

Simulated surgery in the summative assessment of general practice training: results of a trial in the Trent and Yorkshire regions

JUSTIN ALLEN

ALISON EVANS

JOHN FOULKES

ADRIAN FRENCH

SUMMARY

Background. General practice registrars are now required to undertake a summative assessment of their consulting skills. Simulated surgeries have been developed as an alternative to the existing method of assessing video-recorded consultations.

Aim. To evaluate the simulated surgery assessment method, developed in the General Practice Postgraduate Education Department in Leicester, for use in assessing general practice consultation skills.

Method. General practice registrars in Leicester performed two eight-patient simulated surgeries separated by four weeks. Assessment outcomes were compared to demonstrate the consistency of the method. Pilot surgeries in Yorkshire were videotaped, and then rated by video-raters trained for summative assessment.

Results. The method consistently identified those registrars who were competent and those who were not yet competent in consulting skills. It proved acceptable to candidate doctors and has fewer resource requirements for both examiners and candidates than other consulting skills assessment methods.

Conclusion. The method developed in Leicester and successfully transferred to Yorkshire is feasible on a large scale, and offers an acceptable alternative to other consulting skills assessment methods. In this study it consistently identified competent from incompetent candidate doctors.

Keywords: simulated surgery; summative assessment; training; Trent; Yorkshire.

Introduction

IN May 1993 the Joint Committee on Postgraduate Training for General Practice (JCPTGP) published its plans for a system of summative assessment to be completed by doctors before a Certificate of Training could be issued.¹ There are four elements to this assessment, one of which is competence in consulting skills: doctors must demonstrate that they perform at or above

the minimum standard required for independent practice.

The use of simulated patients for the assessment of clinical competence is becoming increasingly accepted^{2,3,4} and is now offered as an alternative in the membership examination of the Royal College of General Practitioners.⁵ Studies carried out in the USA,^{6,7} New Zealand,⁸ Holland,⁹ and the United Kingdom^{5,10-12} have shown the method to be superior to written tests, to be reliable and valid, and to be acceptable to doctors and patients. Professional concerns that patients would be unreliable in the assessment of doctors are not borne out by previous studies.^{7,8,13-15}

A method for assessing whether a doctor has the consulting skills required in general practice has been under development since 1993 in the General Practice Postgraduate Department, Leicester. It has aimed to demonstrate that general practitioner registrars who successfully complete the assessment have reached the standard of minimum competence, and those who fail have not yet reached that standard. This paper describes how the standards are set, reports the results of system testing, and considers the acceptability of the method to registrars.

Method

Development of simulated surgeries

Simulations are developed from real consultations using video-recordings, after gaining informed consent from the patients concerned. Patient simulators work with the doctor who performed the consultation, using the video and the doctor's knowledge of the patient, to develop the simulation. The system of training patient simulators for this assessment technique has been fully described.^{10,11} As well as being trained for their simulation, the patient simulators are also trained in the application of the marking schedules.

In the assessment process, candidate doctors carry out a simulated surgery of eight consultations, each lasting 10 minutes, with patient simulators in a real general practice setting.² At the end of each consultation the simulator completes two marking sheets. The first, a medical checklist, is set by the examiners and covers selected areas of the consultation from the doctor's viewpoint, and the second is a subjective evaluation by the simulator. This process takes five minutes, during which the candidate doctors complete an 'escape' sheet, giving their view of their strengths and weaknesses in the previous consultation.

Standard setting

Pass/refer standards are set for each simulation for the medical checklist by an expert group of working general practitioners, who are familiar with the technique and who have themselves consulted with the patient simulator. The checklist is then tested for each simulation in a series of pilot surgeries, undertaken by registrars at all stages of training. The registrar must pass six out of the eight medical checklists in order to pass the surgery. The results are scrutinized for problems in the performance of the medical checklists (e.g. statements that are ambiguous or non-discriminating), and the lists are revised if necessary to ensure

J Allen, FRCGP, associate adviser in general practice and coordinator; and J Foulkes, PhD, educational consultant; Leicester Simulated Patient Project. A French, MRCP, course organizer, Leicester Vocational Training Scheme, and Research Fellow, Leicester Simulated Patient Project, General Practice Postgraduate Education Department, Leicester General Hospital. A Evans, FRCGP, senior lecturer in general practice, Division of General Practice and Public Health Medicine, University of Leeds.

Submitted: 26 February 1997; accepted: 21 January 1998.

© British Journal of General Practice, 1998, 48, 1219-1223.

that they are set at the level of minimum competence. Pass rules are considered in an iterative process by the expert group and amended if it appears that the standard has not been set correctly. An example of a medical checklist is given in Figure 1.

In addition to completing the medical checklist as a process of recording what happens in a consultation, the patient simulator fills in a patient satisfaction questionnaire (Figure 2), designed to capture the patient simulator's feelings and attitudes.

The 10 questions are scored by allocating one mark for each 'Yes' and a half-mark for each 'Not Sure'. The maximum possible score is therefore 10 points for each consultation, and 80 points in an eight-consultation surgery. The pass mark was set, during extensive piloting of the method over the past three years, at 72 out of 80 for this checklist. The process was again an

iterative one, in which the scores of the candidate doctors were reviewed and compared with expectations, and a consensus on the pass mark developed.

Testing the system

The standard-setting process was formalized by refining the medical and patient satisfaction checklist scores during single surgery pilots with registrars groups from the Trent Region, Yorkshire, and Wrexham. In all, 24 simulated surgeries, each involving eight patients, were carried out in seven different locations during the development of the method; 64 candidate doctors took part.

When all the pass marks had been set, the assessment was test-

CLINICAL CHECKLIST			
Patient			
Doctor			
<u>Circle Yes or No</u>			
1	I understand that the bleeding is not caused by serious disease	yes	no
2	I accepted some treatment for my sinuses	yes	no
3	I told the doctor my mother had died recently	yes	no
4	I was offered help with my grief	yes	no
5	I understand no further investigations are needed	yes	no
6	I may come to see the doctor again if I wish	yes	no
7	I understand that how I feel is normal after my loss	yes	no

Figure 1. Clinical checklist.

PATIENT RATING SCALE		
Doctor:..... Patient:.....		
The statements in each group should be scored separately, and should be used in your decision.		
	Satisfactory	Not satisfactory
A. COMMUNICATION	<input type="text"/>	<input type="text"/>
I had adequate opportunity to express my problems Yes No Not sure		
The nature of my problem was explained to me. Yes No Not sure		
I was able to discuss what needed to be done to help me. Yes No Not sure		
The doctor used language I could understand. Yes No Not sure		
B. DOCTOR/PATIENT RELATIONSHIP	<input type="text"/>	<input type="text"/>
I was treated with respect. Yes No Not sure		
The doctor was sensitive to my feelings. Yes No Not sure		
I felt at ease with the doctor. Yes No Not sure		
C. PROFESSIONAL ATTRIBUTES	<input type="text"/>	<input type="text"/>
I felt the doctor was competent. Yes No Not sure		
I trust this doctor. Yes No Not sure		
D. OVERALL	<input type="text"/>	<input type="text"/>
I would consult this doctor again. Yes No Not sure		
COMMENTS:		

Figure 2. Patient satisfaction checklist.

ed in two ways. Single surgery pilots using 15 candidate doctors at different stages of training were carried out to determine if the result matched the expectation for each; these pilots took place in Leicester. Test-retest reliability was then determined in a double-surgery pilot, in which a further 15 candidate doctors were asked to undergo two simulated surgery assessments, using different simulations, within four weeks. In this way, candidate variables were kept to a minimum. The hypothesis in testing the method for consistency was that the same candidates passing and failing the assessment would be identified in each of two simulated surgeries. A mix of inexperienced and experienced candidate doctors was used. From these double surgeries it was possible to analyse data on 239 consultations. The decision rule used by the expert group was that the candidate doctor had to pass both elements of the assessment in order to demonstrate competence.

Rating by video assessors

A further check on the standard-setting process was carried out in Yorkshire. Videotapes of simulated surgeries carried out by 11 of the candidate doctors who took part in the Yorkshire pilot surgeries were assessed by two raters trained for the summative assessment of videotapes, and the results compared. The candidate doctors chosen to have their videos rated either were in the final six months of their training, or had not reached the minimum standard required in the simulated surgery.

Results

The results of the single surgery pilots are given in Table 1. Of the 15 candidates, 11 were in their final year of training; one of these failed the assessment. Of the remaining four, three failed on the medical checklist alone; one was in the first year of training and two in the second year. The remaining candidate doctor, who failed both elements, was in the first year of training. The results were in line with the expectations of the teachers of these candidate doctors, and supports the view that more experience in general practice improves competence in consulting skills.

The results of the double surgeries are given in table 2. Of the 15 candidates, seven passed on both occasions, seven failed on both occasions, and one failed on the first occasion but passed on the second. In the first of the two surgeries, eight passed the

medical checklist assessment. Of these eight, one failed the patient satisfaction assessment and therefore the assessment as a whole. When the same candidates took the second simulated surgery, all eight again passed the medical checklist assessment, and the same one candidate failed the patient satisfaction assessment.

Six candidates failed the medical checklist assessment. Of these six, three passed the patient satisfaction assessment on both occasions; one failed on both occasions; and two failed on the first occasion but passed on the second. The results of the failed candidates are analysed further in table 3.

Only one of the fifteen candidates had a different medical checklist result on the two occasions, failing the first time and passing the second time. It seems that this was largely due to circumstances: the candidate arrived late and flustered for the first surgery. By the medical checklist rules he passed four of the seven consultations he was able to conduct, which was not equivalent to the six-out-of-eight rule. On the second occasion he passed seven of the eight consultations. This was reflected in his patient satisfaction performance. On the first occasion he scored 64.5 out of a possible 70, which equates to 73.7 out of 80 (a passing score); but on the second occasion he scored a much more convincing 79 out of 80.

Comparison with video-rating

Of the 11 candidate doctors who had their simulated surgeries videotaped and rated, two were referred by both assessors. They had passed only one and three, respectively, of the eight simulated consultations. One other registrar who passed three simulated consultations was passed by one assessor and referred by the other. The remaining eight registrars passed five or more simulated consultations; six of these were passed by both assessors, while two were passed by only one of the assessors.

Acceptability

The results of the Yorkshire study are published more fully elsewhere.¹² In Leicester, 85% of candidate doctors taking part in pilot studies felt that the patient's viewpoint should be part of summative assessment, and 93% found marking by patient simulators acceptable.

Table 1. Results of single surgery assessments of trainees at different stages of training.

	Number	Passed overall	Failed medical list	Failed patient list
Final year	11	10	1	0
Second year	3	0	3	0
First Year	1	0	1	1

Table 2. Summary of the results of double surgeries (15 candidates).

	Passed overall	Passed medical list	Passed patient list
First surgery	7	8	11
Second surgery	8	9	13

Table 3. Analysis of failed candidate scores in medical and patient satisfaction checklists.

	Failed both checklists	Failed medical checklist only	Failed patient checklist only
First surgery (8)	4	3	1 ^a
Second surgery (7)	1	5	1 ^a

^aSame candidate.

Discussion

The results of the double surgeries presented here demonstrate the test-retest reliability of the method in identifying those who do not achieve minimum competence. Although candidate numbers were small, each of the 15 doctors submitted evidence on four hours of consulting (10 minutes per consultation, with five minutes for writing up after each one); a total of 239 consultations were held. The numbers of candidates assessed to establish reliability in this project are larger than those in the studies on method reliability published by Fraser¹⁶ (five candidates) and Campbell¹⁷ (10 candidates). It is interesting to note that Campbell found limited agreement between video assessors on the scoring of consultations, but greater consistency in the decision on overall competence. This parallels our decision to consider the reliability of the overall assessment rather than of individual simulations.

The overall concept of 'test consistency' was used for a number of reasons. The reliability of any measure used in summative assessment reduces to a determination of how much of the variation in the set of scores is due to differences between the individuals being tested and how much to inaccuracies in the measurement of particular individuals.¹⁸ In any assessment in which an examiner observes a performance and records a judgement, the examiner may not behave consistently and different examiners may not judge the same performance in the same way. This source of measurement error is so obvious that it is often treated as the only source of error, with inter-observer reliability being given undue weight. However, variability among candidates in their responses to different tasks is typically greater than variability among markers in their response to a single performance.¹⁹ Secondly, the pass is achieved by two different systems of marking, one case-specific and the other derived from the simulated surgery overall. During the surgery the simulators are acting on behalf of the examiners in two separate ways — as a recorder of events in the medical checklist, but as a judge of performance in the patient satisfaction checklist. Simulators, provided they have been trained, have been demonstrated by others to be reliable in applying marking schedules in the assessment of clinical competence.^{20,21} This method provides extensive training for the simulators during the development of each consultation. Finally, if the assessment method as a whole demonstrates consistency in identifying competent from incompetent candidates, then reliability of its constituent parts can be inferred.

The most important difference from other assessments of consulting skills is that this simulated surgery assessment includes the patient's perspective in two ways. The most obvious is the patient satisfaction rating of the quality of the doctor-patient interaction, in which the patient simulator gives a verdict on the candidate doctor's skill in communicating and building rapport: a process of judging rather than of recording. The important distinction is that while the patient simulator cannot legitimately judge a doctor's clinical behaviour, but can only act as a recorder of events that are evaluated by others, he or she is the only qualified judge of the doctor-patient interaction seen from the perspective of the patient; this cannot be gained from a video. The inclusion of a structured patient satisfaction element adds a valuable new dimension to an assessment of this type. The issues tested by this part of the method are an important component of all consultations in general practice, are common to all whatever the clinical content, and are thus not case-specific. They are not tested by other methods of consulting skills assessment.

The patient-simulator is also involved in determining the patient's clinical agenda when developing the simulation. This is

what's important to a lay person with a particular problem or set of symptoms. As Middleton²² points out, discovering the patient's agenda is crucial for a successful consultation. Video assessors can make only an informed guess about the agenda of the patient on the video. The simulated surgery is therefore a more valid test of the candidate doctor's communication skills with adult patients. The validity of this simulated surgery method of assessing consultation skills has been discussed elsewhere.²³ The development of simulations based on real consultations gives a high face validity, and this is borne out by the reactions of doctors who consult with the simulators, and by the comments of video assessors who viewed the tapes.

The range of competencies which can be tested is limited in both methods. In theory, any type of consultation may be video-recorded, including those with children and those with either multiple patients or relatives present (situations not covered by simulated surgeries), although a wide range of competencies can be tested. On the other hand, the material submitted on video is selected, and is governed by the case mix seen and by patients' consent to be taped. A registrar who is sensitive to patient wishes may well have difficulty recording consultations relating to emotional or mental health problems. A further problem found by Campbell with video assessment is that the medical correctness of the consultation is not always possible to ascertain.²⁴ This does not arise with simulated patient consultations.

Comparison with video assessment is not straightforward for several reasons, but our work suggests that the standard required in simulated surgeries is equivalent, and it has also been demonstrated that the method can be transferred to other regions¹² and is practicable in terms of time and resources. A great advantage over video is that medical assessor time is only needed at the development stage, and not for the actual surgeries. Perhaps more importantly, it is a standardized test, and therefore avoids the unfairness of different access to video equipment and different consultation settings. Registrars are not able to select only their best material, as with video recordings. This simulated surgery method could be offered as a cheap and convenient alternative to video assessment.

Detailed feedback on strengths and weaknesses can be given to registrars immediately after simulated surgeries. This should help referred registrars to target their efforts to develop their skills. Feedback to the educational system from simulated surgeries can easily be given in quite detailed form, as the competencies that are being tested are pre-determined and standardized. It would be time-consuming and difficult to collect feedback with the same level of accuracy from video assessors.

Quality control is a continuing feature of simulated surgeries. The patient simulators frequently view their reference video to ensure that the simulation remains consistent. Monitoring of results identifies simulations that are not discriminating. Video-recording of simulated surgeries enables the medical examiners to check for consistency, and also enables external review. Joint surgeries with Leicester and Yorkshire simulators have been held to check consistency between regions. It is proposed that 10% of consultations should be videotaped for external quality control.

Conclusion

We believe that the simulated surgery method described here is a fair, valid, and acceptable method for the summative assessment of consulting skills. This study has shown it to be reliable in distinguishing registrars who are competent in consulting skills from those who have not yet reached the level of minimum competence. The method has significant advantages over video-

recording in providing a standardized challenge, in convenience, and in giving feedback to candidates. The research team accepts that further work is needed to confirm these results, and research is currently under way.

References

1. Joint Committee on Postgraduate Training for General Practice. *Report of the Summative Assessment Working Party*. London: JCPT-GP, 1993.
2. Barrows HS. An overview of the uses of standardized patients for teaching and evaluating clinical skills. *Acad Med* 1993; **68**: 443-453.
3. Stillman P, Swanson D, Regan MB, *et al*. Assessment of clinical skills of residents utilizing standardized patients. *Ann Intern Med* 1991; **114**: 393-401.
4. Rethans JJ, Sturmans F, Drop R, van der Vleuten C. Assessment of the performance of general practitioners by the use of standardized (simulated) patients. *Br J Gen Pract* 1991; **41**: 97-99.
5. Bingham L, Burrows P, Caird R, *et al*. Simulated surgery - using standardized patients to assess the clinical competence of GP registrars - a potential clinical component for the MRCGP examination. *Education for General Practice* 1996; **7**: 102-111.
6. Klass D, Fletcher E, King A, *et al*. Developing a standard patient test of clinical skills at the National Board of Medical Examiners. In: *Approaches to the assessment of clinical competence*. Dundee: Centre for Medical Education, University of Dundee, 1992.
7. Vu N, Marcy M, Colliver J, *et al*. Standardised patients' accuracy in recording clinical performance check-list items. *Med Educ* 1992; **26**: 99-104.
8. Thomson AN. Reliability of consumer assessment of communication skills in a postgraduate family practice examination. *Med Educ* 1994; **28**: 146-150.
9. Pieters IM, Touw-Otten FWW, de Melker RA. Simulated patients in assessing consultation skills of trainees in general practice vocational training: a validity study. *Med Educ* 1994; **28**: 226-233.
10. Rashid A, Allen J, Thew R, Aram G. Performance based assessment using simulated patients. *Education for General Practice* 1994; **5**: 151-156.
11. Thew R, Worrall P. The selection and training of patient simulators for the assessment of consultation performance in simulated surgeries. *Education for General Practice* 1998. (In press.)
12. Sutcliffe R, Evans A, Pierce B, Bahrami J. Simulated surgeries — feasibility of transfer from region to region. *Education for General Practice* 1998. (In press.)
13. Rethans JJE, van Boven CPA. Simulated patients in general practice: a different look at the consultation. *BMJ* 1987; **294**: 809-812.
14. Stillman PL, Ruggil JS, Rutala PJ, Sabers DL. Patient instructors as teachers and evaluators. *Med Educ* 1980; **55**: 186-193.
15. Van der Vleuten CPM, Swanson DB. Assessment of clinical skills with standardised patients: state of the art. *Teaching and Learning in Medicine* 1990; **2**: 58-76.
16. Fraser RC, Mckinley RK, Mulholland H. Consultation competence in general practice: testing the reliability of the Leicester assessment package. *Br J Gen Pract* 1994; **44**: 293-296.
17. Campbell LM, Howie JGR, Murray TS. Use of videotaped consultations in summative assessment of trainees in general practice. *Br J Gen Pract* 1995; **45**: 137-141.
18. Thorndike RL. Reliability. *Educational Measurement*. Washington, DC: American Council on Education, 1951. p595.
19. Coffman WE. Essay examinations. *Educational Measurement*. 2nd edn. Washington, DC: American Council on Education, 1971. p281.
20. Colliver J, Verhuist S, Williams R, Norcini J. Reliability of performance on standardised patient cases: a comparison of consistency measures based on generalisability theory. *Teaching and Learning in Medicine* 1989; **1**: 31-37.
21. Stillman P, Swanson D, Smees S, *et al*. Assessing clinical skills in residents with standardised patients. *Ann Intern Med* 1986; **105**: 762-771.
22. Middleton J. The exceptional potential of the consultation revisited. *J R Coll Gen Pract* 1989; **39**: 383-386.
23. Allen J, Rashid A. What determines competence within a general practice consultation? Assessment of consultation skills using simulated surgeries. *Br J Gen Pract* 1998; **48**: 1259.
24. Campbell LM, Howie JGR, Murray TS. Summative assessment: a pilot project in the West of Scotland. *Br J Gen Pract* 1993; **43**: 430-434.

Acknowledgements

The authors would like to thank the teams of simulators in Leicester and Yorkshire, and also the registrars, trainers, and course organizers who gave up their time to assist in the development of this method.

Address for correspondence

Dr J Allen, General Practice Postgraduate Education Department, Leicester General Hospital, Gwendolen Road, Leicester LE5 4PW.

Early Abortion is a simple medical procedure and we aim to make it as woman-friendly as possible.

DAYCARE EARLY ABORTION SERVICE

new

Fast appointments, one hour stay in a relaxed environment

Centres Nationwide **0800 716 390**

GP Information pack **0171 574 7356**

Also Vasectomy and Female Sterilisation
internet: <http://www.mariestopes.org.uk>

MARIE STOPES

• CARING CLINICS SINCE 1921 •

A reg. charity providing reproductive healthcare worldwide



Reg No.
FS 32585

The Royal College of Psychiatrists

Management of Imminent Violence

Clinical Practice Guidelines to Support Mental Health Services
Clinical Practice Guidelines Work Group

This report publishes the findings of the most comprehensive and systematic research yet into the management of violence in clinical settings. It sets guidelines for clinical practice to be implemented in hospitals and psychiatric units throughout the UK and will be of interest to colleagues overseas who are seeking guidance in this area. *Management of Imminent Violence* will be invaluable to all professions and grades working in acute and community mental health settings as well as to primary care teams and 'informal' carers of people with mental health problems. **Price £20.00**

Please order from: Book Sales, Royal College of Psychiatrists, 17 Belgrave Square, London SW1X 8PG. Tel: 0171 235 2351, ext. 146.

Advances in Psychiatric Treatment

Journal

March 1998 Issue: Cycles of Abuse

Recent conferences on Cycles of Abuse have given a clear indication that psychiatrists from all specialities seek continued education, support and training about this area of practice. It is also an area of concern for GPs and nursing staff. This issue succinctly presents current knowledge and the part that mental health professionals can play in the prevention of cycles of abuse. Containing nine articles in total, the issue will be of interest to all professionals involved in the care of children, particularly those involved in caring for children with mental health problems or learning difficulties. **Price £12.00**

Please order from: Publications Subscription Department, Royal Society of Medicine Press Ltd, PO Box 9002, London W1A 0ZA, UK.
Tel: 0171 290 2927/8; Fax: 0171 290 2929.