Patients' reports or clinicians' assessments: which are better for prognosticating?

Paddy Stone,[1] Bridget Gwilliam,[1] Vaughan Keeley,[2] Chris Todd,[3] Matthew Gittins,[4] Laura Kelly,[5] Stephen Barclay,[6] Chris Roberts[4]

[1]Division of Population, Health Sciences and Education, St George's University of London, London, UK
[2]Department of Palliative Care, Royal Derby Hospital, Derby, UK
[3]School of Nursing, Midwifery & Social Work, Jean McFarlane Building, The University of Manchester, Manchester, UK
[4]Health Sciences, School of Community Based Medicine, Jean McFarlane Building, The University of Manchester, Manchester, UK
[5]Department of Palliative Medicine, Surrey and Sussex Healthcare NHS Trust, Surrey, UK
[6]General Practice and Primary Care Research Unit, University of Cambridge, Institute of Public Health, Cambridge, UK

Correspondence to
Paddy Stone, Division of Population, Health Sciences and Education, St George's University of London, London SW17 ORE, UK; pstone@sgul.ac.uk

Accepted 24 April 2012
Published Online First 31 May 2012

ABSTRACT

Background The Prognosis in Palliative care Scale (PiPS) predicts survival in advanced cancer patients more accurately than a doctor's or a nurse's estimate. PiPS scores are derived using observer ratings of symptom severity and performance status. The purpose of this study was to determine whether patient-rated data would provide better prognostic estimates than clinician observer ratings.

Patients and methods 1018 subjects with advanced cancer no longer undergoing tumour-directed therapy were recruited to a multi-centre study. Prognostic models were developed using observer ratings, patient ratings or a composite method that used patient ratings when available or else used observer ratings. The performance of the prognostic models was compared by determining the agreement between the models' predictions and the survival of study participants.

Results All three approaches to model development resulted in prognostic scores that were able to differentiate between patients with a survival of 'days', 'weeks' or 'months+'. However, the observer-rated models were significantly ($p<0.05$) more accurate than the patient-rated models.

Conclusions A prognostic model derived using observer-rated data was more accurate at predicting survival than a similar model derived using patient self-report measures. This is clinically important because patient-rated data can be burdensome and difficult to obtain in patients with terminal illnesses.

INTRODUCTION

Terminally ill patients, their families and healthcare providers want sufficient information to allow them to adequately 'prepare' for the end of their lives.[1] This includes a desire to know about when they will die, and how they could expect their physical condition to deteriorate. Unfortunately, clinicians are poor at predicting survival in patients with advanced cancer.[2] Prognostic questions arise on a day-to-day basis in clinical practice and any method that is superior to (or independent of) clinician estimates would be highly valued.

The Prognosis in Palliative care Scale (PiPS) was developed as a result of a large prospective multi-centre study involving over 1000 patients with advanced cancer.[3] PiPS can be used in both competent and incompetent patients and is able to predict survival in terms of 'days', 'weeks' or 'months+'. Alternate forms of the scales are available for use in patients in whom blood tests are available (PiPS-B) and in patients in whom such

results are unavailable (PiPS-A). Both scales are at least as accurate as a multi-professional estimate of survival and PiPS-B is significantly better than an individual doctor's or nurse's prediction. Both scales use clinician proxy ratings about the presence or absence of key symptoms, level of physical functioning and overall health of the patient.

However, there are limitations to using proxy ratings for subjective measures and as part of the PiPS study we wanted to determine whether or not clinician ratings could be substituted for patient reported outcomes without adversely affecting the accuracy of the scales. On a practical level, this is clearly important because many palliative care patients are incompetent and a prognostic scale that was only reliable when completed by a competent patient would have limited clinical utility. We therefore undertook a planned subanalysis of the PiPS data set to determine whether patient reports or clinician estimates were superior in terms of prognostic accuracy.

PATIENTS AND METHODS

A detailed report of the methods used in the PiPS study has been published elsewhere.[3] Briefly, consecutive referrals to participating palliative care units were screened for inclusion. Eligible patients had locally advanced or metastatic cancer and were no longer undergoing disease-modifying treatment. Both competent and incompetent patients were recruited. Competence to participate in the research was assessed by the attending clinician using Department of Health guidance.[4] Competent patients gave written informed consent and the assent of carers was obtained for incompetent patients. A core data set was collected on all participants and an extended data set was collected from competent patients.

Core data set

The researcher collected data on the presence or absence of the following symptoms: pain, breathlessness at rest, loss of appetite, dry mouth, difficulty swallowing and tiredness. Performance status was assessed using the Eastern Cooperative Oncology Group (ECOG)[5] score. ECOG scores vary between 0 and 4 (0 = normal functional abilities, 4 = confined to a bed or chair and requires all care). Global health status was recorded using a study-specific 7-point scale (1 = extremely poor health, 7 = normal health). Other information collected by the research team was obtained directly from patients' notes (eg, demographic and disease-related variables)

or from direct observation (eg, pulse rate or the presence of peripheral oedema).

Extended data set

Competent patients were asked to provide self-reports on all of the proxy domains assessed by clinicians. Thus, paired data (patient and clinician) were available for key symptoms, performance status and global health score.

Other assessments

Nurse, doctor and multi-professional estimates of survival were also obtained along with patient-reported estimates of survival. These results are presented elsewhere. Participants were also 'flagged' with the National Health Service Information Centre, so that the research team was informed when the patient died.

Statistical methods

In order to create a prognostic score that could estimate survival in terms of 'days', 'weeks' or 'months', it was necessary to develop two separate prognostic models, one model to predict survival up to 2 weeks (14 days) and another model to predict survival up to 2 months (56 days). These two models could then be combined to predict whether a patient would survive for 'days' (ie, less than 2 weeks), 'weeks' (between 2 weeks and 2 months) or 'months+' (ie, 2 months or greater).

The original PiPS observer-rated models (PiPS/OR) were developed using backward stepwise logistic regression. In order to correct for over-optimism during model development, the models were further evaluated using a bootstrap technique.[6][7] The PiPS/OR-14 models were used to predict survival up to 14 days both in patients in whom blood results were available (PiPS/OR-B14) and in those in whom blood results were not available (PiPS/OR-A14). Separate models were also created to predict survival up to 56 days (PiPS/OR-A56 and PiPS/OR-B56, respectively). The performance of each of these models was assessed by calculating the area under the receiver operating characteristic curve (AUC). Finally the models were combined to provide an estimated prognosis in one of three distinct categories; 'days' (<14 days), 'weeks' (14–55 days) and 'months+' (>55 days). The performance of these combined models was assessed by calculating the absolute agreement between the model prediction and actual survival and by use of the linear-weighted κ-statistic which gives an average measure of chance corrected agreement for the two thresholds.[8] The relative prognostic value of the models was compared using the AUC statistic, the linear weighted κ and the level of absolute agreement between model predictions and actual survival.

Two further groups of models were also created. The PiPS patient-rated models (PiPS/PR) were developed by using the self-report data from competent patients. The PiPS composite-scoring models (PiPS/CM) were created using a substitution rule. When patient generated data were available then this was used; however, when patient-rated data were missing then observer-rated data were used instead. As with PiPS/OR, both 14-day and 56-day survival models were created and a combined categorical model was produced ('days', 'weeks' and 'months+').

Quadratic weighted κ was used to evaluate the concordance between patient-rated and observer-rated symptom scores. Although weighted κ scores are the preferred method of comparing the concordance between predicted survival and actual survival, they do not lend themselves easily to an intuitive

interpretation of meaning. Landis and Koch[9] have proposed the following guide for interpreting the κ coefficient: ≤0 = poor, 0.0–0.20 = slight, 0.21–0.40 = fair, 0.41–0.60 = moderate, 0.61–0.80 = substantial and 0.8–1.0 = almost perfect. However, these cut-off points are by nature rather arbitrary and do not take into account factors such as the weighting applied (linear or quadratic), the relative prevalence of the categories or the number of categories being compared.

RESULTS

Paired data (matched patient and observer ratings) were available for 708 subjects (mean age = 71.7 years, proportion of men 399/708, 56%; median survival 50 days). The levels of agreement between patients' and clinicians' ratings are shown in table 1. Concordance was the lowest for 'dry mouth' (quadratic-weighted κ=0.27) and the highest for ECOG performance status (quadratic-weighted κ=0.68).

In total, 12 different prognostic models were created: four models derived using observer ratings only (PiPS/OR-A14, PiPS/OR-A56, PiPS/OR-B14 and PiPS/OR-B56), four models using patient-rated data only (PiPS/PR) and four models using composite data (PiPS/CM). The variables included in each of the models are shown in table 2. The PiPS/OR models had greater AUCs than either the PiPS/PR or PiPS/CM models (table 2), although it should be noted that these models were developed in slightly different populations (due to the differing pattern of missing data).

The models for 14-day and 56-day survival were then combined to produce an overall prognostic prediction in terms of 'days' (<14 days), 'weeks' (14–55 days) or 'months+' (>55 days) for each of the three types of data (observer-rated, patient-rated or composite). The accuracy of these models was evaluated by comparing the agreement between the model predictions and actual survival (table 3). Both of the observer-rated models provided statistically superior agreement with actual survival compared with the patient-rated models (p<0.001). The models developed using a composite rating method were not significantly better than predictions derived using observer ratings alone.

DISCUSSION

Statement of principal findings

We have found that observer ratings about the presence or absence of key symptoms, performance status and global health status are at least as good as patient-reported or composite scores in terms of prognostic ability. This is an important finding. Observer ratings are much easier to obtain and do not rely on the competence or otherwise of patients. Moreover, even in competent patients, obtaining patient-reported data

Table 1 Agreement between patient and observer ratings

	Quadratic-weighted κ	95% CI
Anorexia	0.47	0.40 to 0.54
Dyspnoea	0.40	0.33 to 0.47
Dry mouth	0.27	0.20 to 0.34
Dysphagia	0.37	0.29 to 0.44
Pain	0.52	0.46 to 0.59
Tiredness	0.26	0.17 to 0.36
ECOG	0.68	0.66 to 0.72
Global Health status	0.34	0.31 to 0.39

ECOG, Eastern Cooperative Oncology Group.

Table 2 Variables included in each of the models

	PiPS/OR				PiPS/PR				PiPS/CM			
	A14	A56	B14	B56	A14	A56	B14	B56	A14	A56	B14	B56
AMTS	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓
Pulse	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Any site of mets	✓	✓	✓		✓		✓		✓	✓	✓	
Liver mets	✓	✓				✓				✓	✓	
Bone mets	✓		✓		✓		✓		✓			
Anorexia	✓	✓	✓						✓			
Dyspnoea	✓								✓			
Dysphagia	✓											
ECOG PS	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓
Global health	✓	✓	✓	✓	✓	✓		✓	✓	✓		
Breast cancer		✓								✓		
Prostate cancer		✓		✓		✓	✓			✓		✓
Weight loss		✓				✓				✓		✓
Dry mouth					✓		✓		✓		✓	
Pain						✓				✓		✓
Tiredness				✓						✓		✓
WBC			✓	✓			✓	✓			✓	✓
Platelets			✓				✓	✓				✓
Urea			✓	✓			✓	✓			✓	✓
ALT			✓				✓					
CRP			✓	✓			✓	✓			✓	✓
Platelets				✓							✓	
Neutrophils				✓								✓
Lymphocytes				✓			✓				✓	✓
ALP				✓								✓
Albumin				✓				✓				✓
Sodium							✓				✓	
Bilirubin							✓				✓	
AUC	0.850	0.773	0.837	0.782	0.700	0.688	0.780	0.751	0.829	0.752	0.807	0.776
N	960	946	634	646	752	748	591	598	963	951	633	640

A14/A56, PiPS model developed for use in all patients (no blood results) to predict 14-day (A14) or 56-day (A56) survival; ALP, alkaline phosphatase; ALT, alanine transaminase; AMTS, abbreviated mental test score; AUC, area under the receiver operating characteristic curve; B14/B56, PiPS model developed for use in patients with blood results available to predict 14-day (B14) or 56-day (B56) survival; CRP, C reactive protein; ECOG PS, Eastern Cooperative Oncology Group Performance Status; mets, metastases; PiPS, Prognosis in Palliative care Scale; PiPS/CM, PiPS composite-scoring model; PiPS/OR, PiPS observer-rated model; PiPS/PR, PiPS patient-rated model; WBC, white blood count.

Table 3 Performance of the combined models

	Combined A-models			Combined B-models		
	Absolute agreement between model prediction and actual survival	Linear-weighted κ	p Value comparing κ scores with PiPS/OR	Absolute agreement between model prediction and actual survival	Linear-weighted κ	p Value comparing κ scores with PiPS/OR
PiPS/OR	57.3%	0.441	–	57.3%	0.373	–
PiPS/PR	51.5%	0.215	<0.001*	55.3%	0.314	0.013‡
PiPS/CM	56.8%	0.420	0.227†	57.8%	0.388	0.789§

*N=745.
†N=954.
‡N=594.
§N=638.
PiPS, Prognosis in Palliative care Scale; PiPS/CM, PiPS composite-scoring model; PiPS/OR, PiPS observer-rated model; PiPS/PR, PiPS patient-rated model.

imposes an extra burden on an already vulnerable population and does not improve prognostic ability.

Strengths and weaknesses

This was a large study that was specifically designed to answer the question as to whether patient reported outcomes provided better prognostic information than observer ratings or a mixture of the two. As far as we are aware, no previous study has systematically addressed this issue. The strengths of our study include the large sample size, the prospective design, the inclusion of both competent and incompetent patients and the use

of bootstrap statistical methods to minimise over-optimism during model development. However, there were a number of limitations. No specific instructions were given to clinicians about how to obtain the proxy ratings. Clinicians were simply asked to indicate whether they believed the patient to be experiencing the presence or absence of particular symptoms. Some clinicians may have obtained these data by directly asking the patient themselves, from direct observation of patient behaviour, by discussion with other members of the healthcare team or by information obtained from carers. This unstructured approach was adopted because it reflected the reality

of the clinical situation and one of the key features of the design of the PiPS study was an attempt to maintain the ecological validity of the research process. Nonetheless, this lack of standardisation means that we cannot give specific advice about the best way to obtain proxy ratings for prognostic decisions and it is possible that other clinicians may not perform as well at providing ratings as the clinicians who participated in this study.

For similar reasons, our desire to keep all assessments to a minimum meant that key terms were not precisely defined. Thus for instance, clinicians were asked to judge whether or not a patient was experiencing pain (yes/no), but no instructions were given about what level of pain should be considered significant, how long or how frequently the patient was experiencing pain for or whether it was adequately controlled with analgaesia. Given the relatively large number of symptoms that were assessed during this study it was considered that the use of more structured symptom assessment tools (such as the Brief Pain Inventory)[10] would have imposed too great a burden on both patients and clinicians and (even if it worked in a research setting) would be likely to lead to an unwieldy prognostic instrument.

It should also be remembered that none of the prognostic models described in this paper has yet undergone external validation in an independent data set.[11] As such, the assessment of the performance of each of the models can only be considered to be provisional.

Relationship to other studies

A number of studies have previously investigated the reliability or otherwise of using proxy ratings by clinicians or carers when patient self-report data are unavailable.[12–14] The Palliative Care Outcome Scale is a widely used audit tool for patients with advanced disease.[15] It consists of 10 items covering pain, other physical symptoms, psychological distress, social concerns and quality of life. There is both a patient-rated and a staff-rated component to the scale. During the validation of Palliative Care Outcome Scale, 145 subjects (32% of the study population) were able to complete the patient-rated component of the scale. Patient-rated scores were found to correlate acceptably well with staff-rated measures (linear-weighted κ values were >0.3 for 8/10 items).

Martin and colleagues[16] have recently reported on the prognostic significance of nutritional variables and patient-reported performance status in patients with advanced cancer. They studied 1767 palliative cancer patients who completed the Patient Generated Subjective Global Assessment (PG-SGA). The PG-SGA is a nutritional screening tool including self-report data on weight change, dietary intake and gastrointestinal symptoms.[17] It also includes a self-report version of the ECOG performance status. Clinicians completed the Palliative Performance Scale (PPS) for each of the study participants. The PPS is an observer-rated performance status measure which has been specifically developed for use in patients with advanced disease.[18] The authors reported that the most parsimonious predictive model included only two variables, diagnosis and performance status. Patient-reported and clinician-rated performance status measures were equally good at predicting survival in this population, leading Martin and coworkers[16] to recommend that patient-reported performance status may be of considerable practical utility as a prognostic tool. In contrast, we found that patient-reported outcomes (including patient-rated ECOG scores) provided less accurate prognostic information than clinician ratings. Moreover, we found that many

patients (24% of study participants) were unable to complete patient rating scales because of cognitive impairment.

The difference between our findings and Martin's may at least partly be explained by differences in the disease burden of the participants. The median survival of patients in Martin's study was 3.2 months (90 days), compared with 34 days for the 1018 subjects in our own study[3] and the proportion of patients with ECOG scores between 0 and 2 was 52% in Martin's study compared with our figure of 29%.

Unanswered questions

Quality of life measures are known to have prognostic significance across a variety of different cancers.[19] In terms of straightforward quality of life assessment, the patient's perspective is usually taken to represent the 'gold standard'. One might therefore have expected that observer ratings of symptoms, performance status and overall health would be a poor substitute for patient-generated data. In this large, multicentre, prospective study we found that observer ratings are acceptable and may in fact be superior to patient-reported data for the purposes of prognostication. This is important because it is not always possible or practical to obtain patient-reported data from patients with advanced cancer, either because of debility, fatigue or cognitive impairment.

However, it is not entirely clear why observer ratings should be so effective at providing prognostic information. Is it possible that patients are not the best judges of their own performance status? Theoretically, performance status (as measured by ECOG) could be 'objectively' assessed using real-time activity monitoring. Actimeters are portable accelerometers[20] which can record the proportion of the day that subjects are active, and have been used extensively in studies of fatigue, sleep disturbance and circadian rhythms.[21–23] Patients with 'paradoxical insomnia'[24] under-report the amount of sleep that they experience. Might real-time activity monitoring also reveal that some patients are similarly inaccurate in estimating their level of day-time activity? Inclusion of actimeters in future prognostic studies may provide a valuable insight into whether patients' or clinicians' perspectives more closely mirror reality.

Although activity monitoring may provide an objective measure of performance status, no such 'external' measure exists with which to assess the veracity of patients' reports of symptoms or global health status. It must be assumed therefore that patients' reports have the legitimacy of representing the way that the patients' themselves truly feel about their situation. However, just because patients' reports may be the most important source of information about the subjective experience of being ill, it does not necessarily follow that their own assessments will carry more prognostic significance. It is conceivable that clinicians' assessments are more valuable precisely because they are made through the filter of a healthcare professionals' clinical experience. Thus, some patients may report that they are not experiencing pain when in fact they have merely grown accustomed to a low level of background discomfort which is now considered simply normal. This phenomenon of resetting the internal calibration for symptom appreciation has been termed a 'response shift'.[25] Patients' self-reports may also be affected by numerous factors that are difficult to quantify such as cultural background or personal and family history.[26] In these circumstances, perhaps a clinician's assessment that the patient is in fact in pain may be more relevant than the patient's own perspective. The restricted nature of the data collected in this study makes it impossible

to explore these issues further. However, future researchers might want to consider undertaking some nested qualitative interviews with both patients and clinicians exploring how they arrived at their overall assessment about the symptom prevalence and global health status.

REFERENCES

1. **Steinhauser KE**, Christakis NA, Clipp EC, et al. Preparing for the end of life: preferences of patients, families, physicians, and other care providers. *J Pain Symptom Manage* 2001;**22**:727–737.
2. **Glare P**, Virik K, Jones M, et al. A systematic review of physicians' survival predictions in terminally ill cancer patients. *BMJ* 2003;**327**:195–198.
3. **Gwilliam B**, Keeley V, Todd C, et al. Development of prognosis in palliative care study (PiPS) predictor models to improve prognostication in advanced cancer: prospective cohort study. *BMJ* 2011;**343**:d4920.
4. **Department of Health.** Reference guide to consent for examination or treatment (Second edition). London: COI 2009.
5. **Oken MM**, Creech RH, Tormey DC, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. *Am J Clin Oncol* 1982;**5**:649–655.
6. **Chen CH**, George SL. The bootstrap and identification of prognostic factors via Cox's proportional hazards regression model. *Stat Med* 1985;**4**:39–46.
7. **Schumacher M**, Holländer N, Sauerbrei W. Resampling and cross-validation techniques: a tool to reduce bias caused by model building? *Stat Med* 1997;**16**:2813–2827.
8. **Roberts C**, McNamee R. Assessing the reliability of ordered categorical scales using kappa-type statistics. *Stat Methods Med Res* 2005;**14**:493–514.
9. **Landis JR**, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159–174.
10. **Cleeland CS**, Ryan KM. Pain assessment: global use of the Brief Pain Inventory. *Ann Acad Med Singap* 1994;**23**:129–138.
11. **Altman DG**, Vergouwe Y, Royston P, et al. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009;**338**:b605.
12. **Higginson IJ**, Gao W. Caregiver assessment of patients with advanced cancer: concordance with patients, effect of burden and positivity. *Health Qual Life Outcomes* 2008;**6**:42.
13. **Groenvold M**, Klee MC, Sprangers MA, et al. Validation of the EORTC QLQ-C30 quality of life questionnaire through combined qualitative and quantitative assessment of patient-observer agreement. *J Clin Epidemiol* 1997;**50**:441–450.
14. **Sprangers MA**, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease: a review. *J Clin Epidemiol* 1992;**45**:743–760.
15. **Hearn J**, Higginson IJ. Development and validation of a core outcome measure for palliative care: the palliative care outcome scale. Palliative Care Core Audit Project Advisory Group. *Qual Health Care* 1999;**8**:219–227.
16. **Martin L**, Watanabe S, Fainsinger R, et al. Prognostic factors in patients with advanced cancer: use of the patient-generated subjective global assessment in survival prediction. *J Clin Oncol* 2010;**28**:4376–4383.
17. **Bauer J**, Capra S, Ferguson M. Use of the scored Patient-Generated Subjective Global Assessment (PG-SGA) as a nutrition assessment tool in patients with cancer. *Eur J Clin Nutr* 2002;**56**:779–785.
18. **Kates WRP**, Singer HSMD. Sex chromosomes, testosterone, and the brain. (Editorial). Neurology 2000;54:2201.
19. **Montazeri A.** Quality of life data as prognostic indicators of survival in cancer patients: an overview of the literature from 1982 to 2008. *Health Qual Life Outcomes* 2009;**7:**102.
20. **Sadeh A**, Hauri PJ, Kripke DF, et al. The role of actigraphy in the evaluation of sleep disorders. *Sleep* 1995;**18**:288–302.
21. **Fernandes R**, Stone P, Andrews P, et al. Comparison between fatigue, sleep disturbance, and circadian rhythm in cancer inpatients and healthy volunteers: evaluation of diagnostic criteria for cancer-related fatigue. *J Pain Symptom Manage* 2006;**32**:245–254.
22. **Berger AM.** Patterns of fatigue and activity and rest during adjuvant breast cancer chemotherapy. *Oncol Nurs Forum* 1998;**25:**51–62.
23. **Ancoli-Israel S**, Liu L, Marler MR, et al. Fatigue, sleep, and circadian rhythms prior to chemotherapy for breast cancer. *Support Care Cancer* 2006;**14**:201–209.
24. **ICSD-2.** International classification of sleep disorders, 2nd ed. Diagnostic and coding manual. American Academy of Sleep Medicine 2005.
25. **Sprangers MA.** Response-shift bias: a challenge to the assessment of patients' quality of life in cancer clinical trials. *Cancer Treat Rev* 1996;**22 Suppl A:**55–62.
26. **Solano JP**, Gomes B, Higginson IJ. A comparison of symptom prevalence in far advanced cancer, AIDS, heart disease, chronic obstructive pulmonary disease and renal disease. *J Pain Symptom Manage* 2006;**31**:58–69.

BMJ Supportive &
Palliative Care

Patients' reports or clinicians' assessments: which are better for prognosticating?

Paddy Stone, Bridget Gwilliam, Vaughan Keeley, et al.

Updated information and services can be found at:

http://spcare.bmj.com/content/2/3/219.full.html

These include:

References	This article cites 24 articles, 7 of which can be accessed free at: **http://spcare.bmj.com/content/2/3/219.full.html#ref-list-1**
Email alerting service	Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.
Topic Collections	Articles on similar topics can be found in the following collections Editor's choice (10 articles)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/