

Use of standardized patients in the assessment of medical practice

Robyn M. Tamblyn, PhD

Since the inception of the Canadian Task Force on the Periodic Health Examination in 1976, we have had a mechanism in Canada of identifying preventive interventions that, if instituted by Canadian physicians, are expected to improve the health of the country's population. Yet substantial disparities appear to exist between optimal institution of preventive care recommendations and actual practice.

In articles presented in this issue, Dr. Brian Hutchison and colleagues have attempted to overcome some of the limitations of previous studies by means of blinded evaluation of practice using standardized patients (pages 185 and 197). "Standardized patients" are people trained to present, accurately and reproducibly, the problems that would be evident in a real patient.¹ Theoretically, blinded assessment of practice by means of standardized patients should provide a number of scientific advantages over more conventional methods of practice assessment. First, researchers can select the exact problem they wish to study and can prospectively establish and collect information on case-specific standards of performance. Second, the need to adjust for differences in the kinds of patients seen by different physicians is eliminated because the "same patient" is presented to all physicians. Third, the problems of incomplete documentation with chart audit and overestimation in self-reporting are addressed by having the standardized patient record the actions taken by the physician during the visit. When this information is coupled with chart documentation, as well as the referral, test and prescription recommendations generated by the visit, a more comprehensive and accurate picture of a physician's practice should emerge.

The use of standardized patients has a place in the methodological armamentarium of health care research, but like all methods it has pitfalls. The "value-added" component only prevails if 5 conditions are met: a specific question is posed that can be assessed by a limited number of standardized patient cases, blinding is maintained, the patient's presentation is standardized, the standardized patient accurately records relevant details of the visit, and the aspect of performance to be assessed can be evaluated during the first visit.

The study of physician performance by Hutchison and colleagues is a good example of the use of standardized patients to evaluate a specific research question — the determinants of adherence to evidence-based preventive guidelines. It was feasible to train the standardized patients to assess physician performance, there were no major physical findings that needed to be standardized, and there was minimum risk that the problem presented would have led to invasive interventions during the visit or deferral of care of real patients. These latter issues must be considered in judging the suitability of standardized patient assessment and the ethical and legal implications of blinded practice assessment. If the physician believes that the standardized patient has an acute problem, one that requires immediate intervention, then not only is the standardized patient at risk, but resources may be inappropriately diverted from real patients. For example, in one case involving an elderly person with abdominal pain, the physician requested emergency abdominal ultrasonography, necessitating overtime pay to hospital technicians.² In a second case, a physician wanted to give cortisone injections to a standardized patient presenting with chronic hip pain.²

The main pitfall in using standardized patients to answer questions about practice behaviour is the technical restriction that physicians with closed practices (in the study reported in this issue, 48% of the 480 physicians who responded to the initial survey) have to be excluded. When this limitation is coupled with low participation rates, estimates of the prevalence of preventive care as well as their determinants may be biased.



Editorial

Éditorial

Dr. Tamblyn is with the Departments of Medicine and of Clinical Epidemiology and Biostatistics, McGill University, Montreal, Que.

Can Med Assoc J 1998;158:205-7

‡ See related articles pages 185 and 197



For example, in this study, if physicians with closed practices had a higher patient retention rate and better continuity of care, then adherence to preventive care guidelines might have been underestimated, and the association between studied predictors and practice might also be biased. Because of these problems, standardized patients are probably most valuable for evaluating the effectiveness of interventions designed to change practice. The study population can be restricted to practices in which blinded assessment can be conducted. Although this restriction limits the generalizability of results, it should not bias comparisons between physicians assigned randomly to control and intervention groups.

Physicians can be successfully blinded to the identity of the standardized patient (unblinding or detection rates range from 0% to 18% in different studies³). Detection of the standardized patients is more likely to occur in relatively closed practices (where new patients are uncommon) or when the standardized patients enter the practice by an atypical route. For example, if most of a physician's new patients come to the practice after discharge from a particular hospital, those who enter through another route are immediately suspect. This means that entry into the practice must, in most instances, be tailored to the specific circumstances of the individual physician. The exceptions to this rule are high-volume settings such as walk-in clinics or emergency departments, where there is a diffuse referral base. The most important issue in relation to blinding is that performance in visits where standardized patients have been detected is systematically better than in blinded visits² (as was the case in this study). For this reason, unblinded visits are usually excluded from the analysis, or an indicator variable for blinding status is included in the multivariate analysis, as was done in this study.

The unique advantage of the standardized patient method is that the "same patient" is seen by all physicians. The key question that needs to be asked in this type of study is "How standardized was the standardized patient?" Three parameters are assessed to answer this question. First, did the standardized patients present the same information to every physician? Second, and of equal importance, was the information provided under the same conditions? For example, if the physician had asked about diet, would all of the standardized patients have said the same thing? Alternatively, would some standardized patients but not others have spontaneously volunteered information about their diet and asked for guidance? This type of variation in performance is not uncommon, and it has an impact on physician performance.³ Third, were other problems presented by the standardized patients that were not part of the "official case"? Performance standards are designed to be case specific, so it is not surprising that "unplanned" deviations in the content of the case bias performance measurement. For example, in one study the diagnostic scores

for physicians who saw standardized patients with unplanned problems — an upper respiratory tract illness, a rash or uncontrolled hypertension — additional to the one they were trained to present (NSAID gastropathy) were 10% lower than for physicians who saw only the planned problems.³

In the articles by Hutchison and colleagues, too little information is provided to determine whether the stories presented to the physicians by the standardized patients were in fact standardized. It would appear that few physicians perceived the reason for the visit as an annual examination (of 249 diagnostic codes assigned to claims for these visits, 29 [11.6%] were for an annual examination), and for 105 of the 249 diagnostic codes, the physicians were dealing with a specific health problem (Table 4, page 201), even prescribing treatment in some cases. Problems in standardizing the context of the visit might also have influenced adherence to some of the preventive guidelines. For example, it would be more likely for influenza immunization (a class B recommendation for the 72-year-old man) to be recommended in visits conducted during the immunization season (Sept. 1 through Nov. 30⁴) than in other months. A tanned standardized patient appearing in July would be more likely to receive counselling on sun exposure than a pasty-faced one appearing on a snowy day in December. Physicians are sensitive to the context of the visit — the setting, the personal characteristics of the patient (sex, race, body habitus), the apparent reason for the visit and the patient's response to the physician's recommendations. For example, moderately overweight standardized patients presenting with osteoarthritis of the hip were found to be more likely to receive weight loss counselling as the primary nonpharmacological intervention than a recommendation for physiotherapy. Female standardized patients were found to be more likely to receive a prescription for medication than men presenting with the same problem.⁵ The extent to which these factors influence performance should be reviewed in considering the suitability of standardized patients for a particular study and in developing the selection and training protocol.

The physician's activities in relation to patient counselling, guidance and advice are poorly documented in the medical chart,^{2,6} and thus adherence to preventive care guidelines will be underestimated in studies that use chart audit assessment. Standardized patients can be trained to record accurately the questions asked by the physician during the visit, the physical examination done and the advice offered.⁷ The only hitch occurs when standardized patients engage in a coaching role — the physician hasn't done something listed on the visit checklist, so the standardized patient helps out, volunteering information or dropping cues to motivate the physician to action. This introduces a type of recording error, whereby the standardized patient indicates that the physician has asked about



something or given advice when in fact it was the standardized patient who volunteered the information.³ Monitoring this potential source of bias is extremely difficult, so most investigators conduct periodic unblinded quality-control checks, as in the study presented in this issue. However, in blinded assessment of performance, we have found that standardized patients differ in their recording accuracy and that this has an impact on the measurement of physician performance — the more errors in recording, the higher the physician's performance score. Screening standardized patients for recording accuracy before training or using alternative methods of data collection, such as audiotaping,^{7,8} represent 2 options for minimizing this source of bias.

Although the visit record completed by the standardized patient can be used to document actions taken by the physician, it does not provide information about why the action was taken or how the information was interpreted (e.g., salient findings or diagnosis). The medical chart for the visit is usually retrieved to provide these data, and in some instances a follow-up interview is also conducted.⁹ The authors of this study report that 22.4% of class C interventions and 21.8% of class D interventions were offered or performed. Yet the authors do not know why the physicians conducted these aspects of the physical examination or requested these investigations. Was it part of routine screening, as suggested, or were the mouth exam, thyroid palpation, serum glucose test and urinalysis done to rule out suspected problems?

Blinded assessment of standardized patients is realistically feasible only if the actions that should (or should not) be taken by the physician can be judged on the basis of the first visit. With preventive care, this expectation may not be realistic. During their first encounter with a new patient physicians may wish to establish a relationship and request past charts; they may then dedicate follow-up visits to the institution of relevant preventive care. Some family physicians are also reticent to invest in comprehensive primary care management until they are confident that the patient will remain in their practice. In this study the option of a follow-up visit was available, an option presumably instituted when the physician scheduled a follow-up appointment. My hat is off to the investigators for accommodating the added headaches of arranging follow-up visits. If the option of a follow-up visit is planned as part of study protocol, then details related to the second visit have to be organized and standardized. Test results and consultation results have to be produced and submitted to the physician through the usual route, and the standardized patient's response to treatment instituted or counselling advice offered on the first visit must be programmed and standardized.¹⁰ In the end, though, the addition of follow-up visits raises the question of whether all physicians saw the "same patient."

On a more general note, the validity of quality-of-care assessment rests on the assumption that the standards used to

judge the "quality" of practice are justified by evidence that they are related to health outcome. In theory, this means that the outcomes of patients seen by physicians who adhere to the standard should be better than the outcomes of patients seen by physicians who do not adhere to the standard. One of the strengths of this study was that the standards used to judge performance were based on a critical review of the available literature and consensus on the resulting recommendations. However, the authors would likely be the first to acknowledge that the strength of the evidence supporting the recommendations of the Canadian Task Force on the Periodic Health Examination varies considerably. For example, there is very good evidence that mammography screening reduces the risk of death from breast cancer in women aged 50 to 69 years and further that a physician's recommendation for regular screening is an important determinant of the likelihood that women will undergo screening.¹¹ In contrast, the evidence supporting the effectiveness of physician counselling for seat-belt use, diet counselling and smoke alarms is not very solid, and for some recommendations there is no evidence that physicians can be effective in changing these aspects of behaviour.¹¹ If we are to realize the benefits of prevention in our society, we need to carefully distinguish the preventive care interventions that are most effectively and efficiently delivered by physicians (e.g., mammography screening) from those that may be better instituted by other means (e.g., seat-belt legislation, home-safety standards).

References

1. Barrows HS. *Simulated patients (programmed patients) the development and use of a new technique in medical education*. Springfield (IL): Charles C Thomas; 1971.
2. Norman GR, Neufeld VR, Walsh A, Woodward CA, McConvey GA. Measuring physicians' performances by using simulated patients. *J Med Educ* 1985;60:925-34.
3. Tamblyn RM, Grad R, Gayton D, Petrella L, Reid T. Impact of inaccuracies in standardized patient portrayal and recording on physician performance during blinded clinic visits. *Teach Learn Med* 1997;9:25-38.
4. Fedson DS, Wajda A, Nicol JP, Hammond GW, Kaiser DL, Roos LL. Clinical effectiveness of influenza vaccination in Manitoba. *JAMA* 1993;270:1956-61.
5. Tamblyn R, Berkson L, Dauphinee WD, Gayton D, Grad R, Huang A, et al. Unnecessary prescribing of NSAIDs and the management of NSAID-related gastropathy in medical practice. *Ann Intern Med* 1997;127(6):429-38.
6. Rethans JJ, Martin E, Metsemakers J. To what extent do clinical notes by general practitioners reflect actual medical performance? A study using simulated patients. *Br J Gen Pract* 1994;44:153-6.
7. Carney PA, Dietrich AJ, Freeman DH, Mott LA. The periodic health examination provided to asymptomatic older women: an assessment using standardized patients. *Ann Intern Med* 1993;119(2):129-35.
8. Hoppe RB, Farquhar LJ, Henry R, Stoffelmayr B. Residents' attitudes towards and skills in counseling: using undetected standardized patients. *J Gen Intern Med* 1990;5:415-20.
9. Jennet PA, Tambay JM, Hogan D, Elford WR, Baumber JS, Crutcher RA, et al. The office management of elderly patients with non-steroidal anti-inflammatory drug-induced gastropathy: a descriptive study using standardized patients. *Ann R Coll Physicians Surg Can* 1996;29(8):462-6.
10. Tamblyn RM, Abrahamowicz M, Berkson L, Dauphinee WD, Gayton D, Grad R, et al. First-visit bias in the measurement of clinical competence with standardized patients. *Acad Med* 1992;67(10):S22-4.
11. Canadian Task Force on the Periodic Health Examination. *The Canadian guide to clinical preventive health care*. Ottawa: Health Canada; 1994. Cat no H21-117/1994E.

Reprint requests to: Dr. Robyn M. Tamblyn, Division of Clinical Epidemiology, Rm. R4.21, Royal Victoria Hospital, 687 Pine Ave. W, Montreal QC H3A 1A1