

---

## Research Briefs

---

# Caught in the Act? Prevalence, Predictors, and Consequences of Physician Detection of Unannounced Standardized Patients

*Carol E. Franz, Ron Epstein, Katherine N. Miller, Arthur Brown, Jun Song, Mitchell Feldman, Peter Franks, Steven Kelly-Reif, and Richard L. Kravitz*

---

**Objective.** To examine the prevalence, predictors, and consequences of physician detection of unannounced standardized patients (SPs) in a study of the impact of direct-to-consumer advertising on treatment for depression.

**Data Sources.** Eighteen trained SPs were randomly assigned to conduct 298 unannounced audio-recorded visits with 152 primary care physicians in three U.S. cities between May 2003 and May 2004.

**Study Design.** Randomized controlled trial using SPs. SPs portrayed six roles, created by crossing two clinical conditions (major depression or adjustment disorder) with three medication request scripts (brand-specific request, general request for an antidepressant, or no request).

**Data Collection.** Within 2 weeks following the visit, physicians completed a form asking whether they “suspected” conducting an office visit with an SP during the past 2 weeks; 296 (99 percent) detection forms were returned. Physicians provided contextual data, a Clinician Background Questionnaire. SPs filled in a Standardized Patient Reporting Form for each visit and returned all written prescriptions and medication samples to the laboratory.

**Principal Findings.** Depending on the definition, detection rates ranged from 5 percent (unambiguous detection) to 23.6 percent (any degree of suspicion) of SP visits. In 12.8 percent of encounters, physicians accurately detected the SP before or during the visit but they only rarely believed their suspicions affected their clinical behavior. In random effects logistic regression analyses controlling for role, actor, physician, and practice factors, suspected visits occurred less frequently in HMO settings than in solo practice settings ( $p < .05$ ). Physicians more frequently referred SPs to mental health professionals when visits aroused high suspicion ( $p < .05$ ).

**Conclusions.** Trained actors portrayed patient roles conveying mood disorders at low levels of detection. There was some evidence for differential treatment of detected standardized patients by physicians with regard to referrals but not antidepressant

prescribing or follow-up recommendations. Systematic assessment of detection is recommended when SPs are used in studies of clinical process and quality of care.

**Key Words.** Standardized patients, physician–patient communication, health care delivery

---

Standardized patients (SPs) are people trained to portray patient roles so that practicing physicians cannot distinguish them from real patients (McLeod et al. 1997; Rosen et al. 2004). Research designs using high-quality, unannounced (or covert) SPs may be a “gold standard” for clinical quality assessment in the outpatient arena (Peabody et al. 2000).

A low SP detection rate is often accepted as a proxy for high-quality role portrayal. In a review of 11 SP studies through 1997, detection rates ranging from 0 to 42 percent were reported (Beullens et al. 1997); in our analysis of studies since 1997 rates ranged up to 70 percent (Rethans et al. 1991; Tamblyn et al. 1992; Gallagher et al. 1997; Grad et al. 1997; McLeod et al. 1997; Brown et al. 1998; Carney and Ward 1998; Hutchison et al. 1998; Tamblyn 1998; Woodward et al. 1998; Carney et al. 1999a, b; Glassman et al. 2000; Luck et al. 2000; Epstein et al. 2001, 2005; Gorter et al. 2002; Luck and Peabody 2002; Beaulieu et al. 2003; Maiburg et al. 2004). A table reviewing detection prevalence rates and methods for assessing detection since 1997 is available in an online appendix. Few studies reported on the prevalence of suspicion (i.e., some uncertainty) versus detection. Approaches to assessing detection varied widely: some researchers simply relied on participating physicians to report detected visits (Rethans et al. 1991; McLeod et al. 1997; Peabody et al. 2000; Maiburg et al. 2004); others actively assessed suspicion or detection by informing the physician of an SP visit (2 days to 1 year postvisit), then determined whether the physician identified the SP (Gallagher et al. 1997; Carney and Ward 1998; Hutchison et al. 1998; Carney et al. 1999a, b; Epstein et al.

---

Address correspondence to Carol E. Franz, Ph.D., University of California San Diego, Department of Psychiatry, 9500 Gilman Dr., #0738, La Jolla, CA 92093. Ron Epstein, M.D., and Arthur Brown, M.S., are with the University of Rochester, Family Medicine Center, Rochester, NY. Katherine N. Miller, B.A., is with the Center for Nonviolent Communication, Oakland, CA. Jun Song, M.S., Peter Franks, M.D., and Richard L. Kravitz, M.D., M.S.P.H., are with the University of California Davis, Center for Health Services Research in Primary Care, Sacramento, CA. Mitchell Feldman, M.D., M.Phil., is with the University of California San Francisco, Division of General Internal Medicine, San Francisco, CA. Steven Kelly-Reif, M.D., is with the The Kaiser Permanente Medical Group/Point West, Sacramento, CA.

2001, 2005; Luck and Peabody 2002). Rarely have the effects of detection on outcomes been examined (Tamblyn et al. 1992; McLeod et al. 1997; Hutchinson et al. 1998), or the factors affecting detection been systematically collected. Although detection is likely affected by SP training, contextual, geographic, and cultural factors may also be important (Brown et al. 1998; Epstein et al. 2001). Minimizing and adjusting for detection are critical for valid inferences from SP studies. A priori standardization of the methodology for defining detection is also important.

To explore these issues we used data from the Social Influences on Practice Study (SIPS). SIPS examined the effects of patient (SP) prompting for medication requests on physician behavior (Kravitz et al. 2005). Here, we address three issues: (1) the prevalence of detection in the SIPS; (2) the factors predicting detection; and (3) the effect of detection on treatment decisions.

## METHODS

### *Design Overview*

SPs were trained to portray six roles; roles involved a combination of a mood disorder (depression or adjustment disorder), a musculoskeletal disorder (carpal tunnel syndrome or low back pain), and a medication request type (brand-specific, general, or none). Physicians were randomly assigned two visits involving different clinical presentation/request type combinations. Before consenting, physicians were told the study would involve conducting office visits with two unannounced SPs several months apart, that each SP would present with a combination of common symptoms, and that the purpose of the study was to assess social influences on practice and competing demands on primary care. Physicians agreed to be covertly audio recorded; consents were obtained a minimum of 10 weeks before a visit. Institutional review boards at all participating institutions approved the study protocol. See Kravitz et al. (2005) for complete study details.

SPs conducted visits from May 2003 to May 2004. To reduce detection, the two SP visits were separated by at least 8 weeks. In addition, enrollment was limited to no more than two physicians sharing the same waiting room/station; SPs did not return to the same waiting room. Following each visit, with use of the audio recording, SPs reported key features of the visit using a standardized questionnaire (Standardized Patient Reporting Form [SPRF]). At the end of the study, physicians completed a Clinician Background Questionnaire (CBQ), and were then debriefed. Training staff monitored

SPs' performances and their reliability on the SPRF for within role and between site consistency and accuracy throughout the study.

### *Measures*

*Suspicion/Detection.* Ten to 14 days after an SP visit, physicians were faxed a form informing them an SP *may* have conducted an office visit in the previous 2 weeks and asking "During the past two weeks, did you suspect that you conducted an office visit with a Standardized Patient?" Physicians reported the extent of their suspicion on a 1–5 scale (from "definitely" to "definitely not"). Physicians who responded "definitely," "probably," or "uncertain" completed additional items about the identity of the SP, timing, and reason for suspecting the SP. Physicians reported the realism of the SP portrayal (1 = very realistic to 4 = very unrealistic), and the extent to which they treated the SP like a "real patient" (exactly alike; minor differences; major differences). Physicians were encouraged to make additional comments.

*Measures of Risk Factors for Detection.* Based on prior SP studies, we identified role (medical condition, request), actor (individual SP), physician (age, gender, training), and contextual variables from the dataset to estimate their effect on detection. Physician and contextual data were derived from the CBQ; contextual variables included practice setting (solo, group, HMO, university affiliated), clinical busyness (10 or more patients in a typical half-day clinic), and whether the practice was closed to new patients for 1 month or more at any time during the previous year (yes/no).

*Outcome Measures.* We examined three treatment outcomes key to the SIPS study: antidepressant prescribing, mental health referrals, and follow-up plans. Referrals (yes/no) and follow-up interval (< 1 month versus  $\geq$  longer versus none) were obtained from the SPRF. Good agreement between the SPRF and an independent review of 36 randomly selected visit audio recordings was observed (mean  $\kappa$ , 0.82). Study staff coded prescribing based on prescriptions and samples given to SPs.

### *Statistical Analyses*

Univariate analyses used *t*-tests or  $\chi^2$  tests. Generalized linear-mixed models (GLMM) were used both to predict detection and to assess the consequences of detection controlling for role, SP, physician, and practice site (McCulloch

and Searle 2001). The 18 SPs were entered as a random effect (not significant in any analysis). The GLMM accounts for the study design involving physicians nested within practice sites and SP visits nested within physician. Analyses were performed using SAS, version 8.2.

## RESULTS

Eighteen SPs made 298 visits to 152 physicians in Sacramento ( $n = 101$ ), San Francisco ( $n = 96$ ), and Rochester ( $n = 101$ ); six physicians saw only one SP. Two hundred visits (67 percent) were to general internists and 98 (33 percent) to family physicians, while 201 (67 percent) were to male physicians and 97 (33 percent) to female physicians. The average age of participating physicians was 46 (SD = 9.8, range 30–81); physicians had practiced medicine for an average of 15 years (SD = 9.5, range 2–47). Physicians returned 99 percent (296) of the detection faxes.

### *Prevalence of Detection*

In 15 (5 percent) visits, physicians responded “yes, definitely” that they conducted an SP visit in the last 2 weeks, suspected the SP before or during the visit, and accurately identified the SP. Using a more liberal definition (yes definitely, yes probably, or uncertain that they had seen an SP over the past 2 weeks), the suspicion rate was 23.8 percent. In two visits, physicians misidentified real patients (one male, one black female) as SPs (Table 1).

Most common reasons for detection included “something about the way the person behaved during the visit” (45 percent), and “having a closed practice” (35 percent). Written comments explaining suspicion before or during the visit included: “The presentation was too classic,” “She seemed to be easily satisfied with the explanation I gave, unlike my other patients,” “too picture perfect, wouldn’t do blood work.” Explanations for suspicion after the visit included comments such as “The request for medical records was returned as unknown,” or “Didn’t follow-up with [behavioral health/nerve conduction/blood work].” “My staff told me” and “closed practice” were given as reasons for suspicion both before and after the visit. SPs mildly demurred requests for blood work or additional tests, saying they were pressed for time and would return to the office later. Physicians accepted SPs explanations that they had seen a gynecologist in the past year and in no instance applied pressure for gynecological exams.

Table 1: Detection Rates, Timing of Detection and Physicians' Perception of Impact of Detection on SP Care Based on Detection Fax 2 Weeks Following Visit

<i>Suspected an SP Visit</i>	<i>N (%)</i> , 296	<i>Timing of Suspicion*</i>	<i>N (%)</i>	<i>Impact on Care†</i>	<i>N (%)</i>
Yes, definitely	22 (7%)	Before/during visit	15/22 (68%)	No impact on care	10/15 (67%)
		After visit	7/22 (32%)	Some impact on care	5/15 (33%)
Yes, probably	35 (12%)	Before/during visit	23/35 (66%)	No impact on care †	20/22 (91%)
		After visit	12/35 (34%)	Some impact on care	2/22 (9%)
Uncertain	13 (4%)	Before/during visit*	4/12 (33%)	No impact on care	4/4 (100%)
		After visit	8/12 (67%)	No impact on care	8/8 (100%)
No, probably not	50 (17%)				
No, certainly not	176 (60%)				

\*One physician did not complete timing question on detection form.

†One physician did not complete impact question on detection form.

### *Predictors of Detection*

We operationalized detection in two ways. The “degree of suspicion” (DOS) measure categorized physician detection fax responses into three groups: high suspicion visits (physician responded “yes, certainly” or “yes, probably” on detection fax; HSV;  $N = 57$ , 19 percent), moderate suspicion visits (physician responded “uncertain” or “no, probably not” on detection fax”; MSV;  $N = 63$ , 21 percent), and no suspicion visits (physician responded “No, certainly not”; NSV;  $N = 176$ , 60 percent), regardless of timing or accuracy of suspicion. “Meaningful detection” was defined as occurring if the physician responded “yes certainly” or “yes probably” that they suspected an SP visit, the SP was identified accurately, and suspicion was aroused before or during the visit. The assumption underlying the meaningful detection measure was that suspicions aroused before or during the visit would be more likely to influence treatment outcomes.

Meaningful detection occurred in 38 encounters (12.8 percent). Physicians rated these encounters as less realistic than other suspected visits (mean 1.82 versus 1.39,  $p < .009$ ). Physicians were marginally more likely to say there were minor or major differences in how they treated the meaningfully detected SPs ( $p = .057$ ). However, there were no significant differences in prescribing, referral, or follow-up when physicians who reported treating the detected SPs “just like real patients” were compared with those who stated they “treated detected SPs differently” ( $p > .20$ ).

Meaningful detection occurred in 1.69 percent (1/59) of visits at an HMO, 12.3 percent (9/73) of visits at solo practices, 16.1 percent (20/124) of visits at group practices, and 20 percent (8/40) of visits at university-affiliated practices. Having a closed practice was marginally associated with meaningful detection ( $p < .10$ , data not shown). In regressions that grouped suspected and detected visits together ( $N = 70$ ), practice setting (but not having a closed practice) was significant ( $F = 2.90$ ,  $p < .05$ ); physicians practicing in HMOs were less likely to detect visits than physicians in solo practices.

### *Effect of Detection of Physician Behavior*

Random effects logistic regressions analyzed whether detection affected the primary outcome measures of the SIP study: prescribing, referrals, or follow-up. Regressions were performed separately for DOS and meaningful detection as well as for each of the three physician behaviors (Table 2), controlling for role, actor, physician, and contextual variables. With the DOS measure, high suspicion SP visits but not moderate or no suspicion visits were associated

Table 2: Mixed Effects Regression Controlling for Meaningful Detection or Degree of Suspicion Measure, Request, Condition, Request  $\times$  Condition Interaction, Physician Gender, Speciality, Practice Type, Age Group, Number of Patients per Half-Day Session, Whether the Practice Was Closed, and if the Suspected Visit Was by the First or Second SP

Physician Treatment Behaviors	Meaningful Detection				Degree of Suspicion			
	Odds Ratio	Standard Error	Confidence Intervals (Lower, Upper)	p Value	Odds Ratios: High Suspicion Visits (HSV)/Moderate Suspicion Visits (MSV)*	Standard Error	Confidence Intervals	p Value
Prescribed antidepressant	0.59	0.40	- 1.32; 0.26	.19	HSV versus NSV 0.61	0.37	- 1.21; 0.22	.18
Follow-up < 1 month	1.23	0.38	- 0.54; 0.95	.59	MSV versus NSV 0.76	0.36	- 0.97; 0.43	.45
Mental health referral	2.10	0.40	- 0.04; 1.53	.07	HSV versus NSV 0.94	0.34	- 0.73; 0.61	.85
					MSV versus NSV 0.69	0.34	- 1.04; 0.29	.27
					HSV versus NSV 2.35	0.36	0.16; 1.55	.02
					MSV versus NSV 1.58	0.34	- 0.22; 1.14	p = .18

The SP identifier was entered as a random effect.

\*Note: Comparison group is nonsuspected (NSV) visits.



with a significantly greater likelihood of referral ( $p < .05$ ). There was a marginally significant main effect of meaningful detection on mental health referrals ( $p < .10$ ). Detection was not associated with prescribing or follow-up.

## DISCUSSION

Unannounced SP visits potentially facilitate more realistic assessments of physician behavior than do techniques where physicians know they are observed, for example, with real patients. Research using self-assessment or chart review suggests these sources yield unreliable information about medical practice (Peabody et al. 2000; Gorter et al. 2002; Luck and Peabody 2002; Biernat et al. 2003). However, high detection rates threaten the validity of SP studies. High detection rates suggest poor SP role performance and introduce the potential for physician performance bias. Thus, adequate evaluation of SP detection rates is critical. In our study, we required that all physicians return the detection form, regardless of detection, and collected complete data on practice and physician characteristics.

Detection rates ranged from 5 to 23.6 percent, depending on the definition of detection. These rates are within the range found in prior research using unannounced SP visits. No role or actor characteristics predicted detection. Controlling for physician and contextual characteristics, detection was least likely to occur in HMO settings. In the HMO practices, physicians and their local staff had little control over patient flow or scheduling (appointments were scheduled centrally), possibly allowing SPs to be less conspicuous. Medical staff in other settings tended to be protective of physicians' schedules and sometimes disclosed the SP to the physician. Although in some studies, physicians gave "closed practice" as a reason for detecting an SP (Epstein et al. 2001, 2005), in this study physicians in closed practices were only marginally more likely to detect SPs. Unlike other studies, we assessed closed status for *all* participating physicians rather than just among those who reported being suspicious. Thus, we were able to empirically test hypotheses about the impact of contextual and physician characteristics on detection. Out of the 167 visits that occurred in closed practices, only 24 (14 percent) were detected; solo practices were less likely to be closed to new patients (41 percent) than HMO-based practices (86 percent closed). Solo and closed practices pose a challenge for SP research as new patients are relatively infrequent and SPs often require the assistance of practice staff to arrange a visit, increasing their vulnerability to detection. Omitting these practices, however, would limit generalizability of

study findings. These results pose a problem for SP research aimed at clinical quality assessment as these same practices may also have less institutional oversight.

Although desirable as an indicator of the success of SP training and role portrayal, low detection rates in the SIPS limited our statistical power to examine factors affecting detection. (Tamblyn 1998). Other limitations of the study include the uncertain generalizability of our findings to other practice types, clinical presentations, and other geographic areas of the country. Certain groups of patients or medical conditions may be atypical in some clinical settings, increasing the risk of detection or differences in treatment. SP research, though, could provide a unique window into clinical process for such office visits. Finally, physician behaviors affected by detection may be subtle and not captured by global indicators such as those we analyzed.

In summary, unannounced SP visits are a powerful tool for assessing clinical performance because they represent a relatively fixed clinical “stimulus” and avoid unwanted influences introduced when overtly observing or audio recording physicians. With appropriate training and quality control procedures, we have demonstrated that trained actors conducting unannounced office visits can convincingly portray patient roles to capture actual physician behavior during everyday practice at moderately low levels of detection. Finally, we recommend that researchers evaluate the impact of announced and unannounced SPs on physician behavior, and adjust for detection in data analyses. This is particularly important as quality assurance and recertification exercises increasingly incorporate SP-based assessments. In addition, we recommend developing a protocol as a step toward formulating a consistent and systematic approach to SP detection. Such a protocol might include (a) assessment of suspicion, and practice setting characteristics from all participating physicians within a reasonable timeframe; (b) information on the timing of suspicion; and (c) presentation of detection data in ways that elucidate the joint effects of degree and timing of suspicion.

## ACKNOWLEDGMENTS

The authors are grateful to the following individuals for their many and varied roles in making the SIP study work: Debbie Sigal, Lesley Sept, Ph.D., Michelle McCullough, Rahman Azari, Ph.D., Wayne Katon, M.D., Patricia Carney, Ph.D., Edward Callahan, Ph.D., Michael Wilkes, M.D., Ph.D., Fiona Wilson, M.D., Debra Roter, Ph.D., Jeff Rideout, M.D., Robert Bell, Ph.D., Debora

Paterniti, Ph.D., W. Ladson Hinton, M.D., Lisa Meredith, Ph.D., Debra Gage, Mimi Hocking, Alison Venuti, Diane Burgan, Linda Nalbandian, Katherine Li, Vania Manipod, Sheila Krishnan, Henry Young, Ph.D., and Phil Ramonondi, M.D. Special thanks are due to Blue Shield of California, the UCD Primary Care Network, Western Health Advantage (Sacramento), Kaiser Permanente (Sacramento), Brown & Toland (San Francisco), and Excellus Blue Cross (Rochester). We are deeply indebted to the 18 superb actors (SP), and to the participating physicians and their office staffs whose effort, patience, and good humor made this study possible. Supported by a grant (5 R01 MH064683-03) from the National Institute of Mental Health.

## REFERENCES

- Beaulieu, M. D., M. Rivard, E. Hudon, D. Saucier, M. Remondin, and R. Favreau. 2003. "Using Standardized Patients to Measure Professional Performance of Physicians." *International Journal for Quality in Health Care* 15: 251-59.
- Beullens, J., J. J. Rethans, J. Goedhuys, and F. Buntinx. 1997. "The Use of Standardized Patients in Research in General Practice." *Family Practice* 14: 58-62.
- Biernat, K., D. Simpson, E. Duthie, D. Bragg, and R. London. 2003. "Primary Care Residents Self Assessment Skills in Dementia." *Advances in Health Sciences Education* 8: 105-10.
- Brown, J. A., J. Abelson, C. A. Woodward, B. Hutchison, and G. R. Norman. 1998. "Fielding Standardized Patients in Primary Care Settings: Lessons from a Study Using Unannounced Standardized Patients to Assess Preventive Care Practices." *International Journal for Quality in Health Care* 10: 199-26.
- Carney, P. A., A. J. Dietrich, M. S. Eliassen, M. Owen, and L. W. Badger. 1999a. "Recognizing and Managing Depression in Primary Care: A Standardized Patient Study." *Journal of Family Practice* 48: 965-72.
- Carney, P. A., M. S. Eliassen, G. L. Wolford, M. Owen, L. W. Badger, and A. J. Dietrich. 1999b. "How Physician Communication Influences Recognition of Depression in Primary Care." *Journal of Family Practice* 48: 958-64.
- Carney, P. A., and D. H. Ward. 1998. "Using Unannounced Standardized Patients to Assess the HIV Preventive Practices of Family Nurse Practitioners and Family Physicians." *Nurse Practitioner* 23: 56-8.
- Epstein, R. M., P. Franks, C. G. Shields, S. C. Meldrum, K. N. Miller, T. L. Campbell, and K. Fiscella. 2005. "Patient-Centered Communication and Diagnostic Testing." *Annals of Family Medicine* 3: 415-21.
- Epstein, R. M., J. C. Levenkron, L. Frarey, J. Thompson, K. Anderson, and P. Franks. 2001. "Improving Physicians' HIV Risk-Assessment Skills Using Announced and Unannounced Standardized Patients." *Journal of General Internal Medicine* 16: 176-80.

- Gallagher, T. H., B. Lo, M. Chesney, and K. Christensen. 1997. "How Do Physicians Respond to Patients Requests for Costly, Unindicated Services." *Journal of General Internal Medicine* 12: 663–8.
- Glassman, P., J. Luck, E. M. O’Gara, and J. W. Peabody. 2000. "Using Standardized Patients to Measure Quality: Evidence from the Literature and a Prospective Study." *Joint Commission Journal on Quality Improvement* 26: 644–53.
- Gorter, S., J. J. Rethans, D. Van Der Heide, A. Scherpbier, H. Houben, C. Van Der Vleuten, and S. Van Der Linden. 2002. "Reproducibility of Clinical Performance Assessment in Practice Using Incognito-Standardized Patients." *Medical Education* 36: 827–32.
- Grad, R., R. Tamblyn, P. J. McLeod, L. Snell, A. Illescas, and D. Boudreau. 1997. "Does Knowledge of Drug Prescribing Predict Drug Management of Standardized Patients in Office Practice?" *Medical Education* (31): 132–7.
- Hutchison, B., C. A. Woodward, G. R. Norman, J. Abelson, and J. A. Brown. 1998. "Provision of Preventive Care to Unannounced Standardized Patients." *Journal of the Canadian Medical Association* 158: 185–93.
- Kravitz, R. L., R. M. Epstein, M. D. Feldman, C. E. Franz, R. Azari, M. S. Wilkes, L. Hinton, and P. Franks. 2005. "Influence of Patients’ Request for Direct-to-Consumer Advertised Antidepressants: A Randomized-Controlled Trial." *Journal of the American Medical Association* 293: 1995–2002.
- Luck, J., and J. W. Peabody. 2002. "Using Standardized Patients to Measure Physicians’ Practice: Validation Study Using Audio Recordings." *British Medical Journal* 325: 679–83.
- Luck, J., J. W. Peabody, T. R. Dresselhaus, M. Lee, and P. Glassman. 2000. "How Well Does Chart Abstraction Measure Quality? A Prospective Comparison of Standardized Patients with the Medical Record." *American Journal of Medicine* 108: 642–9.
- Maiburg, B. H. J., J. J. Rethans, I. M. Van Erk, L. M. H. Mathus-Vliegen, and J. W. Van Ree. 2004. "Fielding Incognito Standardized Patients as ‘Known’ Patients in a Controlled Trial in General Practice." *Medical Education* 38: 1229–35.
- McCulloch, C. E., and S. R. Searle. 2001. *Generalized, Linear, and Mixed Models*. New York: Wiley.
- McLeod, P. J., R. M. Tamblyn, D. Gayton, R. Grad, L. Snell, L. Berkson, and M. Abrahamowicz. 1997. "Use of Standardized Patients to Assess between-Physician Variations in Resource Utilization." *Journal of the American Medical Association* 278: 1164–8.
- Peabody, J. W., J. Luck, P. Glassman, T. R. Dresselhaus, and M. Lee. 2000. "Comparison of Vignettes, Standardized Patients, and Chart Abstraction." *Journal of the American Medical Association* 283: 1715–22.
- Rethans, J. J., R. Drop, F. Sturmans, and C. Van Der Vleuten. 1991. "A Method for Introducing Standardized (Simulated) Patients into General Practice Consultations." *British Journal of General Practice* 41: 94–6.
- Rosen, J., B. Mulsant, M. L. Bruce, V. Mittal, and D. Fox. 2004. "Actors’ Portrayals of Depression to Test Interrater Reliability in Clinical Trails." *American Journal of Psychiatry* 161: 909–1911.

- Tamblyn, R. M. 1998. "Use of Standardized Patients in the Assessment of Medical Practice." *Canadian Medical Association Journal* 158: 205–07.
- Tamblyn, R. M., M. Abrahamowicz, L. Berkson, W. D. Dauphinee, D. C. Gayton, R. M. Grad, L. M. Isaac, M. Marrache, P. J. McLeod, and L. S. Snell. 1992. "First-Visit Bias in the Measurement of Clinical Competence with Standardized Patients." *Academic Medicine* 67: S22–24.
- Woodward, C. A., B. Hutchison, G. R. Norman, J. A. Brown, and J. Abelson. 1998. "What Factors Influence Primary Care Physicians' Charges for their Services?" *Journal of the Canadian Medical Association* 158: 197–202.

## SUPPLEMENTARY MATERIAL

The following supplementary material for this article is available online:

APPENDIX. Review of prevalence rates and methods for assessing detection in unannounced standardized patient research since Beullens' et al. (1997) review.