

Assessing Clinical Judgment with Standardized Patients

Craig L. McClure, MD, Eric P. Gall, MD, Keith E. Meredith, PhD,
M. Annell Gooden, and John T. Boyer, MD
Tucson, Arizona

Family physicians and general practitioners see the majority of patients with uncomplicated rheumatic disease, yet information on database collection and clinical judgment in such practices is limited. Trained patients with uncomplicated rheumatic disease (standardized patients) were used to evaluate these abilities in 26 family physicians at the University of Arizona College of Medicine in blinded, but previously consented to, brief new encounters. Ability to formulate an assessment and to plan was evaluated as well as ability to collect diagnostic information.

Few physicians explored the psychosocial impact of the illness (4 percent) or the role of depression (0 percent). In the brief encounter with a localized complaint, little inquiry was directed to systemic disease (46 percent). Physicians more uniformly asked about the chief complaint (96 percent) and time of onset (88 percent). Physical examination items most commonly omitted were evaluation of systemic joint involvement (69 percent) and muscle wasting in the involved area (59 percent). Referral occurred on 15 percent of encounters and patient education occurred in 62 percent. Three quarters of physicians developed an adequate assessment and virtually all developed an adequate patient care plan.

Mendenhall, in a commissioned study, estimated that 32 percent of rheumatic disease patients see family physicians or general practitioners, 23 percent see internists, and 17 percent see orthopedic surgeons.¹ Only five percent of all

patients with rheumatic disease are cared for by rheumatologists. The remaining are seen by a variety of other physicians. It has been estimated that in the future there will be an even greater increase in primary care responsibility for these patients.^{2,3} It is therefore important that clinical judgment (defined here as collecting a data base and developing an assessment and care plan) of primary care physicians regarding rheumatic disease be investigated and monitored.

Many attempts have been made to audit primary care physicians' practice methods and clinical judgment.⁴⁻⁸ Chart review is of questionable value because a chart may not accurately reflect

From the Departments of Family and Community Medicine and Internal Medicine, University of Arizona, and the Southwest Arthritis Center, Tucson, Arizona. Requests for reprints should be addressed to Dr. Craig L. McClure, Department of Family and Community Medicine, University of Arizona Health Sciences Center, Tucson, AZ 85724.

what actually occurs in the examination room. Observation of a physician during a patient encounter by a known observer creates an artificial environment that may invalidate the behavior being observed.

Elstein and colleagues,⁹ in their book, *Medical Problem Solving: An Analysis of Clinical Reasoning* present a method employed in the study of medical diagnostic thinking by a group of experienced internists that is interesting in its psychological approach to the problem solving, judgment, and decision-making aspects of medicine. The standardized patient approach concentrates on the external manifestations of clinical problem-solving such as historical items, physical examination performance, discussion of assessment with patient, and development of a patient care plan. As such, it is more similar to the work of Smith and McWhinney¹⁰ and the work of Scherger et al¹¹ than that of Elstein et al.⁹

Whereas Scherger's group used written simulated patients to compare the diagnostic methods of family practice and internal medicine residents, the current report deals with 26 encounters of family physicians in an actual clinical setting. With Smith and McWhinney's approach in 1975, although the two groups of physicians were presented with clinical problems by a programmed patient, the physician was not blinded, and the situation did not simulate a clinical encounter. The present study is an attempt not to extend the psychological investigation and explanation of the processes by which physicians accumulate data or formulate assessments and plans, but rather to present an approach by which all elements of that process may be collected in an unobtrusive, blinded fashion.

Faculty at the University of Arizona have worked to develop patients with authentic disease who are able to record accurately the histories and physical examinations performed on them by health professionals as well as to record information on development of an assessment and health care plan. The term used for such an individual is *standardized patient*. The standardized patient in this case is a patient with uncomplicated rheumatologic disease who has been coached to deliver consistent historical information in each blinded physician encounter and to record reliably data pertinent to the encounter. Thus, standardized

patients function as unrecognized observers who can reliably record elements of patient-physician encounters that are generally unavailable by other means. Furthermore, standardized patients are able in a unique way to gather data on the ability of the physician to develop an assessment and patient care plan.

Standardized patients were used in this study to assess (audit) rheumatologic clinical judgment in a primary care setting, the University of Arizona Family Practice Program. Clinical judgment is used here to refer both to the acquisition of a diagnostic data base and to the problem solving inherent in arriving at a diagnosis and appropriate care plan. Included in this paper are the development of the program, selection and training of the standardized patients, development of the audit checklists, and data obtained on clinical judgment in a Department of Family and Community Medicine.

The use of standardized patients as a method of assessing health care providers represents a natural evolution of patient instructor programs.¹²⁻¹⁵ Such programs have been used at the University of Arizona to teach and assess the skills and knowledge necessary to conduct an adequate physical examination.

Approximately five years ago at the University of Arizona Gall et al¹⁶⁻¹⁸ began using patient instructors with advanced, but stable, rheumatic disease as teachers in the medical school curriculum, the rheumatology fellowship, and resident rotation. The emphasis with physician instructors has been to teach physical examination skills and determine mastery levels, not to evaluate the assessment or plan for a given patient situation. Moreover, as the care of someone with advanced disease is complicated, often requiring specialty input, assessments and plans for the rheumatology physician instructors often called for referral by the primary care physician to a rheumatologist. A need, therefore, existed for patients with common, uncomplicated rheumatic disease to assess skills of a higher order than those of obtaining a history and performing a physical examination, such as, for example, the ability of a primary care physician to develop a differential diagnosis and rational plan for patient care workup. Simultaneously data could be collected regarding how relevant historical and physical data were collected and recorded by physicians.

Methods

Selection of Patients

The five patients selected each had an uncomplicated rheumatic disease commonly seen in a medical clinic setting. Selection interviews included subjective evaluations for intelligence, appropriate ability to respond to unforeseen events in the clinic, reliability of character, and time commitment. The five rheumatic diseases of these patients included mechanical derangement of the knee, early rheumatoid arthritis, temporomandibular joint syndrome, myofascial syndrome (fibrositis), and degenerative disease of the back.

Development of Checklists

The study physicians (one family physician and one rheumatologist) performed a complete history and physical examination pertinent to the patient's disease process. A chief complaint was dictated by the standardized patient's particular problem. The patient history was converted to a standardized patient history by adjustments in time course justifying a patient visit and deletion of irrelevant details. As physical findings other than pain were constant over time, standardized patients were also coached to simulate the same degree of tenderness at each examination. Checklists of significant items developed from the above information were designed primarily to include points believed to be crucial in formulating a diagnosis and care plan; however, some items were inserted for data gathering or teaching purposes. Teaching items were inserted when they were considered to be illustrative of a general approach to the patient problem, even if not essential to diagnosis in the particular standardized patient.

Checklist items were included to measure the physician's ability to state the patient's diagnosis, discuss the prognosis, and to give patient education. Even if a physician was uncertain about the final diagnosis, if he clearly stated to the patient what his considerations were and why he was uncertain, he was still given credit for having discussed the diagnosis. Credit was similarly given for the prognosis. Credit was given for patient education if the patient felt the question "why?" had been answered in regard to development of the disease, recommendation for the treatment

program, and so on.

Information regarding the care plan could also be recorded. Multiple listings for a return-to-clinic appointment were included for ease of information recording. Spaces were allowed for such items as a definite return to clinic, a telephone call, or a return as needed. Similarly, lists of possible treatment choices, referral sites for physicians, and diagnostic and x-ray studies were included.

Three checklists were developed for each patient: history, physical examination, and chart audit, which derived from the former two. Four consultants (two family physicians and two rheumatologists) were selected to review the checklists. Each consultant performed his or her customary initial encounter with the standardized patient, after which the consulting physician reviewed the drafted checklists and suggested additions or deletions along with the rationale for each change. Checklists were then revised based on the researchers' reaction to these critical comments. This process provided content validity for the history, physical examination, and chart audit checklists developed for the standardized patients.

Standardized Patient Training

Following development and validation of the checklists, the standardized patients were coached for approximately 30 hours on (1) delivery of the tailored history and (2) reliable completion of the checklists. An important portion of this training was to prepare the standardized patients to anticipate various clinical approaches and thus to maintain a consistent history even when physicians phrased questions in unexpected ways.

Establishing Reliability and Validity

After training, the standardized patients completed four videotaped practice encounters with physicians who were aware of the patient's identity as a standardized patient, but unaware of the diagnosis. History and physical examination checklists completed by the standardized patient after each examination were referred to as the standardized patient "live" checklists. Two weeks later, checklists completed after viewing the videotaped encounter were referred to as the standardized patient "video" checklists. Com-

parison of the standardized patient live and video checklists provided an index of intrajudge reliability. Discrepancies between live and video checklists were thought most likely to reflect checklist deficiencies, standardized patient training deficiencies, the influence of participating in the encounter rather than viewing it, or standardized patient error.

To help determine the source of these discrepancies, each videotaped examination was viewed independently by three criterion judges including a family physician (CM), a rheumatologist (EPG), and the research assistant (MAG) in the program. Each judge completed history and physical examination checklists from the tape. Discrepancies among individual judges' assessments of each item were discussed while viewing the tape as a group to establish a consensus rating on the performance of each item. This rating became the criterion for judging the performance or nonperformance of each item on the checklists for the given videotape.

Standardized patient reliability was then judged by comparing standardized patient ratings for each item on both the live and video review against the criterion judges' consensus ratings. A ratio of the number of correct items by the standardized patient to the total number of items would yield the percentage of the total information available that was reported by the standardized patient. No standardized patient was allowed to enter the clinic setting until all measurements of reliability were .85 or higher. In other words, the standardized patient must match consensus on 85 percent of items when assessing the tape. Supplemental training was given not only to standardized patients who did not reach .85, but also to all standardized patients for any items on which the standardized patient disagreed with consensus.

Ongoing standardized patient reliability was evaluated on a regular basis using the videotapes. With four tapes, an interval of two months between each viewing of a given tape decreased the effect of memory on checklist completion. Comparison of performance on the video checklist to the consensus achieved by the criterion judges was a measurement of standardized patient reliability as well as a means for continued practice over and above that provided by encounters with physicians in the study situation.

The Study

Twenty-six family physicians, including residents and faculty, were asked in July 1981 to agree to an encounter with a standardized patient in the following year. To decrease the effect of performance anxiety, the physicians, although consenting to the encounter, were blinded to the standardized patient identity. For the encounter an appointment was made for a brief new-patient visit of 20 minutes. Following the examination, the standardized patient completed the validated history and physical examination checklists. The progress note was sent to the study team along with checklists.

The progress note of the standardized patient-physician encounter and the checklists were audited individually by the family physician and rheumatologist of the study team. Data was hence available both on what the standardized patient recorded of the encounter on checklists and what the physician recorded of the encounter on a progress note. The study physicians compared the information obtained by the physician and the physician's assessment and plan. As part of checklist development and standardized patient training, a list of appropriate and inappropriate differential diagnoses were developed for each standardized patient and incorporated into the audit checklist. Validity of a diagnosis being classified as appropriate or inappropriate was reviewed by the consultants who reviewed the checklists as well as by the team physicians. A global checklist was utilized for final recording of adequacy of information gathering for historical and physical examination items. Adequacy of the assessment given the knowledge obtained by the physician was also recorded. An educational feedback session focusing on rheumatologic skills was conducted with the physician by either of the two project physicians.

Results

Initially it was thought that items from the history and physical examination checklist of the standardized patients would not be comparable because of the different diseases involved. In review, it was possible to generate generic categories (eg, chief complaint, onset of symptoms) to

Table 1. Results From History Taking (N is variable)

Question Category	Number of Physicians Asking Question	Number of Physicians Having Opportunity To Ask Question	Percentage
Chief complaint	25	26	96
Time of onset	23	26	88
Onset of events	17	26	65
Onset symptom	11	15	73
Medical care at onset	15	26	58
Previous physician's diagnosis/treatment	9	26	35
Past information (eg, x-ray, physician address)	6	26	24
Past medication	13	26	53
Result of medication	7	23	33
Follow-up	5	19	27
Current symptoms	17	26	65
Precise location of symptoms	15	23	65
Inflammation	12	26	46
Systemic problems	12	26	46
What makes better	13	18	72
What makes worse	21	26	81
Morning vs evening pain	3	22	14
Other medical problems	19	26	73
Psychosocial impact	1	23	4
Sleep difficulty (depression)	0	15	0
Unsolicited information	8	26	31

allow for comparison of physician performance across all of the standardized patients for most categories. During the first year three of the standardized patients participated in the study. Although a total of 26 physicians were assessed, the total number of physicians for each item was not always equal to 26 because a question from each generic category was not always present on each checklist.

Table 1 displays results for history taking. Physician performance in eliciting the chief complaint and time of onset was excellent (25/26 and 23/26, respectively). However, on some other items performance was not so high. Although not all of the standardized patients had problems that would prompt an investigation of depression or sleep dif-

ficulty, no physician investigated this emotional issue. Similarly, only one of 23 physicians investigated the psychosocial impact of the disease on the patient's functioning. Few physicians (3/22) inquired about morning pain or stiffness as a survey question for inflammatory disease, and were similarly remiss on two thirds (16/23) of the cases in investigating the effect of current medications on the symptoms of illness.

Table 2 presents results for the physical examination. Surprisingly, 37 percent of all physicians omitted palpation of the involved painful area. Over two thirds of the physicians omitted examination of even one other joint in the body as a check for systemic rheumatic involvement. On the other hand, visual inspection of the involved area

Table 2. Physician Performance Results From the Physical Examination (N = 26)

Examination Task	Adequate No. (%)	Inadequate No. (%)	Not Performed
Introduction	24 (92)	0 (0)	8
Undress	19 (86)	0 (0)	14
Inspection of involved area	18 (71)	1 (4)	25
Palpation of involved area	13 (52)	3 (11)	37
Range of motion of involved area	16 (61)	2 (8)	31
Muscle wasting in involved area	9 (41)	0 (0)	59
Systemic joint involvement	8 (30)	1 (1)	69
Summary	26 (100)		0
Prognosis given	19 (73)		27
Patient education	16 (62)		38
Return to clinic	25 (96)		4
Appropriate therapy	13 (50)		50
Referral	4 (15)		85

was omitted by only 35 percent of the physicians. Although admittedly difficult to assess, range of motion of the involved joint was done by 69 percent of physicians.

Virtually all physicians made a summary statement to the patient. For a summary statement some communication to the patient concerning the physician's assessment was required. The possibility exists that a physician may not know the exact diagnosis or would have a series of differential diagnoses in mind. As long as diagnosis was discussed, the item was considered to have been performed. Similarly 73 percent of physicians gave a prognosis statement for which was required some statement as to the probable course of the disease and the effect on the patient. All but 4 percent of the physicians directly addressed a return appointment (return as needed, return for certain circumstances, or return at a certain time).

Four categories were evaluated by the rheumatologist and family physician of the project regarding the physician's assessment and plan: (1) adequacy of the amount of information obtained from the history for forming a proper assessment, (2) adequacy of the amount of information obtained from the physical examination for forming a proper assessment, (3) adequacy of the assessment given the reviewer's knowledge of the pa-

tient, and (4) adequacy of the plan, given the physician's assessment of the patient. This last item is the reviewers' impression of the ability of the physician to form a plan consistent with the physician's stated assessment. Each category was rated as (1) totally inadequate, (2) inadequate, (3) borderline, (4) adequate, or (5) totally adequate.

In both the history and physical evaluations, both reviewers felt that no physicians did a totally inadequate job. For example, on the physical examination (Table 3) (9/23) physicians were evaluated as borderline and (12/23) as adequate. No physicians were evaluated as having done either a totally inadequate or totally adequate job. The physician's assessment and plan was generally found to be more adequate than the history and physical information gathered as basis for assessment and plan. For example, reviewer 1 in category 4 (Table 3) found (13/25) adequate and (8/25) totally adequate.

A separate questionnaire administered during feedback sessions after the standardized-physician encounter revealed physician response to participating in the study. The physicians consistently reported that the evaluation performed by the standardized patient and the research team was accurate, 85 percent and 88 percent, respectively. Most physicians were in fact unaware of the iden-

Table 3. Ratings of Patient Assessment and Plan Evaluation by 26 Physicians

Category		Totally			Totally		
		Inadequate	Inadequate	Borderline	Adequate	Adequate	Omitted
Adequacy of the amount of information obtained from the history for forming a proper assessment	Reviewer 1	0	5	8	11	1	1
	Reviewer 2	0	6	8	9	1	2
Adequacy of the amount of information obtained from the physical examination for forming a proper assessment	Reviewer 1	0	4	9	12	—	1
	Reviewer 2	0	8	10	6	—	2
Adequacy of the assessment given the reviewer's knowledge of the patient	Reviewer 1	0	3	4	11	7	1
	Reviewer 2	0	1	4	17	2	2
Adequacy of the plan given the reviewer's knowledge of the patient	Reviewer 1	0	2	2	13	8	1
	Reviewer 2	0	—	5	14	5	2
Adequacy of the plan given the assessment formed by the physician	Reviewer 1	0	3	1	13	8	1
	Reviewer 2	0	—	1	14	9	2

tity of the patient as a standardized patient (88 percent). Generally the program was well received, with only 19 percent finding the program to be an invasion of privacy.

Discussion

This study demonstrates the use of trained patients to evaluate clinical judgment, defined both as database collection and development of an assessment and care plan. This unique approach provides an accurate means of unobtrusively auditing physicians' encounters. Because the physician was unaware of being audited at the moment, behavior in the encounter was assumed to be average for the physician, neither better nor worse than it would be in a routine office visit. This technique can be useful in peer review and eventually in planning training programs once suf-

ficient data have been collected to describe norms of clinical judgment in various settings.

The findings indicate that despite emphasis on behavioral science in family practice, investigation of the social impact of disease on the individual and conversely the role of emotion in causing a disease may not occur as often as would be expected, at least in a brief first encounter. Two items that were performed only a quarter of the time—"where can old records be obtained" (6/26) and "what was the follow-up for initial illness" (5/19)—may have reflected the study setting. The appointment with the physician was a first, brief visit and only 20 minutes long. It may have been that because of time constraints and the local nature of the complaint that these items were deemed of lower priority than other items. One hopes that with a second and third visit to the same physician, the percentage of physicians investigating these items would increase with establishment of a data base and continuity relationship.

As far as the encounter itself, it seems that most physicians, when presented with a problem that seems localized, tended not to explore systemic disease on the initial encounter either in history or physical examination. Again because of study design with the short appointment, it may be that time limitations caused the physician to concentrate only on the localized problem. Nevertheless, study consultants agreed that, as a general approach to a correct differential diagnosis, indicators of systemic or inflammatory disease should be surveyed.

These data lend themselves to the interesting suggestion that there may not be a direct correlation between the amount of information gathered and the adequacy of the assessment and plan. Although previous experience has also shown that skilled physicians often make a reasonable diagnosis without much input from the history and physical examination, not enough information exists at present to determine whether there is a particular level of training at which an apparently less than adequate history and physical examination may yet yield a perfectly adequate assessment and plan. As additional data are collected, one may find that the idea of what is an adequate history and physical examination may be altered.

In summary, data have been collected in an innovative fashion from blinded consenting practicing physicians in the office setting by use of actual patients, who can provide an accurate, reliable, and credible way of auditing physician encounters unobtrusively. Potentially such data may be used in peer review and in planning training programs once sufficient material has been collected. In the family practice academic setting in a brief new encounter, these data would suggest that preconceived ideas regarding family physicians' emphasis, such as investigation of emotional factors, may not be correct.

The standardized patient means of studying clinical judgment has been extended to physicians in internal medicine. Data from these standardized patients is being gathered comparing the practice patterns of comparable academic family physicians and internists at the University of Arizona. With an increased data pool the influence of amount of training on the ability to make an appropriate assessment and plan may be clarified. Additionally, future studies will extend into the

community to investigate practice patterns of community physicians.

Acknowledgment

This study was supported by the University of Arizona College of Medicine and by an Arthritis Center Grant No. 2P60 AM20597-06 from the National Institutes of Health.

References

1. Mendenhall PC: Rheumatology Practice Study Report. Los Angeles, University of Southern California Press, 1978
2. Rudd E, Lockshin M: Education in rheumatology for the primary care physician. *J Rheumatol* 1978; 5:99-111
3. Schumacher HR, Lockshin M: Manpower and fellowship education in rheumatology, 1980. *Arthritis Rheum* 1981; 24:1168-1172
4. Barrows HS: Simulated Patients: The Development and Use of a New Technique in Medical Education. Springfield, Ill, Charles C Thomas, 1971
5. Payne BC: The medical record as a basis for assessing physician competence. *Ann Intern Med* 1979; 91:623-629
6. Payne BC, Lyons T, Dwarshins L, et al: The Quality of Medical Care: Evaluation and Improvement. Chicago, Hospital Research and Educational Trust, 1976
7. Peterson OL, Andreas LP, Span LS, et al: An analytical study of North Carolina general practice 1953-1954. *J Med Educ* 1956; 31(2):1-65
8. Gall EP, Meredith K, Boyer JT, et al: The use of trained patient instructors for teaching assessing rheumatological care. *Arthritis Rheum* 1984; 27:557-563
9. Elstein A, Kagan N, Shylman LS, et al: Medical Problem Solving: An Analysis of Clinical Reasoning. Cambridge, Mass, Harvard University Press, 1978
10. Smith DH, McWhinney IR: Comparison of the diagnostic methods of family physicians and internists. *J Med Educ* 1975; 50:264-270
11. Scherger JE, Gordon MJ, Phillips TJ, et al: Comparison of diagnostic methods of family practice and internal medicine residents. *J Fam Pract* 1980; 10:95-101
12. Stillman PL, Ruggill JS, Sabers DL: The use of practical instructors to evaluate a complete physical examination. *Evaluation Health Professions* 1978; 1(1):49-54
13. Stillman PL, Brown DE, Redfield DL, Sabers DL: Construct validation of the Arizona clinical interview rating scale. *Educ Physiol Meas* 1977; 37:1031
14. Still PL, Ruggill JS, Rutala PJ, et al: Patient instructors as teachers and evaluators. *J Med Educ* 1980; 55:186-193
15. Stillman PL, Ruggill JS, Rutala PJ, et al: Students transforming into an American medical school: Remediating their deficiencies. *JAMA* 1980; 243:129-133
16. Gall EP, Stillman PL, Rutala PJ, et al: The arthritis patient instructor in competency board teaching and evaluation of rheumatic diseases care by health professionals, abstract 21. In *Arthritis Health Professions Annual Meeting Abstracts*. New York, Arthritis Foundation, 1979, p 226
17. Gall EP, Boyer JT, Riggs GE, et al: Physician skills in musculoskeletal exams and care planning as measured by patient instructors, abstract 25. In *Arthritis Health Professional Association Meeting Abstracts*. New York, Arthritis Foundation, 1981, p 152
18. Gall EP, Boyer JT, Meredith KE, Gooden MA: Assessment of physical examination skills by the patient instructor. *Arthritis Rheum* 1981; 24:S-102