

## Evaluation of three short-listing methodologies for selection into postgraduate training in general practice

Fiona Patterson,<sup>1,2</sup> Helen Baron,<sup>2</sup> Victoria Carr,<sup>2</sup> Simon Plint<sup>3</sup> & Pat Lane<sup>4</sup>

**OBJECTIVE** This study aimed to evaluate the effectiveness and efficiency of three short-listing methodologies for use in selecting trainees into postgraduate training in general practice in the UK.

**METHODS** This was an exploratory study designed to compare three short-listing methodologies. Two methodologies – a clinical problem-solving test (CPST) and structured application form questions (AFQs) – were already in use for selection purposes. The third, a new situational judgement test (SJT), was evaluated alongside the live selection process. An evaluation was conducted on a sample of 463 applicants for training posts in UK general practice. Applicants completed all three assessments and attended a selection centre that used work-related simulations at final stage selection. Applicant scores on each short-listing methodology were compared with scores at the selection centre.

**RESULTS** Results indicate the structured AFQs, CPST and SJT were all valid short-listing methodologies. The SJT was the most effective independent predictor. Both the structured AFQs and the SJT add incremental validity over the use of the CPST alone. Results show that optimum validity and efficiency is achieved using a combination of the CPST and SJT.

**CONCLUSIONS** A combination of the CPST and SJT represents the most effective and efficient battery of instruments as, unlike AFQs, these tests are machine-marked. Importantly, this is the first study to evaluate a machine-marked SJT to assess non-clinical domains for postgraduate selection. Future research should explore links with work-based assessment once trainees are in post to address long-term predictive validity.

**KEYWORDS** \*education, medical, graduate; family practice/\*education; Great Britain; \*school admission criteria; evaluation studies [publication type]; problem-based learning; questionnaires; judgement.

*Medical Education* 2009; **43**: 50–57

doi:10.1111/j.1365-2923.2008.03238.x

<sup>1</sup>Department of Psychology, City University, London, UK

<sup>2</sup>Work Psychology Group, Nottingham, UK

<sup>3</sup>GP Dean & Head of the School of Primary Care NHS Education South Central Oxford Deanery, UK

<sup>4</sup>Formerly, Director of Postgraduate General Practice Education, Yorkshire and the Humber Postgraduate Deanery, Sheffield, UK

*Correspondence:* Professor Fiona Patterson, c/o Executive Assistant Sarah Marshall, Department of Psychology, City University, Northampton Square, London EC1V 0HB, UK.

Tel: 00 44 115 940 7978; Fax: 00 44 115 940 3000;

E-mail: f.patterson@city.ac.uk

---

 INTRODUCTION

Unprecedented changes in postgraduate medical training in the UK, as a result of the Modernising Medical Careers initiative,<sup>1</sup> have placed more emphasis on the methodologies used for selecting doctors into specialty training.<sup>2,3</sup> Almost all large-scale recruitment uses a multi-stage process to progressively reduce the applicant pool.<sup>4</sup> This paper compares the effectiveness and efficiency of three methods used to short-list applicants for selection into postgraduate training. Specifically, this paper compares the effectiveness of a machine-marked situational judgement test (SJT) targeting non-clinical domains with the effectiveness of traditional selection methodologies. The results will directly inform optimal selection practices for postgraduate training across all medical specialties, which has been a topic of fierce debate. We present these data to encourage further analysis within the scientific community and to develop a future evidence-based research agenda for selection methodology.

As with any selection methodology, various psychometric and legal criteria must be satisfied, including those pertaining to standardisation, reliability, validity and fairness.<sup>3,5</sup> The method must discriminate fairly between applicants and ensure few 'false negative' decisions (i.e. where applicants who are potentially successful at interview are rejected). Processes must be cost-efficient, especially with large numbers of applicants, where recruiter time and administrative efficiency are important. The methodology should be acceptable to applicants so that procedural fairness is maintained.<sup>5</sup>

### Recruitment into general practice in the UK

This study evaluates three short-listing methodologies for selection into postgraduate training in general practice (GP) in the UK. The selection system is designed to process several thousand applicants per year<sup>6,7</sup> and the methodology comprises three stages. Stage 1 consists of long-listing eligibility checks. Stage 2 comprises short-listing via, firstly, a clinical problem-solving test (CPST), a machine-marked test developed from an existing item bank in which the applicant applies clinical knowledge to solve a problem reflecting a diagnostic process or to develop a management strategy for a patient, and, secondly, structured application form questions (AFQs) comprising open-ended questions targeting non-clinical domains in the person

specification, such as empathy or integrity, to which applicants provide a word-limited response.<sup>8</sup> Once short-listed, applicants attend Stage 3, a selection centre (SC) using work-relevant simulations (patient consultation, group and written exercises) to target both clinical and non-clinical domains from the person specification. This methodology has been validated.<sup>7</sup> However, the short-listing process (Stage 2) is relatively resource-intensive, especially with large-volume recruitment.

Typically, 5–10% of applicants are rejected at long-listing and a further 20–30% are selected out at short-listing. This paper reports on the evaluation of a new short-listing methodology, a situational judgement test (SJT) targeting non-clinical domains, where applicants are presented with written depictions of scenarios they may encounter at work and asked to identify an appropriate response from a list of alternatives. The evaluation considered whether this new approach could improve efficiency by replacing the hand-marked AFQs (which require 30 minutes of assessor time per applicant in addition to invigilation) with a machine-marked measure, using a previously agreed scoring frame.

Recent reviews indicate that the predictive and incremental validity of SJTs is well established.<sup>9–13</sup> Although the SJT methodology has been validated for use in medical school admissions,<sup>14</sup> this is, to the authors' knowledge, its first application in postgraduate specialty selection. Here, the SJT focused on three non-clinical selection criteria: empathy, integrity and coping with pressure. Table 1 provides two example items and illustrates different response formats (ranking and multiple best answer).

Table 2 provides a summary of the content, scoring process and related properties of each of the three short-listing methodologies. All are written tests, completed independently and under invigilated conditions. By comparing the predictive validity and resource requirements of each, the purpose was to identify the most effective methodology for future use.

As the CPST and AFQs were currently used in selection, the SJT was piloted alongside them. Specifically, this paper addresses the following questions.

- 1 The machine-marked CPST is efficient to use. What validity does it offer compared with other short-listing methodologies?

Table 1 Example situational judgement test items using ranking and multiple best-answer response formats

1 **A 55-year-old woman with ischaemic heart disease has smoked 20 cigarettes per day for 40 years. She requests nicotine replacement patches. She has had these previously but has been inconsistent in their use and has often continued to smoke while using the patches.**

**Rank in order the following immediate actions in response to this situation (1 = most appropriate; 5 = least appropriate)**

- A *Emphasise the dangers of smoking but do not prescribe*
- B *Enquire about the difficulties she has with stopping smoking and any previous problems with patches*
- C *Insist on a period of abstinence before prescribing any further patches*
- D *Prescribe another supply of patches and explain how they should be used*
- E *Suggest that nicotine replacement therapy is not suitable for her but explore alternative therapies*

2 **You are looking after Mrs Sandra Jones, who is being investigated in hospital. You are asked by her family not to inform Mrs Jones if the results confirm cancer.**

**Choose the THREE most appropriate actions to take in this situation**

- A *Ignore the family's wishes*
- B *Agree not to tell Mrs Jones*
- C *Explain to the family that it is Mrs Jones' decision*
- D *Ask Mrs Jones whether she wishes to know the test results*
- E *Ask Mrs Jones whether she wishes you to inform the family*
- F *Inform Mrs Jones that her family do not wish her to have the results*
- G *Give the results to the family first*
- H *Give the results to the next of kin first*

- 2 Structured AFQs are relatively time-consuming to score and require trained assessors. Do they add incremental validity over other short-listing measures?
- 3 The SJT is a newly designed, machine-marked methodology in this context. Is it valid and does it have incremental validity over the other measures?
- 4 Which combination of the three measures represents the most effective short-listing methodology?

## METHODS

### Design and procedure

Data were collected during the 2006 selection round for GP training in the UK. At short-listing stage, applicants completed one of four pilot forms of the SJT together with operational assessments consisting of the CPST and AFQs. Applicants gave full consent to participation in the SJT as this was for evaluation purposes only and scores were not used in selection decisions. The SJT was completed in advance of the CPST and AFQs.

Applicants who were successful in being shortlisted were invited to the SC, on the basis of which job offers were made. Performance at the SC was used as an outcome measure, using the mean scores across three simulation exercises, as a primary aim of short-listing is to identify applicants who are likely to perform well at the SC. Performance at the SC has been shown to predict performance 3 months into GP training.<sup>7</sup> The regression for SC results on scores at short-listing from the CPST, AFQs and SJT was examined to establish predictive validity. Hierarchical regression analysis was used to explore incremental validity.

### Psychometric properties of the instruments

The reliability of each assessment methodology was evaluated using Cronbach's  $\alpha$  coefficient and the Spearman-Brown formula where appropriate. The CPST, AFQs and SC show satisfactory levels of reliability (CPST  $\alpha = 0.89$ ; AFQs  $\alpha = 0.78$ ; SC  $\alpha = 0.89$ ). As the SJT was a pilot rather than an operational test, some items did not meet required psychometric standards and were not subsequently included in the bank. Results show 71% of items had sufficient psychometric quality to be included in an operational test. The Spearman-Brown formula was used to estimate the  $\alpha$  reliability of operational length SJT forms from the items of acceptable quality in each pilot form; reliability ranged from 0.80 to 0.83. In preparing SJT items, 10 experienced item writers from the examiner panel for Membership of the Royal College of General Practitioners (MRCGP) in the UK, with no previous involvement in the SJT development process, each responded to two pilot forms in order to provide a five-person concordance sample for the response key. This concordance analysis was undertaken to ensure experts were in

Table 2 Overview of three short-listing methodologies

	Clinical problem-solving test (CPST)	Structured application form questions (AFQs)	Situational judgement test (SJT)
Item content and response format	The CPST evolved from an existing test of clinical knowledge developed by the North Western Region in the UK. Operational papers have around 100 questions and a completion time of 90 minutes. Each paper contains a mix of extended-match and single best-answer formats covering the range of clinical areas defined by the UK training curriculum.	Seven AFQ items are drawn from an existing item bank. Each question targets a single non-clinical competency domain, including empathy, integrity and ability to cope with pressure. An example question is: ' <i>Give an example of how you won the trust of a worried and sceptical patient. Describe what you did, and why, and the effect it had on both you and the patient.</i> ' Applicants are allowed 2 hours to complete seven questions, with a 250-word limit per question. The assessment is completed under invigilated conditions.	The SJT is a newly designed instrument consisting of 50 questions of two different types (rating and multiple best-answer formats; see Table 1). Twenty senior general practitioners (four female, one from a minority ethnic background), with experience in both training and designing assessments, worked with three psychologists to generate, review and refine a bank of questions. Questions target three non-clinical domains: empathy, professional integrity and ability to cope with pressure. 186 items were written and divided into four pilot test forms of 50 items each (14 items were repeated in a second test form).
Scoring process	Papers are machine-marked according to the agreed key.	Each paper is scored independently by two trained assessors working from a validated scoring framework. Assessors are trained during a 1-day workshop and calibrated to enhance reliability.	Papers are machine-marked according to the agreed key.

agreement over the keyed response to each item. Kendall's *W* was computed for each ranking item, which showed 85% had a concordance over 0.6 and 71% above 0.7. This indicates substantial and significant levels of agreement between experts. Items with poor concordance were reviewed by item writers after piloting and only included in the bank if consensus was reached on the response key.

### Sample

This study is based on a total of 524 doctors who applied for GP training programmes in the UK commencing in 2007, who attended an SC as

part of final-stage selection. Of those, 463 applicants had attended short-listing at a location which piloted the SJT alongside the CPST and AFQs. Situational judgement test data were not available for 12% of applicants because one pilot paper was not used as a result of an administrative error on the paper. Applicants were short-listed via the CPST and AFQs. Of the 463 applicants, 178 were males and 275 were female (10 unknown). Overall, 39.0% described themselves as White (UK and Ireland), 46.0% as Asian, 4.6% as Black and 10.4% as belonging to other ethnic groups (European, Arab etc). Approximately 58% of the sample was aged  $\leq 30$  years; 37% were aged 31–40 years and 5% were  $> 40$  years.

## RESULTS

All three short-listing methodologies and the SC showed score distributions close to normal with effective levels of variability (Table 3). There were significant correlations between scores on each of the short-listing methodologies. The correlation between the AFQs and the SJT was  $r = 0.41$  ( $P < 0.001$ ), which is consistent with the similarity of the non-clinical constructs measured by these two instruments (empathy, integrity, coping with pressure). The CPST and SJT also correlated positively ( $r = 0.39$ ,  $P < 0.001$ ). This may reflect the common closed-response format of both these measures (as opposed to the open-response format of the AFQs). The correlation between the CPST and AFQs was  $r = 0.32$  ( $P < 0.001$ ), reflecting the differences in content (clinical versus non-clinical domains) and response method (closed versus open). In summary, this level of correlation suggests that each pair of instruments have both common and independent variance.

In analysing associations between each short-listing methodology and SC scores, both uncorrected correlations and coefficients corrected for restriction of range are reported.<sup>15</sup> In this context, uncorrected correlations underestimate 'true' relationships between predictors and criteria when, as in this case, the predictors are used to select applicants to be included in the sample (to attend the SC) and thus restrict the variation in scores. The uncorrected correlation between AFQs and the SC was  $r = 0.26$  ( $P < 0.001$ ), indicating moderate predictive validity for a selection methodology. Uncorrected correlations were  $r = 0.30$  ( $P < 0.001$ ) for the CPST and SC, and  $r = 0.46$  ( $P < 0.001$ ) for the SJT and SC. The AFQs and CPST were then

corrected for direct restriction of range caused by their use in short-listing. The SJT was corrected for the indirect restriction of range resulting from the correlation of the SJT with the operational short-listing measures (i.e. AFQs and CPST). When corrected for restriction of range, the correlations between each short-listing measure and the SC increased to  $r = 0.40$  (AFQs),  $r = 0.44$  (CPST) and  $r = 0.56$  (SJT). A comparison of short-listing methodologies revealed the SJT to show the strongest association with SC performance and AFQs to show the weakest association, although it was still moderately strong.

Results demonstrate that the strongest prediction of SC performance is a combination of all three measures (uncorrected  $r = 0.51$ ,  $P < 0.001$ ). Hierarchical regression analyses showed all three measures to have significant incremental validity over all other measures ( $P < 0.001$ ) (Table 4). The SJT is the best single predictor (adjusted  $R^2 = 0.213$ ,  $P < 0.001$ ) and offers the most incremental validity over other methodologies. The predictive validity of the SJT is superior to that of both the CPST and AFQs, but all three are effective predictors. This has important implications for the development of selection methodologies in this context, especially in the assessment of non-clinical domains.

Resource-efficiency is an important consideration in developing selection methodologies.<sup>2,16</sup> In terms of cost comparison and efficiency between the methodologies, a utility analysis of the GP recruitment process (including analysis of assessor time and administration costs) produced estimated costs of £50 (€59, \$74) per applicant using the AFQ method compared with actual costs of £20 (€24, \$30) per applicant using a machine-marked test.

Table 3 Summary of descriptives for all measures in the study

<i>n</i> = 463	Selection centre	Application form questions	Clinical problem-solving test	Situational judgement test
Mean	3.15	3.07	82.35	0.71
Standard deviation	0.46	0.26	7.5	0.57
Min, max	1.50, 4.00	2.1, 3.7	64, 98	- 1.24, 1.97

Situational judgement test scores were standardised within the trial group in order to combine scores across the different trial forms

Table 4 Predictive validity results for the clinical problem-solving test, structured application form questions and situational judgement test

Predictors	Adjusted R <sup>2</sup>	R <sup>2</sup> change	F (d.f. 1, d.f. 2)*
CPST	0.090		46.85 (1,461)
AFQs	0.066		33.43 (1,461)
SJT	0.213		125.74 (1,461)
CPST and AFQs	0.138		37.93 (2,460)
Incremental prediction of CPST over AFQs		0.074	39.62 (1,460)
Incremental prediction of AFQs over CPST		0.049	26.42 (1,460)
CPST and SJT	0.232		70.81 (2,460)
Incremental prediction of CPST over SJT		0.021	12.69 (1,460)
Incremental prediction of SJT over CPST		0.143	86.12 (1,460)
AFQs and SJT	0.235		71.87 (2,460)
Incremental prediction of SJT over AFQs		0.170	102.91 (1,460)
Incremental prediction of AFQs over SJT		0.024	14.35 (1,460)
CPST, AFQs and SJT	0.252		52.94 (3,459)
Incremental prediction of AFQs over CPST and SJT		0.022	13.38 (1,459)

Performance at the selection centre was used as the criterion variable, using the mean score across three simulation exercises in the selection centre  
 \* All values were significant at  $P < 0.001$   
 CPST = clinical problem-solving test; AFQs = application form questions; SJT = situational judgement test

## DISCUSSION

### Main findings

All three short-listing methodologies independently show sufficient predictive validity for selection purposes. Corrected correlations over  $r = 0.3$  between predictor and criterion indicate a moderately strong level of prediction.<sup>17</sup> Although there is scope for improvement, the existing selection process using the CPST and AFQs is an effective methodology. The CPST focuses on clinical knowledge and problem-solving ability and the AFQs focus on non-clinical domains, including empathy, integrity and ability to cope with pressure. The results show that AFQs add significant incremental validity over the CPST in predicting SC outcome. This might be expected as AFQs target non-clinical domains, whereas the CPST focuses on clinical problem-solving skills. The SJT shows the strongest validity in predicting SC performance. This concurs with other studies using SJTs in college admissions procedures, where this methodology is best suited to assessment of non-clinical domains.<sup>14</sup> However, this is the first application of an SJT in postgraduate selection and it demonstrates that the methodology is effective in this context.

Although each methodology independently predicted later SC performance, the most accurate overall prediction was obtained using the newly designed SJT in combination with the other measures.

### Interpretation of results

The degree of predictive validity is an important evaluative standard for any selection procedure. However, although sufficient validity is achieved, the AFQ method is relatively costly to implement, especially for large-volume recruitment such as this, as it requires substantial marking resources (two trained assessors score each response independently) in addition to invigilation. The SJT provides similar measurement properties to the AFQ process but does not carry the associated marking costs. When considering efficiency in addition to predictive validity, results show the optimum short-listing methodology to be a combination of the CPST and SJT, as both are machine-marked. It must, however, be acknowledged that SJTs (like CPSTs) require investment in development costs and, because they are machine-marked, are most cost-beneficial when there are large applicant numbers.

## Limitations

Although the validity of the three short-listing measures in predicting performance in the final stage of selection has been clearly demonstrated, further research is needed to examine whether the measures also predict performance in training. This is currently being explored in a separate validation study. One limitation refers to the loss of some data (12%) for administrative reasons. However, the sample size remains adequate and there is no evidence to suggest a systematic bias in the data lost. As the SJT was a pilot and candidates were aware that scores would not contribute to selection decisions, it could be argued that candidates may not have addressed this paper with the same degree of seriousness, which might have attenuated its observed reliability and, therefore, validity. Additionally, because the SJT was a pilot paper rather than an operational test, the validity coefficients may, again, be attenuated.

---

## CONCLUSIONS

The results have important implications for developing selection systems for large-volume recruitment to optimise both effectiveness and efficiency. Although there are limitations to this study, as outlined above, the evaluation of a machine-marked SJT to assess non-clinical domains is an important innovation in this context. This is relevant to current proposals for selection into UK postgraduate training for all medical specialties regarding the use of machine-marked tests for short-listing purposes.<sup>18</sup> However, recent evidence shows although there are common competencies across specialties, there are also differing priorities in the selection criteria, especially in non-clinical domains.<sup>19</sup> Research is required to examine whether an SJT is an appropriate methodology for use in other specialties. Future research should explore applicant reactions and perceptions of fairness<sup>5,20</sup> and other crucial evaluative standards in judging the quality of selection methodologies. In addition, further research should explore whether assessments are prone to coaching or practice effects.<sup>14</sup> Results at selection should be linked with other work-based assessment methodologies once trainees are in post to explore predictive validity in the long-term.

---

*Contributors:* FP contributed to the original conception and design of the study, data collection, analysis and interpretation, and the write-up of the paper. HB and VC contributed to the overall study design, data collection,

analysis and interpretation, and the write-up of the paper. PL and SP contributed to the conception of the original study, and data collection and interpretation, and commented on the paper. All authors approved the final manuscript for publication.

*Acknowledgements:* Gai Evans is acknowledged for her contribution to data collection via the General Practice National Recruitment Office. Our gratitude to the various contributors to item writing for all three short-listing methodologies is acknowledged.

*Funding:* this research was funded by the Committee of General Practice Education Directors (COGPED) in the UK via the National Recruitment Office.

*Conflicts of interest:* FP, HB and VC are employed by Work Psychology Ltd. This organisation has advised the UK Department of Health on selection and recruitment issues. SP has subsequently become the GP Adviser in the DH MMC team.

*Ethical approval:* this study was approved by City University, London.

---

## REFERENCES

- 1 *Modernising Medical Careers*. London: Department of Health. <http://www.mmc.nhs.uk/>. [Accessed 2 April 2008.]
- 2 British Medical Association Board of Medical Education. *Selection for Specialty Training*. London: BMA 2006. <http://www.bma.org.uk/ap.nsf/Content/selectionforspecialtytraining>. [Accessed 2 April 2008.]
- 3 Patterson F, Lane P. Assessment for recruitment. In: Jackson N, Jamieson F, Khan A, eds. *Assessment in Medical Education and Training*. Oxford: Radcliffe Publishing 2007;62–73.
- 4 Arnold J, Silvester J, Patterson F, Robertson I, Cooper C, Burnes B. *Work Psychology: Understanding Human Behaviour in the Workplace*, 4th edn. Harlow: Financial Times Prentice Hall 2004.
- 5 Patterson F, Ferguson E. *Selection for Medical Education and Training. Understanding Medical Education Series*. Edinburgh: Association for the Study of Medical Education 2007.
- 6 Plint S, Gregory S, Evans G. Recruitment to GP specialty training. *BMJ Career Focus* 2007;**335**:gp73–5.
- 7 Patterson F, Ferguson E, Norfolk T, Lane P. A new selection system to recruit general practice registrars: preliminary findings from a validation study. *BMJ* 2005;**330**:711–4.
- 8 Lamsdale C, Wood R, Mulrooney C. An alternative to an assessment centre on six pieces of paper? *Int J Select Assess* 1999;**7** (3):170–6.
- 9 Chan D, Schmitt N. Situational judgement and job performance. *Hum Perform* 2002;**15**:233–54.
- 10 Rahman M. Tackling situational judgement tests. *BMJ Career Focus* 2007;**334**:gp189–90.
- 11 McDaniel MA, Morgeson FP, Finnegan EB, Campion MA, Braverman EP. Use of situational judgement tests

- to predict job performance: a clarification of the literature. *J Appl Psychol* 2001;**86**:730–40.
- 12 Chan D, Schmitt N. Situational judgement tests. In: Anderson N, Evers A, Voskuijl O, eds. *Blackwell Handbook of Selection*. Oxford: Blackwell Publishing 2005;219–42.
  - 13 McDaniel MA, Nguyen NT. Situational judgement tests: a review of practice and constructs assessed. *Int J Select Assess* 2001;**9**:103–13.
  - 14 Lievens F, Buyse T, Sackett PR. The operational validity of a video-based situational judgement test for medical college admissions: illustrating the importance of matching predictor and criterion construct domains. *J Appl Psychol* 2005;**90**:442–52.
  - 15 Mendoza JL, Mumford M. Corrections for attenuation and range restriction on the predictor. *J Educ Statist* 1987;**12**:282–93.
  - 16 Boudreau JW. Selection utility analysis: a review and agenda for future research. In: Smith M, Robertson IT, eds. *Advances in Selection and Assessment*. London: John Wiley & Sons 1989;227–58.
  - 17 Murphy KR, Bartram D. Recruitment, personnel selection and organisational effectiveness. In: Robertson IT, Callinan M, Bartram D, eds. *Organizational Effectiveness: the Role of Psychology*. New York, NY: John Wiley and Sons Ltd 2002;85–114.
  - 18 Tooke J. *Aspiring to Excellence: Findings and Recommendations of the Independent Inquiry into Modernising Medical Careers*. London: MMC Inquiry 2007. <http://www.mmcinquiry.org.uk/draft.htm>. [Accessed 2 April 2008.]
  - 19 Patterson F, Ferguson E, Thomas S. Using job analysis to identify core and specific selection criteria for three secondary care specialties. *Med Educ* 2008;**42**:1195–1204.
  - 20 Plint S, Patterson F, Lane P. Developing a national selection process for UK GP specialty training – a model for success? *J Postgrad Med*; Submitted.

*Received 27 November 2007; editorial comments to authors 6 February 2008, 29 April 2008; accepted for publication 28 May 2008*