Virtual standardized patients: an interactive method to examine variation in depression care among primary care physicians

Lisa M. Hooper[1], **Kevin P. Weinfurt**[3], **Lisa A. Cooper**[4], **Julie Mensh**[5], **William Harless**[6], **Melissa C. Kuhajda**[2], and **Steven A. Epstein**[5]

[1]Department of Educational Studies in Psychology, Research Methodology, & Counseling, The University of Alabama, Tuscaloosa, AL, USA

[2]Department of Psychiatry and Behavioral Medicine; The University of Alabama School of Medicine, Tuscaloosa, AL, USA

[3]Department of Psychiatry and Behavioral Sciences and Department of Psychology and Neuroscience, Duke Clinical Research Institute, Durham, NC, USA

[4]Welch Center for Prevention, Epidemiology, and Clinical Research, Johns Hopkins University, Baltimore, MD, USA

[5]Department of Psychiatry, Georgetown University School of Medicine, Washington DC, USA

[6]Interactive Drama Inc., Bethesda, MD, USA

Abstract

Background—Some primary care physicians provide less than optimal care for depression (Kessler *et al., Journal of the American Medical Association* 291, 2581–90, 2004). However, the literature is not unanimous on the best method to use in order to investigate this variation in care. To capture variations in physician behaviour and decision making in primary care settings, 32 interactive CD-ROM vignettes were constructed and tested.

Aim and method—The primary aim of this methods-focused paper was to review the extent to which our study method – an interactive CD-ROM patient vignette methodology – was effective in capturing variation in physician behaviour. Specifically, we examined the following questions: (a) Did the interactive CD-ROM technology work? (b) Did we create believable virtual patients? (c) Did the research protocol enable interviews (data collection) to be completed as planned? (d) To what extent was the targeted study sample size achieved? and (e) Did the study interview protocol generate valid and reliable quantitative data and rich, credible qualitative data?

Findings—Among a sample of 404 randomly selected primary care physicians, our voice-activated interactive methodology appeared to be effective. Specifically, our methodology – combining interactive virtual patient vignette technology, experimental design, and expansive open-ended interview protocol – generated valid explanations for variations in primary care physician practice patterns related to depression care.

Keywords

depression care; interactive vignette methodology; physician decision making; primary care; virtual patients

Correspondence to: Lisa M. Hooper, Department of Educational Studies in Psychology, Research Methodology, & Counseling, The University of Alabama, Box 870231, 315 B Graves Hall Tuscaloosa, Alabama 35487, USA. lhooper@bamaed.ua.edu.

Introduction

Some primary care physicians provide less than optimal care for depression (Kessler *et al.*, 2004), yet depression is one of the most common disorders encountered in the primary care setting (Saver *et al.*, 2007). It is estimated that between 20% and 40% of individuals who have had a myocardial infarction during the past six months will have comorbid depression (Goldman *et al.*, 1999). Any effort towards reducing the barriers to effective depression treatment in general medical settings would have far-reaching positive effects for patients, providers, and the health care system at large (Rost *et al.*, 2005). To this end, our survey study was undertaken to determine the extent to which patient and physician factors are associated with physician care and decision making for depression among the medically ill. Because of the cost, time, and needed sample size to carry out an experimental research design study, we were also interested in testing an interactive method to measure physician practice patterns.

The current paper considers the effectiveness of a new interactive methodology used to capture physician practice behaviour. The primary objective of this paper is to add to the literature base by providing a detailed description of our interactive methodology and an assessment of its effectiveness. Thus, we intend for this report to facilitate the use and duplication of the described methodology in future primary care research.

Background

Measurement of physician behaviour and practice patterns

Researchers agree that many valid and reliable methods exist for assessing physician practice in the context of medical university settings (Kirwan *et al.*, 1983; Rethans and Saebu, 1997; Glassman *et al.*, 2000; Luck and Peabody, 2002). However, researchers debate the validity and reliability of using these same methods in actual practice settings (Colliver and Schwartz, 1997; Feldman *et al.*, 1997; Peabody *et al.*, 2000; 2004; Kravitz *et al.*, 2006). Methods used in the past include standardized patients (Colliver *et al.*, 1993; Badger *et al.*, 1994; Colliver and Schwartz, 1997; Kravitz *et al.*, 2006; Epstein *et al.*, in press), chart reviews (Gilbert *et al.*, 1996; McDonald *et al.*, 1997; Tamblyn, 1998; Dresselhaus *et al.*, 2002), pharmaceutical company databases, paper-and-pencil vignettes (Colliver *et al.*, 1993; McKinlay *et al.*, 1997; Tamblyn, 1998; Epstein *et al.*, 2001), and video clinical vignettes (Feldman *et al.*, 1997; McKinlay *et al.*, 1997; Schulman *et al.*, 1999).

In a prospective validation study, Peabody and colleagues examined how accurately clinical vignettes measure physician practice and decision-making patterns compared with medical chart review and with standardized patients, suggesting that the latter two methods are the 'gold standard' (Peabody *et al.*, 2004). Among a sample of primary care physicians ($n = 116$), the quality of physician clinical practice was found to be 73% correct when measured by standardized patients, 68% correct by vignettes, and 63% correct by medical chart review. Peabody and colleagues concluded that clinical vignettes were a valid method to assess quality of clinical care irrespective of symptomatology, severity, or physician medical school training.

McKinlay and Feldman's video methodology (Feldman *et al.*, 1997; McKinlay *et al.*, 1996; Norcini, 2004) has been established as a reliable and valid method of assessing the influence of patient and provider characteristics on physicians' decision-making and practice patterns. McKinlay *et al.* (1996) assert that their video vignette methodology has provided an excellent combination of realism, feasibility, and financial expedience.

Using the video vignette methodology, Schulman and colleagues tested how 720 physicians each interacted with one randomly selected video vignette to study the influence that race and gender may have on primary care physician recommendations for chest pain management

(Schulman *et al.*, 1999). The use of video vignettes allowed the researchers to hold constant patient symptomatology, language, emotion, and appearance. Video patient's race and gender were found to have significantly affected physician treatment and management of chest pain. For example, African-Americans and women presenting with chest pain were less likely to be referred for diagnostic tests than were White Americans and men. Schulman and colleagues concluded that, given the internal validity and level of control evinced in the design of the study, they were able to draw important conclusions about their findings, compared to other studies that had not used an experimental study design.

In another study, Dresselhaus and colleagues concluded that studies using medical record review to assess physician behaviour and practice patterns overstate patient care (ie, 'overestimate the quality of care' provided by physicians (Dresselhaus *et al.*, 2002: 291)). They found false positives 6.4% of the time; actual medical chart reports were a less-accurate or valid representation of what physicians do in their day-to-day practice than structured reports by standardized patients.

Limitations in methods to measure physician practice patterns in prior studies

Methods used in studies examining physician practice and decision-making patterns are not without limitations. Many of these methods, particularly medical chart abstraction, are limited by incomplete data, potential recording bias, and inaccuracies (Dresselhaus *et al.*, 2002). These methods may also be limited by a lack of documentation regarding the reasoning and rationale for diagnosis, assessment, and treatment decisions.

Research studies using live standardized patients are traditionally very costly and afford less control over presenting symptomatology than standardized vignettes. Additionally, and related to cost, live patients may represent a narrow spectrum of primary care practice (Badger *et al.*, 1994). Nor do they allow for a semi-structured interview assessing physician reasoning and rationale for diagnosis, assessment, and treatment decisions. On the other hand, Norcini (2004) contends that performance on clinical vignettes may be skewed by physicians' tendency to respond in an 'ideal fashion,' unlike how they would during a typical hectic day in the office.

These limitations aside, many researchers continue to uphold 'live standardized patients' as a valid method for investigating physician behaviour and practice patterns. Other researchers contend that clinical vignettes are a more valid and reliable method of capturing physician practice (Peabody *et al.*, 2004). The purpose of this paper's method is to provide a detailed description of a new voice-activated interactive methodology and to examine to what extent it was effective in capturing physicians' described behaviour and decision making.

Current study: a new interactive methodology

Our study, Physicians Decisions for the Depressed Medically Ill (PD Study), is among the few to use virtual patients in computerized interactive clinical vignettes (McKinlay *et al.*, 1997; Schulman *et al.*, 1999; Epstein *et al.*, in press). The methodology in the current study is undergirded by Feldman and McKinlay's (Feldman *et al.*, 1997; McKinlay *et al.*, 1998; 2002) video methodology. However, our interactive CD-ROM patient vignette methodology is different from their methodology in that we used voice-activated interactive virtual patients with whom the physicians in this study interacted. As with paper-and-pencil and video vignettes, our physician-participants were exposed to identical patient data on which diagnostic, assessment, and treatment decisions were based. Consequently, responses are comparable and specifically related to the physician and not to some unknown, unmeasured variable.

The following section describes the method employed in our study. Importantly, we also delineate specific details related to the development of each part of the interactive methodology so that it may be duplicated in future studies.

Development of CD-ROM interactive clinical vignettes

Thirty-two unique virtual programs were created to provide a venue for primary care physicians to engage in voice-activated interactive conversations with virtual patients presenting with clear depression symptomatology. Although the actors in these 32 scenarios did not differ in their presentation of depression, they did differ in race, gender, comorbid medical illness, treatment preference, and attribution style. These five factors were dichotomized and stratified into 32 distinct factor combinations ($2^5 = 32$; see Table 1) in conjunction with the patients' presentation of clearly recognizable depression.

The research team employed a rigorous set of procedures creating authentic and medically accurate scripts for actors who were hired from the Washington, DC, metropolitan area to enact the role of a 55-year-old patient (African-American or White male or female (see Figure 1)) with a clear presentation of moderate depression in the context of a routine primary care office visit. Over a period of six months, individual interviews were conducted with primary care patients who had suffered depression and a myocardial infarction separately and concomitantly. Focus groups and interviews were also conducted with physicians to ensure the representativeness of the virtual dialogue on which the CD-ROM vignettes were based. Additionally, audiotaped conversations from actual physician–patient encounters were used to create preliminary scripts (Cooper *et al.*, 2003). Scripts were reviewed and edited by a panel comprising internationally recognized experts from survey and primary care research, psychiatry, cardiology, and mental health.

Interactive dialogue—Harless and his staff from Interactive Drama Incorporated (IDI) developed the interactive technology, determining what modifications of the original scripts would create the best virtual conversations. The research team and Harless and his staff collaborated on the final scripts used by the actors (see Table 2). Once the interactive software was complete, the research team conducted a two-month pilot test with volunteer physician-participants to assess the usability of the technology.

The interactive technology, consisting of a special virtual dialogue technology program, a microphone, speakers, and a laptop, had pre-scripted questions – which appeared on the bottom of the laptop screen below the virtual patient – that each physician 'asked' of the virtual patient. Thus, the 7-min interactive dialogue was an exchange between the physician-participant and the virtual patient. All physicians activated the dialogue by clicking on the microphone and reading aloud a common question that likely would be asked of most patients coming into their office. The virtual patient's response was consistent with the factors under study (see Table 2). Talking into the microphone, the physician would ask the identical pre-scripted questions. An example of the voice-activated interactive sequence is as follows:

Physician: 'What can I do for your today?'

Patient: 'Well I haven't been feeling like myself lately. I'm really tired all the time and I thought I should see a doctor. I just don't have any energy.'

Physician: 'How long has this been going on?'

Patient: 'I started to feel this way three months ago, after my heart attack. I am okay now, but my recovery has been very slow. Things have been difficult for me recently. The doctor said I would be fine but, I'm not back to normal yet.'

The resultant 7-min dialogues were taped and transferred to CD-ROM to be used during the study interview with physician-participants.

Development of the research interview protocol

Interviewers—Two clinical research team members, trained during the piloting phase, conducted the semi-structured interviews after the physician interacted with the virtual patient. Approximately half of the randomly selected physicians were assigned to each. Inter-rater reliability was established through blind coding of each other's audiotaped interviews. Supervision and a review of all surveys were completed throughout the study.

Semi-structured interview—Although a semi-structured interview was used, the researchers closely followed a manualized interview guide standardizing the process for all participants. All physicians were asked to respond to the interview questions as if the patients were actually in their office. The interview sequence began with a brief scripted introduction of the study and the informed consent. The interview included questions about assessment (ie, based on the information presented in the case vignette, what additional information would be essential for you to obtain about this patient's clinical condition?), diagnosis (ie, other than the diagnoses indicated in the medical chart, what are the diagnoses that you are considering for this patient?), and treatment (ie, what are your leading treatment recommendations for this patient?). The interviewer also asked physicians their reasoning behind their decision making related to assessment, diagnosis, and treatment recommendations (ie, please explain the reasoning behind your treatment plan).

Research design

Our research study employed a 2×2×2×2×2 factorial design (see Table 2) examining combinations of five (two-level) experimental patient factors: race (African-American or White), gender (female or male), comorbid medical illness (healthy or post-myocardial infarction), treatment preference (accepting or reluctant), and attribution style (psychological or somatic). Each primary care physician-participant was randomly assigned one CD-ROM case vignette of a virtual patient with explicit depression.

Participant recruitment

The randomly selected nationally representative sample of physicians was derived from a database provided by the American Medical Association. Physicians listed in the Primary Metropolitan Statistical Areas of Washington, DC, and Baltimore, Maryland, were stratified by ethnicity (African-American or non-African-American) and practice type (family physician or internist) for the purpose of planned statistical analyses. Following the initial randomization process, physicians were selected in blocks of 48.

Physician recruitment was achieved in several steps (see Figure 2). First, participation invitations were mailed and included a brief study description, explanation of participant duties, participation risks and benefits, and a postcard refusal option. After 12 days elapsed (adequate time for investigators to receive postcard refusals), physicians were telephoned to determine eligibility, willingness to participate, and availability for the interview. On average, six calls were needed to secure a time to interview the physicians. One strategy employed in the PD Study, differing slightly from similar studies (Main *et al.*, 1993;Feldman *et al.*, 1997), was to speak directly with the physician upon initial contact. Researchers avoided speaking with or scheduling the interview through a gatekeeper, a method that decreased the number of calls required to make physician contact.

Physicians, once reached and agreeing to participate, were informed that their responses would be kept confidential, that only group data would be reported, that interviews would be conducted at convenient times in their offices, that this study has no affiliation with a pharmaceutical company, and that physician time would be compensated with $125. Once the interview was scheduled, the research team interviewer received one randomly selected virtual case vignette out of a possible 32.

Final study sample

Participants were 404 primary care physicians. Specifically, we sent mailings to 942 physicians inviting them to participate in the study. Thirty-six physicians returned the postcards, indicating a refusal to participate. Of the 906 remaining physicians, 418 (44%) were eligible and agreed to participate, 340 (36%) were ineligible, and 184 refused to participate (20%, ie, 36, refused by postcard and 148 refused by phone). Of the 418 physicians who were interviewed, 14 were later removed from the data pool (ie, their interview records were discarded) because of large amounts of missing data or significant technology problems.

Participants ranged in age from 29 to 88 years, with the total study sample's mean age being 47.66 (SD = 10.15) years. Race and ethnicity were diverse, with participants reporting non-Hispanic White (48%, $n = 194$), non-Hispanic Black (33%, $n = 133$), Asian-American (12%, $n = 51$), or other race/ethnicity (7%, $n = 28$) as their primary racial/ethnic identification.

Procedures

All study procedures were approved by the Georgetown University Institutional Review Board. The research study protocol (see Figure 2) included the following: (1) the informed consent form and an introduction to the study; (2) the approximately 7-min interaction dialogue with the physician-participant and virtual patient on the study laptop; (3) the semi-structured interview; (4) a paper-and-pencil survey about the physician-participant demographics (gender, birth date, race, and specialty); knowledge, attitudes, and beliefs regarding depression care; and opinions about the believability of the virtual patient and the similarity of the virtual patient to the patients whom the physician sees in his or her everyday practice.

Instrumentation

Demographic questionnaire—The instrument created for the study asks participants to describe themselves (gender, birth date, race, and specialty), their practice patterns (board certification, administrative duties), and their attitudes and knowledge about depression care. Physicians were also asked about the believability of the virtual patient, as well as 'how real the virtual patient' seemed to them.

Semi-structured interview—A manualized interview guide was used to standardize the protocol process for all participants. All physicians were asked to respond to the interview questions as if the patients were actually in their office. After the physician interacted with the virtual patient, the interview was conducted, which included questions about assessment (based on the information presented in the case vignette, what additional information would be essential for you to obtain about this patient's clinical condition?), diagnosis (other than the diagnoses indicated in the medical chart, what are the diagnoses that you are considering for this patient?), and treatment (what are your leading treatment recommendations for this patient?). We also asked physicians their reasoning behind their decision making related to assessment, diagnosis, and treatment recommendations (eg, please explain the reasoning behind your leading diagnoses).

Results

To determine the meaningfulness of our study outcomes, we examined the following questions:

1. Did the CD-ROM technology work?

2. Did we create believable virtual patients?

3. Did the research protocol enable interviews (data collection) to be completed as planned?

4. To what extent was the targeted study sample size achieved?

5. Did the study interview protocol generate valid and reliable quantitative data and rich, credible qualitative data?

The following sections discuss our findings for these questions.

Effectiveness of the CD-ROM technology

The current study was one of few studies to use interactive software to better understand physician practice patterns. The development and piloting of the technology took over 12 months to ensure that all parts were operationable and believable. However, it remained unclear whether the technology would be effective once the study moved into the active phase. We measured effectiveness of the CD-ROM technology (ie, special virtual dialogue technology program, a microphone, speakers, and a laptop) by the number of interviews that had to be cancelled or discarded because of faulty technology or physician resistance to using the interactive software. Our study found that 98% ($n = 399$) of the participants used the software successfully, with only 2% ($n = 8$) of interviews having complications. No physicians refused to 'interact' with the virtual patient on the laptop.

Although we did not track the number of unsolicited qualitative responses regarding the assessment method and interactive technology, it was overwhelmingly positive. Many physicians reported that they found the technology to be a fun and interesting way to get to know the virtual patients and their presenting problems. A very few physicians verbalized that they found the technology to be unnecessary and would have preferred a more traditional paper case vignette.

Believableness and representativeness of the study's virtual patients

In order to make meaning of and to determine the usefulness of the study's findings, we assessed how 'believable' or 'real' the virtual patient was and how 'similar' the virtual patient was to patients whom the physician-participants see in their everyday practice. With regard to the first question, physicians were asked if the virtual patient seemed *real* to them. On a Likert scale of 1 (strongly disagree) to 5 (strongly agree), 90% of the physicians either agreed (37%) or strongly agreed (53%), 7% were neutral, and the remaining 3% of the physicians either disagreed (2%) or strongly disagreed (1%). When asked to rate how *similar* the virtual patient was to patients seen in their practice, on a Likert scale of 1 (not at all similar) to 5 (very similar), over 85% of the physicians indicated that the virtual patient was 'very similar' (42%) or 'similar' (43%), 12% were neutral, and the remaining 3% of the physicians suggested that the patient was dissimilar (2.5%) or not at all similar (0.5%) to the patients who they see in their office most days.

Effectiveness of overall study research protocol

We recognized that conducting research with physician-participants can be challenging because of their reported work schedules. Thus, the research protocol was conducted in the physicians' offices in the context of their regular workdays. To determine whether the overall

study research protocol was effective, we considered the following points of data: (1) the number of incomplete study protocols (ie, the physician started the interview but practice was too busy to allow for the completion of the interview); (2) the degree to which physicians refused to participate *after* agreeing to be a part of the study; (3) the number of on-site cancellations; and (4) the average length of time to complete the protocol. Study findings suggest that the research protocol worked well: there were no incomplete interviews; all physicians with the exception of one who originally agreed to be a part of the study went on to complete the protocol; very few interviews were cancelled (2%; $n = 9$); and only a small number lasted beyond the 1-h reserved time slot.

Effectiveness of recruitment strategy in meeting sample size goal

To determine whether the recruitment strategy was successful, we considered the extent to which we were able to reach our target study sample size of $n = 500$. Additionally, we considered the degree to which we were able to gain access to and schedule the physicians to be a part of the study.

For this study, 418 physician-participants (67% participation rate) were recruited over an 18-month period. Although less than our planned sample, the final study sample still provided greater than 80% power to detect differences between predictor and outcome variables.

Compared with the few other studies that have employed this methodology, our recruitment strategy was successful (Peabody *et al.*, 2004). One of the major study obstacles of surveying primary care physicians is gaining direct access to the physician to discuss physician willingness to participate (J. McKinlay, personal communication, February 2002). In this study, six calls (mean = 5.84; SD = 3.90) were needed to speak with and schedule an interview; our study's results (ie, six calls) are consistent with or slightly better than other studies (Main *et al.*, 1993).

Quantitative and qualitative data

This was a factorial study, with *a priori* hypotheses examining causative factors related to depression care in a primary care setting. The overall planned analyses were quantitative in nature. We also collected rich qualitative data to examine via the audiotaped physician interviews. The success of our research team interviewers, research interview protocol, and data management methods resulted in very little missing data.

Quantitative data and findings—A strength of this study is the experimental research design, enabling more precise measurement of physician and patient characteristics that may influence diagnosis, assessment, and treatment of depression. This study allowed us to examine the main effects of provider factors (eg, ethnicity and practice type) averaged over all combinations of the experimentally manipulated patient factors (eg, race, gender, comorbid medical illness, treatment preference, and attribution style). This section describes the primary care physician's response to the interview questions related to assessment, diagnosis, and the treatment of a patient with depression. The major results of this study can be found elsewhere and thus are only briefly described here (Epstein *et al.*, in press).

Almost all study physician-participants put forward a preliminary diagnosis of a depressive disorder. In fact, physicians indicated depression as their leading diagnosis 97% ($n = 395$) of the time. Other diagnoses given were hypothyroidism, adjustment disorder, coronary artery disease, anaemia, and anxiety disorder.

Physicians varied regarding the additional testing and assessments they identified as essential prior to treatment. For example, 37% wanted to ask the patient about prior depressive episodes,

2% wanted information about a history of bipolar disorder; 54% wanted to order a thyroid study, and 36% would have asked about current or previous suicidal thoughts and behaviours.

Forty-seven per cent of the physician-participants indicated that they would refer the patient to a mental health provider, 22% suggested that they would do counselling themselves, and 85% recommended an antidepressant as their first-line treatment of depression. Importantly, some physicians emphasized the significance of communicating educational messages about the medicine in addition to writing the prescriptions. Examples include adverse effects of the medication (eg, gastrointestinal side effects, sleep disturbances, sexual side effects), directions to take the medication (eg, take medication every day, take the medication for six months even if feeling better), and limits of the medication treatment (eg, therapeutic delay).

Qualitative data—Extensive supplemental qualitative data were collected in addition to the planned quantitative data. Qualitative data, generated from open-ended questions directed towards physician decision-making approaches on depression care, have yet to be examined, although we believe they will facilitate our ability to contextualize our quantitative findings and assist in data interpretation. An example of the rich qualitative data derived from our methodology is evidenced in the brief extract, which illustrates the primary care physician's response to the exchange with the interactive patient following the patient's stated reluctance to receive mental health treatment. The physician describes what he would say to the patient if the patient were in his office.

> I see this quite commonly in my patients and what I normally do is explain to them that there is nothing to be embarrassed or ashamed about if they have depression. I usually go through a scale and try to talk to them about where people fall on the serotonin scale. People with very high levels of serotonin are less likely to develop significant depression; people with low levels of serotonin are often depressed or anxious all the time without a social stressor. And folks with medium to low levels of serotonin can do just fine, but when there is a psychosocial stressor in their lives they can be thrown easily into depression or anxiety. And that this is a chemical thing that we are able to treat with medication… I would push her very strongly to take the medication.

Study limitations

This study has several limitations. While other studies have established that physician-participants often respond to clinical vignettes as they would to an actual patient (McKinlay *et al.*, 1997; Peabody *et al.*, 2004; Epstein *et al.*, in press), the reliance on virtual patients as the method of measurement is nonetheless a limitation of this study. Because the depression vignette was purposefully straightforward and explicit, it may in fact be less complicated than patients who present with depression in the physician-participants' everyday practice. A related potential limitation is the semi-structured interview, which may have yielded responses unrepresentative of actual practice patterns. Finally, a limitation of the study is the participation rate of only 67%. Although that rate is good compared with other survey research, it still limits the generalizability of the findings.

Implications and future directions in primary care research

Building on and extending the scant literature to date (McDonald *et al.*, 1997; Tamblyn, 1998; Epstein *et al.*, in press), our study methodology generated preliminary and valuable explanations for the varied practice patterns related to depression care among primary care physicians. We believe our ability to capture and uncover these variations is a direct reflection of our interactive research method involving the use of an experimental factorial design, virtual

standardized patients presented via interactive computer technology, and an expansive open-ended interview protocol. Towards this end, we spent a significant amount of time (approximately one year) creating believable virtual case vignettes derived from real-life physician–patient encounters, focus groups comprising physicians and patients, and individual interviews. We employed the same rigorous method to create the on-site research interview protocol that elucidated physicians' approach to depression care and their decision making that underlies this care. Finally, we used specific recruitment techniques that served to enhance our study participation. For example, detailed participation invitations were mailed describing the study, explaining participant duties and risks and benefits to participation, and providing a postcard refusal option. The unique recruitment enhancing feature, differing slightly from similar studies (Main *et al.*, 1993; McKinlay *et al.*, 1997), was our insistence upon speaking directly with the physician upon initial contact. Researchers avoided speaking with or scheduling the interview through a gatekeeper, decreasing the number of calls required to make physician contact.

We have described the comprehensive work carried out in the PD Study to create a reliable, believable, and interactive method to measure depression care among the physician-participants. Our study was undertaken to better understand the extent to which patient factors inform primary care physician diagnosis and treatment of depression separately and in conjunction with a serious medical illness (myocardial infarction). Additionally, this study was carried out to understand the extent to which physician characteristics further explain the variation in depression care among primary care physicians. Knowledge gleaned from this survey study (specific and comprehensive results reported elsewhere; see Epstein *et al.*, in press) will improve the measures currently used in the assessment of quality of care among primary care physicians treating patients with depression and comorbid medical conditions.

Our interactive research method, given its potential for improved cost-effectiveness, utility in collecting a comprehensive range of unconfounded data, and valuable physician responses, taken together with studies by Feldman, McKinlay, Schulman, and others (Feldman *et al.*, 1997; McKinlay *et al.*, 1998; 2006; Meredith *et al.*, 1999; Schulman, 1999; Veloski *et al.*, 2005; Srinivasan *et al.*, 2006), add considerably to the converging body of evidence supporting this rigorous research method. Researchers interested in assessing physician practice and decision-making variation may wish to consider this methodology.

Acknowledgments

References

Badger LW, deGruy FV, Hartman J, Plant MA, Leeper J, Anderson R, et al. Patient presentation, interview content, and the detection of depression by primary care physicians. Psychosomatic Medicine 1994;56:128–35. [PubMed: 8008799]

Colliver JA, Schwartz MH. Assessing clinical performance with standardized patients. Journal of the American Medical Association 1997;278:790–91. [PubMed: 9286845]

Colliver JA, Vu NV, Marcy ML, Travis TA, Robbs RS. Effects of examinee gender, standardized-patient gender, and their interaction on standardized patients' ratings of examinees interpersonal and communication skills. Academic Medicine 1993;68:153–57. [PubMed: 8431238]

Cooper LA, Gonzales JJ, Gallo JJ, Rost KM, Meredith LS, Rubenstein LV, Wang N, Ford DE. The acceptability of treatment for depression among African-American, Hispanic, and White primary care patients. Medical Care 2003;41:479–89. [PubMed: 12665712]

Dresselhaus TR, Luck J, Peabody JW. The ethical problem of false positives: a prospective evaluation of physician reporting in the medical record. Journal of Medical Ethics 2002;28:291–94. [PubMed: 12356955]

Epstein SA, Gonzales JJ, Weinfurt K, Boekeloo B, Yuan N, Chase G. Are psychiatrists' characteristics related to how they care for depression in the medically ill? Results from a national case-vignette survey. Psychosomatics 2001;42:482–90. [PubMed: 11815683]

Epstein SA, Hooper LM, Weinfurt KP, DePuy V, Cooper LA, Mensh JA. Primary care physicians' evaluation and treatment of depression: results of an experimental study using video vignettes. Medical Care Research and Review. 2008

Feldman HA, McKinlay JB, Potter DA, Freund KM, Burns RB, Moskowitz MA, et al. Nonmedical influences on medical decision making: an experimental technique using videotapes, factorial design, and survey sampling. Health Services Research 1997;32:343–65. [PubMed: 9240285]

Gilbert EH, Lowenstein SR, Koziol-McLain J, Barta DC, Steiner J. Chart reviews in emergency medicine: where are the methods? Annals of Emergency Medicine 1996;27:302–308.

Glassman PA, Luck J, O'Gara EM, Peabody JW. Using standardized patients to measure quality: evidence from the literature and a prospective study. Joint Commission Journal on Quality Care Improvement 2000;26:644–53.

Goldman LS, Nielsen NH, Champion HC. Awareness, diagnosis, and treatment of depression. Journal of General Internal Medicine 1999;14:569–80. [PubMed: 10491249]

Kessler RC, et al. Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization world mental health surveys. Journal of American Medical Association 2004;291:2581–90.

Kirwan JR, Chaput de Saintonge DM, Joyce CR, Currey HL. Clinical judgment in Rheumatoid arthritis: I. Rheumatologists' opinions and the development of "paper patients". Annals of Rheumatoid Disorders 1983;42:644–47.

Kravitz RL, Franks P, Feldman M, Meredith LS, Hinton L, Franz C, Duberstein P, Epstein RM. What drives referral from primary care physicians to mental health specialists? A randomized trial using actors portraying depressive symptoms. Journal of General Internal Medicine 2006;21:584–89. [PubMed: 16808740]

Luck J, Peabody JW. Using standardized patients to measure physicians' practice: validation study using audio recordings. British Medical Journal 2002;325:809–12.

Main DS, Lutz LJ, Barrett JE, Matthew J, Miller RS. The role of primary care clinician attitudes, beliefs, and training in the diagnosis and treatment of depression. Archives of Family Medicine 1993;2:1061–66. [PubMed: 8111487]

McDonald CJ, Overhage JM, Dexter P, Takesue BY, Dwyer DM. A framework for capturing clinical data sets from computerized sources. Annals of Internal Medicine 1997;141:813–815.

McKinlay JB, Potter DA, Feldman HA. Non-medical influences on medical decision-making. Social Science and Medicine 1996;42:769–776. [PubMed: 8685745]

McKinlay JB, Burns RB, Durante R, Feldman HA, Freund KM, Harow BS, Irish JT, Kasten LE, Moskowitz MA. Patient, physician, and presentational influences on clinical decision making for breast cancer: results from a factorial experiment. Journal of Evaluation in Clinical Practice 1997;3:23–57. [PubMed: 9238607]

McKinlay JB, Burns RB, Feldman HA, Freund KM, Irish JT, Kasten LE, et al. Physician variability and uncertainty in the management of breast cancer: results from a factorial experiment. Medical Care 1998;36:385–96. [PubMed: 9520962]

McKinlay JB, Lin T, Freund K, Moskowitz M. The unexpected influence of physician attributes on clinical decisions: result of an experiment. Journal of Health Social Behavior 2002;43:92–106.

McKinlay J, Link C, Arber S, Marceau L, O'Donnell A, Adams A. How do doctors in different countries manage the same patient? Results of a factorial experiment. Health Services Research 2006;41:2182–200. [PubMed: 17116115]

Meredith LS, Rubenstein LV, Rost K, Ford DE, Gordon N, Nutting P, et al. Treating depression in staff-model versus network-model managed care organizations. Journal of General Internal Medicine 1999;14:39–48. [PubMed: 9893090]

Norcini J. Back to the future: clinical vignettes and the measurement of physician performance. Annals of Internal Medicine 2004;141:813–15. [PubMed: 15545682]

Peabody J, Luck J, Glassman P, Dresselhaus TR, Lee M. A prospective validation study of three methods for measuring quality. Journal of American Medical Association 2000;283:1715–22.

Peabody J, Luck J, Glassman P, Jain S, Hansen J, Spell M, Lee M. Measuring the quality of physician practice by using clinical vignettes: a prospective validation study. Annals of Internal Medicine 2004;141:771–80. [PubMed: 15545677]

Rethans JJ, Saebu L. Do general practitioners act consistently in real practice when they meet the same patient twice? Examination of intradoctor variation using standardized (simulated) patients. British Medical Journal (international edition) 1997;314:1170–74.

Rost K, Pyne JM, Dickinson LM, LoSasso AT. Cost-effectiveness of enhancing primary care depression management on an ongoing basis. Annals of Family Medicine 2005;3:7–14. [PubMed: 15671185]

Saver BG, Van-Nguyen V, Keppel G, Doescher MP. A qualitative study of depression in primary care: missed opportunities for diagnosis and education. Journal of American Board of Family Medicine 2007;20:28–35.

Schulman KA, Berlin JA, Harless W, Kerner JF, Sistrunk S, Gersh BJ, et al. The effect of race and sex on physicians' recommendations for cardiac catheterization. New England Journal of Medicine 1999;340:618–26. [PubMed: 10029647]

Srinivasan M, Hwang JC, West D, Yellowlees PM. Assessment of clinical skills using simulator technologies. Academic Psychiatry 2006;30:505–15. [PubMed: 17139022]

Tamblyn RM. Use of standardized patients in the assessment of medical practice. Canadian Medical Association Journal 1998;158:205–07. [PubMed: 9469142]

Veloski J, Tai S, Evans AS, Nash DB. Clinical vignette-based surveys: a tool for assessing physician variation. American Journal of Medical Quality 2005;20:151–57. [PubMed: 15951521]

Figure 1.
Virtual patients (all photographs are of actors who have given their permission for them to be used)
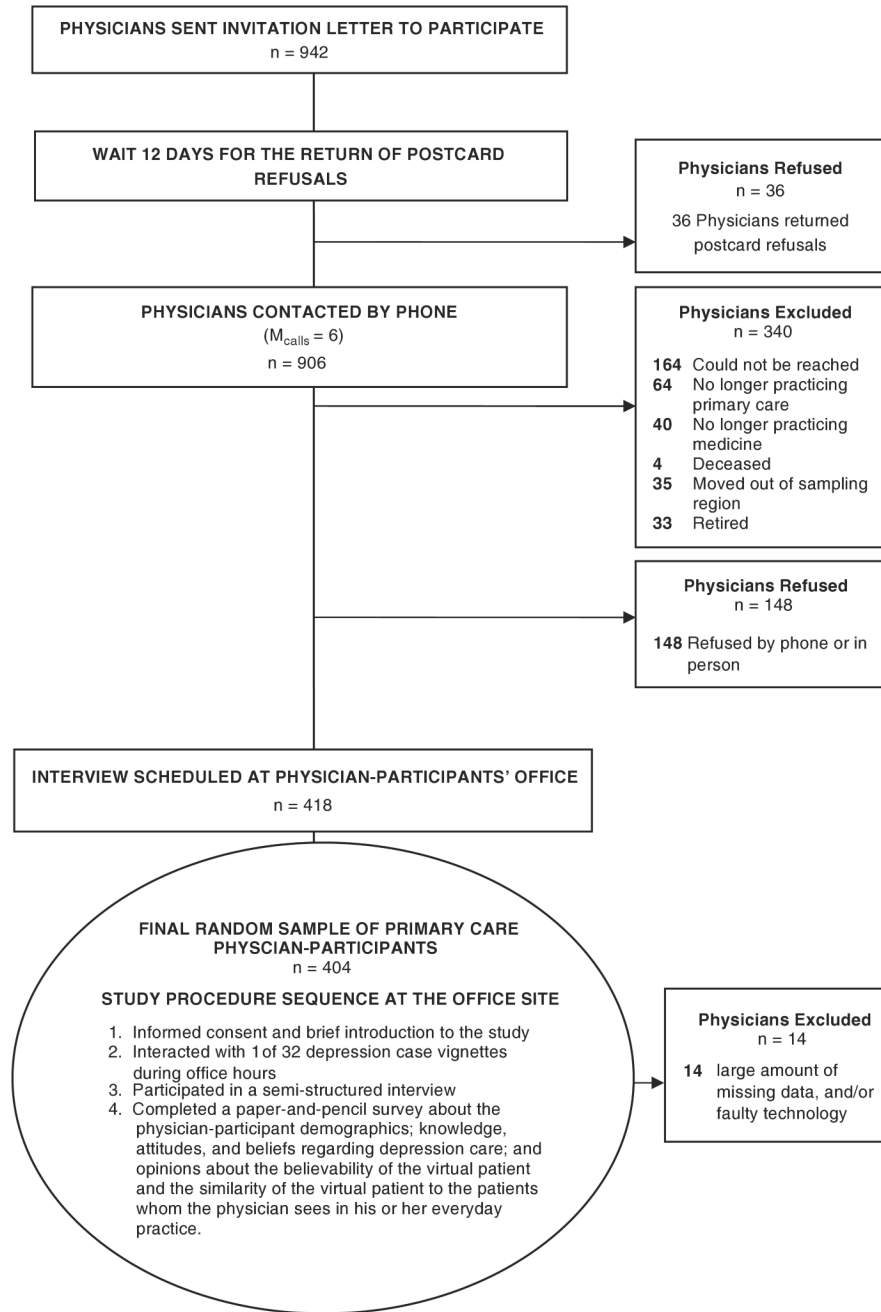
Figure 2.
Physician-participant recruitment flow and procedure

Table 1

Patient vignette variables and levels

Variable 1: medical illness comorbidity/stressor	Variable 2: attributional style	Variable 3: attitude towards mental health treatment	Variable 4: gender	Variable 5: race
(1) Myocardial infarction	(1) Somatic attributional style for depression symptoms	(1) Accepting	(1) Female	(1) African-American
(2) Good health and psychosocial stressor	(2) Psychological attributional style for depression symptoms	(2) Reluctant	(2) Male	(2) White

Table 2

Depression language used in 32 virtual clinical case vignettes

Depression symptomatology	Patient's script
(1) Anhedonia	'I used to walk around my neighborhood every night, but lately I don't have the energy to do that. My husband and I used to enjoy going out to dinner, but I haven't been interested in eating, so we haven't even being doing that. I'm just not interested in doing anything anymore. All I do is watch TV – I just go to work, come home, and watch TV.'
(2) Sleep disturbance	'Well until recently, I've never had trouble sleeping. I go to bed around 11 and I can usually fall asleep right away. But now, I wake up in the middle of the night, and just lie there worrying about things … After an hour or so, I'll get back to sleep, but I never really feel rested anymore.'
(3) Appetite disturbance with no weight loss or weight gain	'My husband and I used to enjoy going out to dinner, but I haven't been interested in eating, so we haven't even being doing that. I'm just not interested in doing anything anymore. All I do is watch TV – I just go to work, come home, and watch TV.'
(4) Tired (low energy)	'I haven't felt like myself lately. I'm really tired all the time and I thought I should see a doctor. I just don't have any energy.'
(5) Loss of concentration	'I'll start to do something, and I'll forget what I was doing … that's just not like me. I can't seem to focus, and I'm having trouble getting my work done. … my boss was telling me about the things I need to do with my department, and my mind wandered off. I missed everything he said.'