The simulated surgery — an alternative to videotape submission for the consulting skills component of the MRCGP examination: the first year's experience

PETER J BURROWS

LIZ BINGHAM

SUMMARY

Background. *The simulated surgery was developed to examine the consulting skills of general practice (GP) registrars by observing their consultations with standardized patients. It was introduced in 1997 as an alternative to videotape submission in the consulting skills component of the Membership of the Royal College of General Practitioners (MRCGP) examination for those candidates who are unable to prepare a videotape.*
Aim. *To describe the methodology of the examination and to report on the first year's experience.*
Method. *The development of the cases and the techniques of marking and standard setting in the simulated surgery are described.*
Results. *Thirty-eight GP registrars took part in pilot examinations and 37 candidates were examined for the MRCGP. The distribution of their marks and the resulting pass/fail decisions are reported. The reliability of the 20-case simulated surgery, using Cronbach's alpha coefficient, is greater than 0.85.*
Conclusion. *The simulated surgery is a feasible, valid, and reliable examination of consulting skills. Cost and manpower requirements remain a problem, but these are being addressed by current plans.*

Keywords: simulated surgery; GP registrars; MRCGP examination; consulting skills.

Introduction

IN 1995, the Council of the Royal College of General Practitioners decided that a consulting skills component should be included in the MRCGP examination. Two methodologies were developed in parallel by the Examination Board: one being based on the viewing by examiners of videotapes of the candidate consulting with his/her patients in their own setting, and the other being based on observation of the candidate in consultation with a series of simulated patients in an examination setting. The latter, which was derived from the Objective Structured Clinical Examination (OSCE),[1] came to be known as the 'simulated surgery'.

It was appreciated from an early stage that there would be candidates who would have insuperable difficulty in providing a videotape of their practice. The simulated surgery is therefore offered as an alternative to the submission of video-recorded consultations to those candidates who can show good reasons why they cannot provide a videotape. They include those whose patients have a religious prohibition against video-recording or withhold their consent, those who do not customarily consult in English, and those who are not currently in active general practice (locums, hospital posts, career breaks).

Candidates in the simulated surgery undertake 20 10-minute consultations with standardized patients who are role-players trained to present the same case to each candidate. The simulated surgery is organized on a circuit similar to an OSCE but the candidates remain within their consulting rooms and are visited in turn by a series of patients. Each patient is accompanied by an examiner who observes the consultation but takes no part in it. The examiner marks the candidate's performance on a marking schedule that is specific to the case. The candidate's case scores are summed to provide an overall examination score for which a passmark is set.

The development of the simulated surgery has been described in detail.[2] A series of eight pilot surgeries was carried out between December 1993 and January 1995, using eight cases to examine 87 GP registrars. The results and psychometric properties of this series have been analysed.[3] The inter-case reliability of this eight-case format, calculated by generalizability theory, was 0.65. We subsequently adopted the 20-case surgery as the standard test. This was done to ensure adequate reliability and coverage of the blueprint. It has allowed us to run consecutive morning and afternoon circuits of 10 cases, using the same examiners and role-players. The organization of the examination and the briefing for candidates is described in Chapter 8 of *The MRCGP Examination*.[4]

Method

Case writing, selection, and content

In the overall examination blueprint, the consulting skills component is required to examine the skills of problem solving in a case-specific context, clinical management, personal care (matching principles to individual patients), verbal communication, and the consultation process. The cases are written by a small group of examiners who are all active in general practice, using a list of topics generated by discussion. The documentation consists of briefing for the role-player, records for the candidate, and a marking schedule for the examiner. The patient briefing tells the role-player their biographical details, their presenting complaint, their concerns and expectations of the consultation, and the information they should give in response to the candidate's questions. The records provide basic information about the patient, such as their name, age, sex, and background history, but the details are deliberately kept to a minimum to avoid distracting the candidate into following irrelevant leads. They may contain important information about prior events such as a consultation with a colleague or the practice nurse, some test or X-ray results, or a letter from a hospital consultant or another practitioner.

P J Burrows, MRCP, FRCGP, convener of the Simulated Surgery MRCGP examination and general practitioner, Romsey, Hampshire. L Bingham, FRCGP, former convener of the Simulated Surgery MRCGP examination, GP adviser to the Health Service Commissioner (Ombudsman), Newbury, Berkshire.

To compile the surgery, 20 cases are selected from the case bank using a blueprint. This procedure aims to provide a full range of patient ages and sexes; a good spread of acute, chronic, and 'service' consultations; a balance between common routine conditions and less common conditions that are important to recognize and treat correctly; and adequate sampling of each of the skills that are to be tested. These skills are classified into five domains, each of which will usually be sampled by the case-marking schedule:

1. *Data gathering*
   Taking a focused and efficient history that elicits relevant information and excludes any potentially serious conditions. Appropriate physical examination using a technique which confirms normality and would reveal any abnormal findings. Making use of information from the records provided.
2. *Doctor–patient interaction*
   Welcome and courtesy.
   Sensitivity to the patient's feelings.
   Facilitating expression of the patient's story by listening skills and using non-verbal cues.
   Awareness of the patient's concerns and expectations.
   Respect for patient's confidentiality and autonomy.
   Empathy.
3. *Communication*
   Explanation of candidate's assessment of the problem.
   Explaining the options for treatment.
   Negotiating the patient's agreement to a management plan.
   Advising and educating the patient.
   Breaking bad news.
   Using appropriate language and checking patient's understanding.
4. *Management*
   Devising a safe and effective management plan.
   Including a range of options where available.
   Rational prescribing, investigation, and referral.
   Sensible use of time and resources.
5. *Anticipatory care*
   Recognizing implications for the patient and others.
   Appropriate follow-up and surveillance.
   Anticipation of future problems.
   Opportunistic health promotion and advice.

Diagnostic skills are tested elsewhere in the examination; diagnoses are usually obvious or provided in the simulated surgery, since a wrong diagnosis will invalidate the assessment of the last three domains. The cases are tried out with mock candidates and examiners who are new to the case, and modified according to the way they run, before being included in a surgery.

Marking, scoring, and standard setting:

Each consultation is marked on a schedule with five items describing key tasks that all doctors would be expected to address. Marks are awarded on a five-point scale using the following grades: 'good', 'satisfactory', 'barely adequate', 'not very

good', and 'unsatisfactory'. Anchor statements are used to define what constitutes 'good' and 'unsatisfactory' performance on each item. The definition of 'good' performance is partially itemized to assist in reliable assignment of the intermediate grades. The grade awarded depends on the examiner's assessment of the candidate's performance within these parameters. Ten different examiners make 100 such judgments for each candidate in the course of the examination. A case score is calculated by allotting numerical values (4–0) to these grades and expressing their sum as a percentage of the maximum obtainable (20) for the case. The final examination score is the mean of the 20 case scores.

A candidate who was graded 'barely adequate' on every item would score 50%, and this was initially taken as the passmark. However, a new procedure for standard setting was introduced at the second sitting of the simulated surgery using a modified version of the 'contrasting groups' method.[5] At the end of the examination, the examiners are provided with printouts of each candidate's grades using the alphabetic symbols ('g', 's', 'b', 'n', and 'u') while remaining blind to the candidate's identity and examination score. They are asked to make an overall pass/fail judgement on the basis of the information given. Disregarding those candidates who are unanimously passed or failed, the number of passes for each of the remainder is plotted against their overall examination score. The data is 'smoothed' by drawing a regression line through these points and the passmark is taken to be the point on this line at which 50% of the judges would pass the candidate. This method is described in detail in another paper.[6]

Results

The 20-case simulated surgery has been held on five occasions, including three pilot and two live examinations. Two different selections of cases were used (Set 1 and Set 2). Seventy-five candidates were examined in the course of nine days as shown in Table 1.

The candidates in the pilot examinations were registrars from the local vocational training scheme courses, with various levels of experience in their general practice training year. The candidates for the live examinations were MRCGP candidates who fulfilled the eligibility criteria for the simulated surgery and had performed sufficiently well in the written papers to be called forward for oral examination. They included overseas candidates from European Union countries, the Middle East, and the Republic of Ireland, as well as United Kingdom candidates with valid reasons for being unable to make a videotape of their consultations.

The distribution of candidates' scores for the two case selections are shown graphically in Figures 1 and 2, and are summarized in Table 2.

The passmark for the first case selection was taken as 50%, which resulted in two of the 51 candidates failing, neither of them in the MRCGP cohort. The passmark for the second case selection was set by the examiners using the contrasting groups method at 58%, resulting in failure for five of the 24 candidates: two of them in the MRCGP cohort. Retrospective application of

Table 1. Information on the 75 candidates examined in the course of nine days.

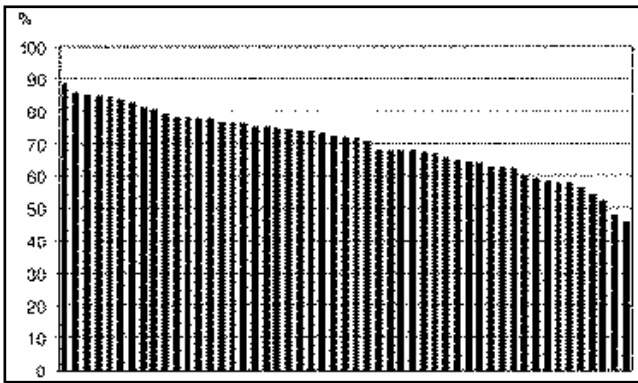| Date	Place	Status	Case set	Days	Candidates
October 1996	Wexham Park	Pilot	1	1	10
February 1997	CHIME - UCH	Pilot	1	2	18
June 1997	Princes Gate	Live	1	3	23
October 1997	Wexham Park	Pilot	2	1	10
December 1997	Princes Gate	Live	2	2	14

Figure 1. Set 1: Distribution of scores rank ordered for 51 candidates.
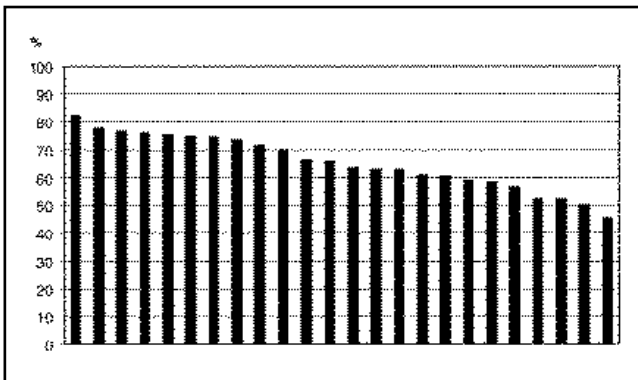

Figure 2. Set 2: Distribution of scores rank ordered for 24 candidates.

Table 2. Distribution of candidates' scores.

Case selection	Candidates (n)	Mean (%)	SD (%)	Range (%)
Set 1	51	69.9	10.2	45.5–88.3
Set 2	24	65.3	10.0	45.5–82.5

this standard to Set 1 would have failed eight candidates.

The reliability of the examination scores was tested by calculating Cronbach's alpha, which gave values of 0.85 for Set 1 and 0.88 for Set 2. The performance of individual cases in a set was monitored by calculating the mean score, standard deviation, and correlation coefficient of the candidates' scores in the individual case with their mean scores in the remaining cases. These were taken respectively as indications of their level of difficulty, degree of discrimination, and consistency with other cases. This information was used in selecting some cases from Set 1, which were known to perform well, for inclusion in Set 2.

Discussion

A satisfactory spread of scores was achieved among the 75 candidates, indicating that the simulated surgery is able to discriminate well between good and poor performance in the examination. The validity of these scores as a measure of competence in consulting skills rests on the combined experience of the case writers and the judgements of the observing examiners. The reliability of the 20-case simulated surgery is good; an alpha coefficient of 0.8 being the generally accepted requirement for a high stakes examination. The alphas of 0.85 (Set 1) and 0.88 (Set 2)

gave us grounds to consider a possible reduction in the number of cases needed to fulfil this requirement.

In our first administration of the simulated surgery, we adopted a passmark of 50%. This was predicated on the notion that a hypothetical candidate who was graded 'barely adequate' on every item in every case would score 50%. It resulted in all the candidates passing the first live round of the examination ($n = 23$; mean score = 70.07%; standard deviation = 8.97%; range 51.75–88.25%). The examiners felt that a number of the candidates scoring below 60% should not have passed. We therefore adopted the modified 'contrasting groups' method to set the passmark for the next round of the examination that yielded a passmark of 58.2%. This resulted in failure of two out of 14 candidates in the second live examination. No inference can be drawn about the passing standard from the failure rate of this cohort because they are not typical of the overall group of candidates taking MRCGP.

The simulated surgery has many strengths in the assessment of consulting skills:

- The examination is standardized so that every candidate is assessed on the same cases.
- The surgery is blueprinted to cover a range of contexts and the cases are selected to test specific skills.
- The content can include sensitive issues where consent for video-recording might be withheld by a real patient, also the management of uncommon but important conditions that are unlikely to be seen on a submitted videotape.
- Physical examination is among the skills that can be examined.
- Bias is minimized by the multi-examiner assessment of each candidate, and scores can be used to discriminate at different levels, including MRCGP pass and merit.

The validity of simulation has not proved to be a problem for most candidates, although some have mentioned the artificiality of a fixed consultation length. Most report that it feels like 'real' general practice, although the format prevents us from including children, emergencies, physical signs, and previously known patients. Challenges for the organizers have included ensuring a constant standard between sittings of the simulated surgery and ensuring an equivalent standard to the video method; strict comparability is not possible since the simulated surgery measures competence ('can do'), while the video measures performance ('does do').

The major problems of the simulated surgery for the MRCGP are manpower, time, and cost: in the current format, 10 examiners and 10 role-players are required for one day for every 10 candidates. The video consulting skills component requires only two examiners for one day to examine 10 candidates. The cost of running the 20-case surgery is about £650 per candidate. This is calculated on the basis of 20 candidates examined over two days:

Eleven examiners/marshals (travel and locum fees @ £300/day)	£6600
Eleven role-players/administration staff (travel and fees @ £200/day)	£4400
Administration and venue costs	£2000
Total (excluding development costs)	£13000

While these costs are comparable to those of standardized patient examinations in other countries (Australia, Canada, Holland, and USA), they are not supportable for more than a small number of candidates in the MRCGP.

The MRCGP examination was redesigned on a modular basis at the beginning of 1998. The individual components — Paper I,

Paper II, Consulting Skills, and Orals — can now be sat separately at different times, and each must be passed to achieve a pass at MRCGP. This has had implications for the simulated surgery: it must examine its allotted area of the examination blueprint comprehensively, and the scores must be sufficiently reliable to make an independent pass/fail decision for the module. It has needed to accommodate more candidates, since the cohort who apply are no longer curtailed by some candidates not performing well enough on the written papers to be called forward to orals and the simulated surgery. At the same time, the cost and manpower requirements of the 20-case format were such that it could not be made more widely available in its previous form. The solution has been to reduce the number of cases examined to 12. This allows the full surgery to be run twice in a day, thereby examining up to 24 candidates with 12 examiners and role-players. Two parallel circuits are run, thus increasing the capacity of the examination to 48 candidates a day, with a consequent reduction in cost to around £350 per candidate.

Forty MRCGP candidates were examined in July 1998 using a selection of cases from Set 2. The mean score was 64.34% (SD = 10.55%; range 88.3–37.9%). The examiners set a nearly identical passmark of 58.0% and a merit cutting score of 78%. Twelve candidates failed and 28 passed, six with merit. The alpha coefficient for this examination was 0.83.

Conclusion

We conclude that the simulated surgery is established as a feasible, valid, and reliable method of assessing the clinical skills of MRCGP candidates who are unable to submit a video-recording of their consultations. It has been successfully used to examine 75 candidates in the first year, and a means of transferring the passing standard between successive sittings of the examination has been developed. There are problems of cost and manpower with the 20-case format; however, the high level of reliability achieved has enabled us to reduce the number of cases to 12 in 1998. This has made it possible to increase the number of circuits in a day and to meet the requirements of the new modular MRCGP.

References:

1. Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ* 1979; **13:** 39-54.
2. Bingham E, Burrows PJ, Caird GR, *et al*. Simulated surgery: a framework for the assessment of clinical competence. *Educ Gen Pract* 1994; **5:** 143-150.
3. Bingham E, Burrows PJ, Caird GR, *et al*. Simulated surgery - using standardized patients to assess the clinical competence of GP registrars - a potential clinical component for the MRCGP examination. *Educ Gen Pract* 1996; **7:** 102-111.
4. Burrows PJ. Consulting skills assessment 2: by simulated surgery. In: Moore R (ed). *The MRCGP Examination: a guide for candidates and teachers.* [Third edition.] London: Royal College of General Practitioners, 1998.
5. Livingston SA, Zieky MJ. *Passing scores: a manual for setting standards of performance on educational and occupational tests.* Princeton, NJ: Educational Testing Service, 1982.
6. Burrows PJ, Bingham E, Brailovsky CA. A modified 'contrasting groups' method used for setting the passmark for a small scale standardized patient examination. *Advances in Health Sciences Education* 1999 (in press).

Address for correspondence:

Dr Peter J Burrows, Maurys Mount, Slab Lane, West Wellow, Romsey, Hants SO51 6BY.