Family Medicine Residents' Performance with Detected Versus Undetected Simulated Patients Posing as Problem Drinkers

Meldon Kahan, MD*,†,‡, Eleanor Liu, MAsc†, Diane Borsoi, MAsc*,†, Lynn Wilson, MD*,†, Joan M. Brewster, PhD**, Mark B. Sobell, PhD††, and Linda C. Sobell, PhD††

*Department of Family Medicine, University of Toronto, Ontario, Canada
†Centre for Addiction and Mental Health, Ontario, Canada
‡Department of Family Medicine, St. Joseph’s Health Centre, Ontario, Canada
**Department of Public Health Sciences, University of Toronto, Ontario, Canada
††Center for Psychological Studies, Nova Southeastern University, Fort Lauderdale, Florida, USA

Abstract: -Background: Simulated patients are commonly used to evaluate medical trainees. Unannounced simulated patients provide an accurate measure of physician performance.

Purpose: To determine the effects of detection of SPs on physician performance, and identify factors leading to detection.

Methods: Sixty-six family medicine residents were each visited by two unannounced simulated patients presenting with alcohol-induced hypertension or insomnia. Residents were then surveyed on their detection of SPs.

Results: SPs were detected on 45 out of 104 visits. Inner city clinics had higher detection rates than middle class clinics. Residents' checklist and global rating scores were substantially higher on detected than undetected visits, for both between-subject and within-subject comparisons. The most common reasons for detection concerned SP demographics and behaviour; the SP “did not act like a drinker” and was of a different social class than the typical clinic patient.

Conclusions: Multi-clinic studies involving residents experienced with SPs should ensure that the SP role and behavior conform to physician expectations and the demographics of the clinic. SP station testing does not accurately reflect physicians’ actual clinical behavior and should not be relied on as the primary method of evaluation. The study also suggests that physicians’ poor performance in identifying and managing alcohol problems is not entirely due to lack of skill, as they demonstrated greater clinical skills when they became aware that they were being evaluated. Physicians’ clinical priorities, sense of responsibility and other attitudinal determinants of their behavior should be addressed when training physicians on the management of alcohol problems.

Key Words: Simulated patients; family medicine residents; medical education; problem drinking
affiliated with the University of Toronto Department of Family Medicine. Fifty-six second-year family medicine residents were recruited at city-wide educational rounds on an unrelated topic, representing almost all of the residents who attended the rounds (there were 100 residents in the entire program). During the seminar, residents provided consent and completed a questionnaire about their attitudes towards and knowledge about problem drinking.

**Chief complaint** - Two SP roles were developed, with hypertension or insomnia as presenting complaints. These problems were chosen because they are common, and often caused or made worse by alcohol use. Both the hypertension and insomnia script were similar except the opening problem.

The hypertensive patients opened the interview by telling the resident that their blood pressure had been found to be high at a walk-in clinic, and they were advised to see a family physician for follow-up. They had just moved and wanted a new doctor closer to their place of work and home, hence their visit today. The patients with insomnia began the interview by telling the resident that they had been having trouble sleeping for several months. They wanted to know why, and what they should do to sleep better.

SPs were instructed not to disclose any aspect of their alcohol consumption unless directly asked. When asked, however, they were to respond honestly. Males consumed 36 American standard drinks per week and females consumed 29 drinks per week, or about 18 drinks above recommended Canadian low risk drinking guidelines (17 drinks per week for males, 11 per week for females).

To control for physician perceptions about drinkers based on demographic variables, both the male and female SPs had a similar role for all seven clinics. Each SP was a married accountant. The roles were reviewed for credibility by several staff family physicians, and pre-tested with family medicine residents. The SPs were instructed to not discuss alcohol unless asked, to respond honestly to questions, and to agree with the residents’ recommendations. Sixteen professional SPs from the Standardized Patient Department completed fifteen to eighteen hours of training for the role.

**Patient encounters** - Each resident was then visited at their clinic by two unannounced SPs, a male with hypertension and a female with insomnia, or vice versa. Male and female SP pairs were assigned to each resident at random. Immediately after the visit, the SP completed a detailed yes/no checklist and two global rating scales: the Patient Satisfaction Questionnaire, and the Alcohol Skills Rating Form.

Each of the sixteen SPs completed an average of seven visits, or one visit for each of the seven different clinics. The clinics varied in size and number of resident subjects. Within each clinic, each of the four scenarios (female or male with hypertension or insomnia) were presented on average four times. In other words, four SPs presented the female insomnia, four different SPs the female hypertension role, and so on.

**Methods to avoid detection** - During the recruitment seminar, residents were informed that one or two SPs would schedule an appointment with them at their clinic, but they were not told the date or reason for the visit. To avoid being recognized, an SP never made more than one visit per clinic (see above). The Ontario Ministry of Health provided standard health cards with fictitious names and addresses. The SPs themselves called to book appointments. When necessary, booking secretaries were informed of the study and asked not to reveal the names of the SPs to the residents. Billing clerks were also told of the study so false claims would not be submitted to the Ministry. The SPs were supplied with false work and home addresses close to the clinic site, and false names and addresses for their current physicians.

The resident’s supervisor was informed of the SP visit schedules, and asked not to observe the interview or intervene in any way. If the resident mentioned the case, staff physicians were asked to avoid the topic of alcohol, and to agree with (or remain noncommittal towards) the management plan presented by the resident.

**Resident post-visit questionnaire** - After all the visits were completed, residents were mailed a survey asking them if they detected either of the two SPs, and if so, to list corroborating details such as name, date of visit, presenting complaint and occupation. They were also asked to specify when they were identified (before, during or after the visit), and to describe what alerted them (e.g., the SP’s appearance or manner, or other staff). They were asked to list the most important factors that helped them detect the SP, and to suggest ways to avoid detection in future studies.

**Analysis** - An SP was designated as detected if the resident correctly listed any one of the following items: SP’s name, visit date, presenting complaint, or
Subjects - Fifty-six second year family medicine residents agreed to participate in the study, representing almost all of the residents who attended the educational sessions, and one-half of the 110 residents in the program.

Detection rates - Of the 56 residents, 52 completed the post visit questionnaire, for a response rate of 93%. Responses on the portion of the survey asking for corroborating details indicated that 45 out of 104 SP visits were detected by the residents (25 male visits and 20 female visits, p >0.05). Based on the investigators’ knowledge of the site’s location and patient population, sites were divided into “inner city”, “middle class” and “mixed”. The inner city sites had a significantly higher detection rate than middle class sites (60% vs. 31.8%; χ² (df =1) = 4.92, p = 0.03) (Table 1).

Performance with Detected vs. Undetected SPs - Residents had substantially higher assessment and management scores and higher global rating scores when they detected SPs than when they didn’t detect the SPs (Table 2). The proportion of visits in which SPs were asked or given advice about alcohol was significantly higher for detected than undetected SPs on five of nine assessment items and four of nine management items (Table 3).

Intra-subject Analysis: Detected vs. Undetected - The intra-subject performance of residents who detected only one SP was analyzed (n=15). Residents had significantly higher scores on the assessment and management checklists and the Alcohol Skills Rating Form with the SP they detected than with the SP they failed to detect (Table 4).
Table 3
Proportion of Visits in which Residents Asked About Alcohol and Gave Advice: Detected Vs. Undetected Visits

| Questionnaire: Reasons for detection | Undetected visits (n=59) | Detected visits (n=45) | Chi square | p   |
|-------------------------------------|-------------------------|------------------------|------------|-----|
| Assessment                          |                         |                        |            |     |
| Specific alcohol consumption        | 47 (79.7)               | 45 (100)               | 8.5        | .004|
| Problem with alcohol                | 27 (45.8)               | 34 (75.6)              | 9.3        | .002|
| Withdrawal symptoms                 | 10 (16.9)               | 13 (28.9)              | 2.1        | .15 |
| Morning drinking                    | 20 (33.9)               | 24 (53.3)              | 4.0        | .047|
| Social problems                     | 28 (47.5)               | 25 (55.6)              | .67        | .41 |
| Health problems                     | 15 (25.4)               | 27 (60.0)              | 12.7       | <.001|
| Drinking and driving                | 12 (20.3)               | 22 (48.9)              | 9.5        | .002|
| Advice and recommendations          |                         |                        |            |     |
| Discussed relationship between alco-| 37 (62.7)               | 41 (91.1)              | 11.0       | .001|
| hol & presenting problem            |                         |                        |            |     |
| Advised reduced drinking            | 36 (61.0)               | 40 (88.9)              | 10.1       | .001|
| Advised abstinence                  | 5 (8.5)                 | 10 (22.2)              | 3.9        | .048|
| Recommended specific weekly or daily| 35 (59.3)               | 35 (77.8)              | 4.0        | .047|
| amount                              |                         |                        |            |     |
| Gave at least one specific tip for  | 18 (30.5)               | 22 (48.9)              | 3.6        | .056|
| reducing consumption                |                         |                        |            |     |
| Recommended formal treatment        | 4 (6.8)                 | 7 (15.6)               | 2.1        | .15 |
| Gave diagnosis/label                | 8 (13.6)                | 10 (22.2)              | 1.3        | .25 |
| Requested blood work                | 25 (42.4)               | 19 (42.2)              | .00        | .99 |
| Arranged follow-up                  | 49 (83.1)               | 40 (88.9)              | .71        | .40 |

Table 4
Within-Resident Comparison of Checklist and Patient Satisfaction Scores of Fifteen Residents who Detected only One SP

| Variable                          | Detected visit | Undetected visit | p   |
|-----------------------------------|----------------|-----------------|-----|
| Checklist items                   |                |                 |     |
| Alcohol assessment                | 7.3 (2.7)      | 3.7 (4.1)       | .006|
| Alcohol treatment                 | 5.5 (2.0)      | 3.5 (2.1)       | .001|
| Global rating scales              |                |                 |     |
| Alcohol Skills Rating Form        | 5.9 (.95)      | 5.1 (1.2)       | .016|
| Patient Satisfaction Questionnaire| 4.2 (.72)      | 3.9 (.83)       | .229|

Post-visit Questionnaire: Reasons for detection. - The portion of the survey on reasons for detection was completed for 36 of the 45 detected visits. The most common reasons were the history, presenting complaint, behavior, appearance, occupation, or remarks of the SP (Table 5). The actions of the SP after the interview also contributed to their detection:

Did not arrange a follow-up visit, refused or did not do blood work. Other contributing factors included the remarks of the supervising MD or other staff. For example, on one occasion, the resident became suspicious when the supervisor, normally quite directive, refused to give the resident any advice.
Thirty-four of the respondents provided written reasons for their detection. Consistent with the list portion of the survey, the most frequently cited reason (n=11) was that the patient’s demographic features or social background did not fit the resident’s practice. For example, well-dressed, professional men or women were not as typical of their practice as unemployed patients, psychiatric patients, or blue collar workers. Another frequently cited reason (n=10) was that the SP “did not act like a real patient”, for example, their responses seemed staged and not genuine.

Eight respondents felt that the patient’s questions or comments about alcohol did not appear realistic; for example, the patient was too forthcoming in providing the alcohol history. Five residents felt the patient role was not credible — too ‘neat’, too much like a case description. In some cases, details alerted the resident (e.g., the patient did not know the proper designation of accountant).

Five residents became suspicious because they rarely saw new patients. Nine residents were alerted by supervisors, other residents or office staff, or previous encounters with the SP. Nine residents were alerted when the SPs didn’t go for blood tests or make follow-up appointments. It is not clear whether...

---

### Table 5
Factors Alerting Resident to SP

| List of factors         | Number (%) |
|-------------------------|------------|
| History                 | 20 (55.6)  |
| Behavior of SP          | 16 (44.4)  |
| Presenting complaint    | 13 (36.1)  |
| No follow-up visit      | 13 (36.1)  |
| Appearance              | 11 (30.6)  |
| Occupation              | 6 (16.7)   |
| Patient’s chart         | 6 (16.7)   |
| Refusal of blood work   | 5 (13.9)   |
| Remarks of SP           | 5 (13.9)   |
| Remarks of supervising MD| 2 (5.6)    |
| Receptionist            | 1 (2.8)    |
| Other MD                | 1 (2.8)    |
| Saw patient previously  | 1 (2.8)    |

*a Residents could choose multiple factors  
*b Divisor = 36, which is the number of residents who checked one or more factors in the list provided

### Table 6
Residents’ Suggestions for Avoiding Detection

1. Run study over longer time period so residents won’t anticipate the SP visit
2. Have the SP reflect the clinic’s patient population
3. Book patients during ‘block time’ when new patients are seen
4. Make the SP’s excuse for seeking a new FP more believable (not ‘I’m new in the city’)
5. Have the SP respond more like a real patient:
6. Give vaguer answers to questions
7. Don’t ask ‘textbook questions’ such as “what are the side effects of alcohol”?
8. Respond less positively to advice
9. Keep follow-up appointments
10. Vary the histories (residents talk)
11. The presenting complaint should not be alcohol-related
12. Have two visits, with the first visit unrelated to alcohol
13. Have the SP present with a minor complaint such as a cold, and mention alcohol casually
14. Ask residents not to speak to each other about the case
their suspicions were aroused during or after the interview. Four residents were definitely alerted after the interview, for example, the SP visit was later discussed with the supervisor or other residents, or because ‘everyone smiled’ afterwards.

Eighteen residents provided suggestions for how to avoid SP detection in future studies (Table 6). The main suggestions were: Train the SPs to act more like real patients (e.g., give vaguer responses, don’t react so positively to advice about alcohol), make the presenting complaint unrelated to alcohol use, book the SP during family medicine block time when new patients are seen, have a better excuse for why the patient has no family doctor; vary the SP’s history within each clinic, and keep follow-up appointments.

Discussion

Simulated patients are increasingly being used to teach clinical skills,12 test physicians’ clinical competence13,14 and evaluate the effectiveness of an educational intervention.15 However, physician performance with detected SPs in OSCE stations correlates only moderately with other measures of clinical performance, such as faculty assessments16-19. One possible explanation is that physicians perform differently when they know they’re being observed. This study confirms that the knowledge that a patient is simulated has a marked effect on physician behavior, and use of known SPs may be a less valid measure of physician behavior than use of undetected SPs.

Factors contributing to detection: - A number of factors contributed to the residents’ detection of SPs. The residents’ suspicions were raised when the SP’s social background did not reflect that of the clinic population. The SPs in this study portrayed middle class professionals, and inner city clinics had significantly higher detection rates as a result. Modifying the SP role to reflect the nature of each clinical setting may reduce detection rates, but this makes it more difficult to make controlled comparisons of physician behavior across clinics. Multi-clinic studies involving unannounced SPs should consider choosing clinics with similar patient demographics.

Residents were suspicious of patients whose behavior did not conform to their expectations of problem drinkers. For example, the SPs were direct and specific in their alcohol histories, and readily agreed to resident’s advice to change. Residents were expecting patients to be evasive, disinterested or resistant to physician advice. SP roles may need to be modified for alcohol and other medical conditions that are accompanied by preconceived physician expectations about patient behavior.

Residents’ previous experiences with SPs enabled them to detect general clues to an SP visit, such as a ‘textbook case scenario’, ‘classic SP questions’, lack of genuineness, greater interest in the subject of alcohol than other parts of the interview, refusal of blood-work or lack of a follow-up interview. A number of valid suggestions were made to address this, for example: make the presenting complaint unrelated to alcohol, have patients accept follow-up visits and blood-work.

In a number of cases, staff alerted residents to the SP visit, despite their prior agreement to avoid this. Closer contact between investigators and clinic staff might minimize this, although this can be a challenge in multi-site clinics with many receptionists, nurses, and staff doctors. Residents identified other residents as a major source of information about SP visits. Future studies should ask residents to refrain from alerting other residents, and should consider varying the SPs’ presentations from clinic to clinic.

Reasons for better performance with detected patients - Residents had checklist scores 50% higher with detected than undetected patients. This suggests that physicians do not always choose to apply their skills and knowledge in clinical encounters. When the resident realized the patient was simulated the visit turned into a focused clinical examination, and the residents were immediately able to display their full range of clinical skills.

The superior performance of residents with detected SPs could in part reflect greater clinical skills in residents who are able to detect SPs; however, performance with detected SPs was substantially better on within-subject comparisons of residents who detected one SP but not the other, suggesting that detection and not resident skill is the determining factor.

Implications for medical education - This study demonstrates that SP encounters measure clinical skill and performance, but not actual clinical behaviour. The latter is most realistically measured by close observation and structured, objective evaluation by the residents’ clinical supervisors.

The study also has implications for the teaching of medical issues which are not always viewed by physicians as a core medical responsibility or priority, such as identification or counseling for substance use. Despite evidence from controlled trials that even brief physician advice will reduce alcohol consum-
tion and alcohol-related morbidity, physicians continue to perform poorly in identifying and managing alcohol problems. The current study suggests that lack of skill does not fully explain physicians' performance with drinkers, as their performance improves markedly when aware that they're being evaluated. Other factors must be at play, such as lack of time per visit, their clinical priorities, their sense of responsibility towards addressing treating alcohol problems, and their confidence and optimism that their interventions will make a difference. Training on management of alcohol problems must address these factors as well as improving clinical skills.

Conclusion

Residents' clinical performance substantially improves when they are aware that an unannounced patient is simulated. SPs are more likely to be detected if they do not fit the demographic profile of the residents’ practice, if they do not conform to the residents’ preconceived notions of how patients with that condition behave, and if they act in ways the resident recognizes as ‘typical’ of SPs. The results suggest that residents’ performance with SPs does not fully reflect their actual clinical behavior, and SP stations should not be used as the primary measure of their performance. The study also suggests that physicians’ poor performance in identifying and managing alcohol problems is not entirely due to lack of skill, but also likely reflects attitudes towards intervening with drinkers.

Acknowledgements

We would like to thank Dr. Lorne Becker for his contribution to the design of this study.

This study received funding and ethics approval from the Addiction Research Foundation division of the Centre for Addiction and Mental Health in Toronto, Ontario.

References

1. Luck, J. and J.W. Peabody, Using standardized patients to measure physicians' practice: validation study using audio recordings. BMJ, 2002. 325(7366): p. 679.

2. Glassman, P., et al., Using standardized patients to measure quality: Evidence from the literature and a prospective study. Journal on Quality Improvement, 2000. 26(11): p. 644-653.

3. Norman, G.R., P. Tugwell, and J.W. Feightner, A comparison of resident performance on real and simulated patients. J Med Educ, 1982. 57(9): p. 708-15.

4. O'Hagan, J.J., L.J. Davies, and R.K. Pears, The use of simulated patients in the assessment of actual clinical performance in general practice. N Z Med J, 1986. 99(815): p. 948-51.

5. Saebu, L. and J.J. Rethans, Management of patients with angina pectoris by GPs: a study with standardized (simulated) patients in actual practice. Fam Pract, 1997. 14(6): p. 431-5.

6. Beullens, J., et al., The use of standardized patients in research in general practice. Fam Pract, 1997. 14(1): p. 58-62.

7. Rethans, J.J., et al., A method for introducing standardized (simulated) patients into general practice consultations [see comments]. Br J Gen Pract, 1991. 41(344): p. 94-6.

8. Woodward, C.A., et al., Measurement of physician performance by standardized patients. Medical Care, 1985. 23(8): p. 1019-1027.

9. Wilson, L., et al., Physician behavior towards male and female problem drinkers: a controlled study using simulated patients. J Addict Dis, 2002. 21(3): p. 87-99.

10. Bondy, S., et al., Low-risk drinking guidelines: the scientific evidence. Can J Public Health, 1999. 90(4): p. 264-70.

11. Anonymous, Patient Satisfaction Questionnaire. 1992, Philadelphia: American board of internal medicine.

12. Madan, A.K., et al., Comparison of simulated patient and didactic methods of teaching HIV risk assessment to medical residents. Am J Prev Med, 1998. ;15(2): p. 114-9.

13. Hilliard, R.I. and S.E. Tallett, The use of an objective structured clinical examination with postgraduate residents in pediatrics [see comments]. Arch Pediatr Adolesc Med, 1998. 152(1): p. 74-8.

14. Kassebaum, D.G. and R.H. Eaglen, Shortcomings in the evaluation of students' clinical...
skills and behaviors in medical school. Acad Med, 1999. 74(7): p. 842-9.

15. Greenberg, L.W., et al., Communicating bad news: a pediatric department's evaluation of a simulated intervention. Pediatrics, 1999. 103(6 Pt 1): p. 1210-7.

16. Stillman, P., et al., Assessment of clinical skills of residents utilizing standardized patients. A follow-up study and recommendations for application [see comments]. Ann Intern Med, 1991. 114(5): p. 393-401.

17. Skinner, B.D., W.P. Newton, and P. Curtis, The educational value of an OSCE in a family practice residency. Acad Med, 1997. 72(8): p. 722-4.

18. Petrusa, E.R., T.A. Blackwell, and M.A. Ainsworth, Reliability and validity of an objective structured clinical examination for assessing the clinical performance of residents. Arch Intern Med, 1990. 150(3): p. 573-7.

19. Ram, P., et al., Assessment of practicing family physicians: comparison of observation in a multiple-station examination using standardized patients with observation of consultations in daily practice. Acad Med, 1999. 74(1): p. 62-9.

20. Fleming, M.F., et al., Brief physician advice for problem alcohol drinkers. A randomized controlled trial in community-based primary care practices [see comments]. JAMA, 1997. 277(13): p. 1039-45.

21. Gentilello, L.M., et al., Alcohol interventions in a trauma center as a means of reducing the risk of injury recurrence. Ann Surg, 1999. 230(4): p. 473-80; discussion 480-3.

22. Kahan, M., L. Wilson, and L. Becker, Effectiveness of physician-based interventions with problem drinkers: a review. CMAJ, 1995. 152(6): p. 851-859.

23. Lang, T., et al., Improving hypertension control among excessive alcohol drinkers: a randomised controlled trial in France. The WALPA Group. J Epidemiol Community Health, 1995. 49(6): p. 610-6.

24. Arndt, S., et al., Screening for alcoholism in the primary care setting: are we talking to the right people? J Fam Pract, 2002. 51(1): p. 41-6.

25. Friedmann, P.D., et al., Screening and intervention for alcohol problems. A national survey of primary care physicians and psychiatrists. J Gen Intern Med, 2000. 15(2): p. 84-91.

26. Poulin, C., I. Webster, and E. Single, Alcohol disorders in Canada as indicated by the CAGE questionnaire. Cmaj, 1997. 157(11): p. 1529-35.

27. Andreasson, S., K. Hjalmarsson, and C. Rehman, Implementation and dissemination of methods for prevention of alcohol problems in primary health care: a feasibility study. Alcohol Alcohol, 2000. 35(5): p. 525-30.

28. Midmer, D., et al., Medical faculty members' perspectives on substance use disorders: A survey and focus-group study. Annals of the Royal College of Physicians and Surgeons of Canada, 2002. 35(8 supplement 1): p. 1-6.

29. Saitz, R., et al., Professional satisfaction experienced when caring for substance-abusing patients: faculty and resident physician perspectives. J Gen Intern Med, 2002. 17(5): p. 373-6.

Correspondence

Dr. Meldon Kahan
St. Joseph's Health Centre
Dept. of Family Medicine
30 The Queensway,
Toronto, Ontario M6R 1B5 Canada
Tel: 416-530-6478
Fax: 416-530-6160
Email: kahanm@stjoe.on.ca