers should consider a few questions when reviewing SP studies.

How does the study account for inter-rater reliability?

Each SP has a unique personality and characteristics that affect how he/she elicits responses from healthcare providers and then interprets these responses as data. The greater the interaction between SP and healthcare providers, the greater the potential for such idiosyncrasies to bias results. Researchers typically try to address such concerns by training SPs to provide standardized responses to common scenarios and by randomly assigning SPs to specific guises and facilities. Yet published studies often neglect to report whether such efforts were successful as measured by kappa statistics or other tests of ‘inter-rater reliability’. Instead, they simply assume there are no differences, because trying to model such effects requires prohibitively complex statistical models with, for example, individual ratings nested within individual SPs, further nested within individual facilities and then treatment condition. Future SP studies should routinely document inter-rater reliability and consider how it may bias results.

How does the study account for learning effects?

In addition to differences among SPs, an individual SP’s data may also change during the course of the study. As they gain experience collecting data, SPs learn which responses are particularly noteworthy to record and may tailor their actions to elicit such responses more quickly and clearly. Such ‘learning effects’ may confound other effects related to time, such as whether appointment waiting times change between spring and summer. One approach to documenting learning effects is to record the length of time involved in each encounter. Studies should document whether the average duration of an SP-provider encounter changes during the course of the study and if so, how this may bias results.

Did the study minimize the burden to participants?

Because providers’ time is a scarce, valuable resource, researchers should consider how to minimize this burden while maximizing the knowledge to be gained from the study. This is a particular concern for SP studies where staff may be unaware of, let alone consenting to their participation in a study. (Typically an administrator agrees to participate on behalf of a medical practice or system and then notifies all staff that a study will be conducted surreptitiously. Because it may be impractical for individual staff members to withhold consent and the procedures for notifying them may be imperfect, many end up unaware of their participation in a study). For this reason, pilot studies can be valuable to SP studies in two respects. First, they can document the burden a study will place on participants by documenting the number and average duration of encounters. Second, they can identify adjustments to the methodology that will save time and effort without sacrificing data quality. Including such information can not only facilitate approval from institutional review boards, but can also benefit other researchers seeking to replicate the study. SP studies should routinely include discussions of how other scholars might replicate their methodology more efficiently. In Li and colleagues’ paper, for instance, it would be helpful to know how future studies might safely cut corners, perhaps by employing less highly-trained SPs or including fewer hospitals and/or guises.

Li and colleagues’ study is a noteworthy example of the growing utility of SP studies and their ability to generate useful data efficiently. Better cross-referencing of these types of studies will help improve the methodological rigour of this approach.

Conflict of interest: None declared.

References


