EVIDENCE BASED PUBLIC HEALTH POLICY AND PRACTICE

Offering a prognosis in lung cancer: when is a team of experts an expert team?

F Kee, T Owen, R Leatham

J Epidemiol Community Health 2007;**61**:308–313. doi: 10.1136/jech.2005.044917

The outlook for patients with lung cancer is poor, so an accurate estimation of prognosis will underpin treatment decisions and allow patients to make personal plans for the future. However, evidence suggests that there is a variation between doctors in their predictions of outcomes and also they tend to be overoptimistic. Two main questions are addressed in this study: whether multidisciplinary team discussion changes prognostic accuracy of individual clinicians; and whether team discussion improves the accuracy of the team's aggregated prediction. A real-time study of 50 newly diagnosed patients discussed by a regional lung cancer team was undertaken. A case proforma informed the completion of a pre-discussion questionnaire by each team member, seeking prognostic predictions at specific time points. This was repeated after team discussion. Medical notes were reviewed at 6 months to establish actual survival status. Group discussion did not significantly change the accuracy of survival predictions for any one clinician, but the team as a whole performed better after case discussion. Predictions which the clinicians were more confident about were found to be no more accurate than those where they were less confident. There is a wide variation in the range and accuracy of prognostic predictions made by individual clinicians, with no consistent improvement after team discussion. As such predictions are integral to decision making, further research on decision-making processes of clinical teams is required.

subsequently published by the Department of Health.²

The outlook for patients with lung cancer is poor. The most common management aim is not to prolong survival but to maintain a good quality of life for as long as possible. Nevertheless, an assessment of prognosis, either explicitly or implicitly, underpins most management decisions, influencing the choice of treatment, its timing and duration. Good prognostic judgment also contributes to the efficiency of healthcare, as expensive medical treatments administered without benefit expose patients to toxicity without gain, and also consume resources that may have been used more effectively elsewhere.³ More importantly, an accurate prognosis allows patients and their families to make the best possible personal plans for the future. Thus, it would be beneficial if patients received similar prognostic advice from all members of the multidisciplinary team.

Studies have shown that doctors tend to make overoptimistic predictions,^{4,5} and also there is variation in the accuracy of prognostication among doctors from different backgrounds.⁴ Most research on group decision making appears in the management science and psychology literature, but there have been no studies on how multidisciplinary team discussion affects estimates of prognosis by cancer specialists. We have conducted a study of decision making in a regional multidisciplinary lung cancer team to address the following questions: does team discussion change the prognostic accuracy of clinicians?; does team discussion improve the accuracy of the team's aggregated predictions?; and how is the confidence of the clinicians related to their prognostic accuracy?

METHODS

The authors obtained the agreement of the Northern Ireland Regional Lung Cancer Team (based at the Belfast City Hospital, Belfast, UK) to study, in real time, decision making on cases that were scheduled for discussion at a weekly multidisciplinary meeting. The meeting was attended by respiratory physicians working in the greater Belfast area and visiting specialist oncologists and thoracic surgeons. Two radiologists were also normally present as members of the team to discuss results of imaging investigations relating to staging and operability. In all, 15 clinicians participated in the study, comprising 9 respiratory physicians, 3 oncologists and 3 thoracic surgeons. Although attendance of individuals varied during

Many reforms in the provision of cancer care were initiated in the UK following the publication of the report of the Chief Medical Officer's Expert Advisory Group on Cancer in 1995.¹ One of the key recommendations was that specialist care should be delivered by multidisciplinary teams. This recommendation was based on the assumption that the discussion of a case by a multidisciplinary team will enhance the quality of patient care by facilitating communication between team members and allowing broader identification of the patient's need, leading to better coordination of services to meet those needs. This approach may be of significant benefit for patients with lung cancer who require expert medical, surgical and oncology input, as well as the care of skilled nurses, physiotherapists and others. However, whether such teams share a common understanding about the patient and actually make better decisions was not directly addressed in the extensive literature reviews

See end of article for authors' affiliations

Correspondence to: Professor F Kee, Epidemiology and Public Health, Queens University of Belfast, Mulhouse Building, Grosvenor Road, Belfast BT12 6BJ, UK; f.kee@qub.ac.uk

Accepted 23 October 2006

the study period owing to other clinical commitments and leave entitlements, all specialities were represented at most of the team meetings.

A convenience sample of 50 cases of newly diagnosed primary lung cancer were recruited between December 1999 and January 2003. The cases reflect the referral practices of the physicians and were selected according to availability of pathology and radiology reports in advance of the scheduled meeting. Exclusions were not made on the basis of stage of disease. Due to time constraints, only one case was studied at each meeting. The total number of patients discussed at each team meeting ranged from 3 to 22.

To study the impact of team discussion, participants' views on each case were elicited before and after the discussion of clinical findings and management options. For this purpose, a research nurse abstracted the necessary clinical details and transcribed them on to a printed proforma (appendix) which was circulated to team members 24 h before the scheduled meeting. Views of the participating clinicians were sought with an accompanying questionnaire which they completed before attending the meeting. Each clinician was asked to estimate the patient's chance of survival with and without treatment at 30 days, 6 months and 1 year, and the chance of the patient experiencing morbidity at least as severe as World Health Organization Performance Status 3 (limited or no self-care and confined to bed or chair >50% of the time), at each of these time points. Only the results for the 6-month time point are reported in this paper.

At the multidisciplinary meeting, the responsible clinician presented the details of the case and participants were able to view the radiological films before group discussion took place. At that time, they were offered the opportunity to change any of their initial questionnaire responses, in light of seeing the images for themselves. The questionnaires were collected by the research nurse before discussion began. Clinicians did not share their questionnaire responses as a basis for discussion, the latter proceeding as normal, unguided by the researchers. Immediately after case discussion, the participants individually completed an identical questionnaire, again eliciting their views on prognosis and rating their confidence in the accuracy of the prediction they had made. The patient proforma and questionnaires were piloted with the members of the team, and revisions made, before commencement of the study.

A research nurse subsequently reviewed the medical notes of each case to determine survival and morbidity status at the 6-month follow-up point.

Statistical methods

The accuracy of a clinician's prognostic predictions over n cases can ordinarily be assessed by deriving a Brier Score (or probability score (PS))⁶ where

$$PS = (1/n) \sum (\text{predicted probability} - \text{actual outcome})^2$$

The actual outcome is scored 1 if the patient survives and 0 if he/she dies. After Arkes *et al.*,⁷ we have used the extended covariance decomposition of this probability score (PS):

$$PS_{(\text{average})} = d(1-d) + (\text{bias})^2 + [d(1-d)(\text{slope})](\text{slope}-2) + \text{scatter}$$

where d represents the base rate of the to-be-predicted event (ie, the proportion of patients who actually survived). Bias is the difference between the average judgment (f) and the survival rate (d). As such, it is a good global measure of judgement accuracy. In our study, bias is the difference between the average estimated probability of survival and the

actual proportion surviving, and may be either positive or negative. Lower absolute values are better. As calibration refers to the correspondence between what is predicted and what actually occurs, bias also represents a good global measure of calibration.⁷

Slope is the average probability of survival assigned to those who did survive minus the average probability of survival assigned to those who did not survive. Higher slope is better because it signifies better discrimination, in that the judge was able to differentiate between the survivors and non-survivors.

Bias and slope indicate the accuracy of the prognostic estimates, whereas scatter is an index of the "noisiness" of the estimates. It is a pooled variance derived from the distribution of probability estimates assigned to those who did survive and the distribution of estimates assigned to those who did not survive.

To assess the effect of team discussion on the accuracy of decision making, the Brier Scores of participating doctors, before and after the team discussion, were combined using a random effects model as described by Laird and Mosteller.⁸ This analysis provided an estimate (and its SE) of the mean of the population of such differences, the doctors in the study being regarded as a random sample.

However, as recommended by others,⁹ merely averaging the estimates of all group members may provide an efficient way to measure group performance. Thus, an aggregated team prediction was derived for each of the 50 cases by averaging the predictions of the participating team members on that day.

On completing each questionnaire, clinicians were asked to indicate their confidence in the accuracy of the prognostications on a scale of 1–9: where 1 is not confident at all and 9 is very confident. A before versus after comparison was carried out using the Wilcoxon signed rank test.

The study was approved by the Queen's University Research Ethics Committee and complies with the principles embodied in the Declaration of Helsinki.

RESULTS

Table 1 illustrates the demographic and clinical profile of the cases included in the study. The sample was closely representative of the age and gender profile of newly diagnosed (1998–2000) patients with lung cancer in Northern Ireland.¹⁰ Follow-up details were obtained on all 50 patients studied. Of these, 24 (48%) had died within 6 months of the team meeting.

Table 2 shows the calculated Brier Scores of each of the participating clinicians, before and after case discussion. A lower absolute Brier Score indicates a more accurate prediction of the eventual outcome. After the meeting, 9 of the 11 consultants had improved their Brier Score (modestly) for prediction of survival at 6 months. However, combining the doctors' results using the technique described by Laird and Mosteller,⁸ the mean of such differences is estimated to be -0.007 (SE 0.004) with 95% CIs of -0.014 to 0.001 . Therefore, on the basis of the available evidence, there are no grounds for rejecting the null hypothesis of zero mean difference.

The Brier Scores representing the aggregated prediction for the team showed a modest decline from 0.22 before the team meeting to 0.21 after discussion, indicating better prediction after the case had been discussed. So, while group discussion did not significantly change the accuracy of survival prediction for any one clinician, individuals and the team as a whole generally performed better.

Table 3 indicates how the prognostic estimates differed in terms of calibration and discrimination. Comparison between indicators before and after team discussion results showed that the team as a whole improved its calibration and

Table 1 Demographic and clinical details of the sample

Gender	
Male	31 (62%)
Female	19 (38%)
Age (years)	
Mean	68.2
Median	70
Range	46–83
WHO performance status	
0	5 (10%)
1	31 (62%)
2	12 (24%)
3	2 (4%)
Cell type (where known)	
Small cell	11 (22%)
NSC	39 (78%)
Squamous cell	27
Adenocarcinoma	1
Undifferentiated NSC	10
Stage at diagnosis (where known)*	
Small cell	
Extensive	4 (14%)
Limited	4 (14%)
NSC	
I	6 (21%)
II	2 (7%)
III	10 (34%)
IV	3 (10%)

NSC, non-small cell; WHO, World Health Organization.

*Reported results of the CT scan were not always available before the multidisciplinary team meeting to allow staging to be included on the proforma. However, all films were viewed and commented on by a radiologist at the meeting, before completion of the pre-discussion questionnaire.

discrimination of prognostic estimates after the team meeting. However, there were no consistent trends across the individual clinicians.

Clinicians were found to have significantly more confidence in their prognostic predictions after the team discussion ($p < 0.001$). Before the team meeting, 32.5% of clinicians scored their confidence level as ≥ 7 , while this increased to 47.4% afterwards.

Figure 1 illustrates the relationship between the confidence of the clinician in his or her initial prediction before team discussion and the eventual accuracy of that prediction, for each patient–doctor episode. Predictions which the clinicians were more confident about were found to be no more accurate than those where they were less confident.

DISCUSSION

Team composition, working methods and workloads are related to the effectiveness of cancer teams, including the quality of clinical care they deliver.^{11–12} However, there have been surprisingly few empirical attempts to understand how and why a multidisciplinary cancer team might offer different advice from that of an individual clinician. This is important if the service is to apply what has been learned to promote better quality of care. One of the few recent studies in the UK¹³ investigated the workings of a gastrointestinal multidisciplinary team, focusing on the number of occasions when treatment decisions were not implemented. It was largely based on record review rather than, as in the present investigation, studying real time prognostication decisions during the team meeting. Given the centrality of the multidisciplinary team to the cancer service reforms, the absence of significant research on the way team-based care affects clinical decision making is surprising. For reasons cited in our introduction, we have chosen to study one aspect of team decision making that relates to formulating a prognosis, as this information is often the highest priority for seriously ill patients, even eclipsing their interest in treatment options.⁵ The perceived severity of illness and its prognosis are also key factors in the advance planning discussions on end-of-life care, which are likely to take place between the doctor and patient.¹⁴ Others have investigated the accuracy of prognostic prediction in lung cancer for individual doctors^{5–14}; however, until now, there have been no studies of the effects of team discussion on these prognostications or on treatment decisions. The findings from this study relating to the correlation between prognostic predictions and subsequent treatment recommendations have been reported elsewhere.¹⁵

Among a group of experienced consultants, we found a wide variation in the range and accuracy of prognostic predictions made for a group of patients in a real-time multidisciplinary team setting. Furthermore, there was no consistent improvement in their individual predictions after team discussion of the cases. On the other hand, the unweighted average prediction of the group did become slightly more accurate after discussion.

We chose to investigate prognostic prediction at 6 months, given that there is significantly greater uncertainty at that time point than at 1 year, when at least 80% of the patients are likely to have died. The fact that the magnitude of the Brier Scores seems modest is, thus, not surprising and is, in fact, little different from studies of clinicians' predictions for other seriously ill patients.^{16–17} An initial hypothesis was that an individual clinician's predictions might be more accurate after having had the benefit of team discussion. Although this was not the case, the performance of the group, as reflected in the arithmetical average prediction, did improve slightly.

Table 2 Brier Scores: before and after case discussion

Clinician*	n	Brier Score		Difference in Brier Score	SE of difference
		Before meeting	After meeting		
A	35	0.179	0.171	–0.008	0.005
B	29	0.246	0.239	–0.007	0.01
C	13	0.196	0.171	–0.025	0.015
D	7	0.256	0.188	–0.069	0.059
E	26	0.268	0.257	–0.011	0.01
F	10	0.141	0.126	–0.015	0.021
G	10	0.267	0.269	0.002	0.002
H	29	0.221	0.214	–0.007	0.014
I	19	0.274	0.243	–0.031	0.018
J	12	0.244	0.278	0.034	0.022
K	23	0.190	0.171	–0.019	0.022

*Individual Brier Scores were only calculated for clinicians who participated in, and completed proformas for, >3 case discussions.

Table 3 Calibration and discrimination of the prognostic estimates

	Before case discussion			After case discussion		
	Bias	Slope	Scatter	Bias	Slope	Scatter
Aggregated group	0.073	0.130	0.030	0.063	0.150	0.030
Most-accurate clinician*	-0.050	0.130	0.016	-0.080	0.220	0.024
Least-accurate clinician*	0.011	0.020	0.047	0.001	0.080	0.037

*Based on Brier Score before case discussion.

There is a substantial non-medical literature on the subject of pooled or group forecasts, and though, in some problem-solving circumstances, a group's performance will turn out to be better than the best team member; groups may have both positive and negative effects, depending on the group dynamics and decision-making processes.¹⁸ Few of these studies have been carried out in a clinical domain, let alone focused on the care of seriously ill patients. Poses *et al*'s¹⁹ study of patients managed in intensive care units found the groups' average prediction to be superior to the best clinician, but did not ascertain the doctors' predictions before and after team discussion. Interestingly, in their discussion, they also argued that averaging procedures may produce better judgments than group discussion, or more complex methods such as the Delphi technique.¹⁹ Group interaction may polarise people, amplifying individual biases and causing them to discount information about outcome prevalence.²⁰

Since these earlier studies did not report whether panel discussion brought about change in Brier Scores, we had no real basis for an a priori power calculation. A post hoc power calculation for the combined estimate procedure would suggest that a change of 0.01 in the mean difference could be detected with a power of 70%, although we would have had insufficient power to detect such an effect for individual doctors. The clinical significance of such an effect is uncertain but on the basis of our 95% CIs we could be reasonably sure that larger effects have been excluded. In future studies, account should be taken of the possibility that power may be affected by variability in the characteristics of patients discussed at the team meetings, whether the effects of team discussion may be different for doctors from different specialties, where the team is on its own learning curve, and the way the team achieves consensus. There may be other important distinctions between doctors who prognosticate well and those who do not, even though the latter group may express very sound global treatment preferences.

Although quantitative judgments have been found to correlate with medical decision making,^{21–22} a multidisciplinary

team discussion obviously serves for more than the refinement of prognosis. Clearly, a number of other judgments will be made. We also studied the clinicians' views about the primary objectives of care (prolongation of survival, maximisation of quality of life or achievement of a good death) and there was some significant divergence of opinion on this fundamental objective, even after group discussion. In at least 19 (38%) of our 50 cases, the consensus on the primary objective of care was not unanimous.

Christensen *et al*,²³ in one of the very few clinical studies on group decision making, found that open group discussion of a patient's case was a less-than-optimal means of pooling members' unshared information in clinical teams. Their findings corroborated others from non-clinical fields^{24–25} and led to the conclusion that methods to impose a structure to team discussion, to make it more systematic, may provide greater opportunity for sharing problem-relevant information and increase the accuracy of diagnostic decision making. Such work suggests further avenues for future research with respect to prognostic judgments and treatment preferences of multi-disciplinary cancer teams. Indeed, there is now a sizeable research tradition that has developed a range of analytical techniques for analysing the product of team discussions, little of which has reached the health services research domain.^{26–27}

Although innovative and based on real cases rather than vignettes, our study is not without shortcomings. Our sample may not have been representative of all patients with lung cancer in Northern Ireland, or all team discussions over the period. We studied a consecutive sample for which we could assemble all or most of the clinical details at least 24 h before the meeting. In this respect, the clinicians usually had more information available to them than they would otherwise have

Figure 1 Confidence in prognostic prediction versus accuracy of prediction for each patient-doctor pairing, before case discussion by the multidisciplinary team (trend line represents the line of best fit).

What this paper adds

- Treatment decisions for cancer patients are made on the basis of predicted prognosis.
- A real-time study of how multidisciplinary team discussion affects estimates of prognosis by cancer specialists is presented.
- The range and accuracy of prognostic predictions varied widely between individual clinicians.
- Team discussion did not result in consistent improvement in prognostic predictions.
- Further investigation of team-based clinical decision

Policy implications

Cancer teams and individual members in these teams may learn more about how to improve their care if they record and monitor how the multidisciplinary team meeting affects the accuracy of their prognostic judgments and treatment choices.

APPENDIX Sample proforma

Please review the case below which will be discussed at the next Multidisciplinary team meeting and give a view (now) on the questions posed.

Age	63 years
Sex	Male
WHO performance status	2
Duration of history	4 week history of haemoptysis. Weight loss in past 6 months. Reduced exercise tolerance due to shortness of breath.
Nature of presentation	Haemoptysis, weight loss, poor appetite, shortness of breath on exertion. Stopped smoking 7 years ago (previously 40 per day).
Primary or most problematic complaint now	Haemoptysis & shortness of breath
Lung function tests	FVC = 55% predicted value; FEV ₁ = 45% predicted value
Usual exercise tolerance	20 yards
Principal clinical findings	Reduced air entry at right upper lobe. Cachexic.
Findings on x ray/CT/MRI scan	Chest x ray: suspicious area of shadowing in right upper lobe. CT scan: large mass lying posteriorly in the right mid zone which appears to involve both the posterior aspect of the upper lobe and also the lower lobe. It has a diameter of 8×6 cm. It extends down onto the right main bronchus and to within about 2 cm of the carina. There are a few small nodes in the pretracheal region. There is some lateral pleural reaction or thickening but no pleural effusion and no bony erosion. There is a nodule in the left lower lobe which is not calcified and measures approximately 12×9 mm. Its significance is uncertain but it could represent a significant nodule. No focal abnormality is seen in the liver or adrenals.
Presumed stage	Stage III
Findings on bronchoscopy	Marked distortion of the right upper lobe with an anterior bulge at the origin of the right upper lobe. The anterior segment was patent but both the posterior and apical segments of the right upper lobe were markedly distorted and extrinsically compressed.
Cell type, differentiation	Bronchial washings and brushings: malignant cells from a non-small cell carcinoma, probably squamous in type.
Other significant comorbidity	History of transient ischaemic attacks and bronchiectasis.
Serum alkaline phosphatase	68 µl/l

FEV₁, forced expiratory volume in 1 s; FVC, forced vital capacity; MRI, magnetic resonance imaging; WHO, World Health Organization.

had. A further criticism may be the possibility of a Hawthorne effect,²⁸ as the knowledge that they were being studied may have influenced the responses. The most-frequent informal feedback given to us by the participants was that the use of the proforma helped them focus the discussion a little more than it would have been otherwise. Therefore, in one sense, the prognostic accuracy observed might tend towards an upper bound, for this group at least.

Although a randomised trial of different team modus operandi would be logistically very difficult, future observational studies should include a number of different teams in different settings and employ a combination of qualitative and quantitative analyses. Incorporating and studying the effects of different communication and discussion mediums would clearly be of increasing value.²⁹

ACKNOWLEDGEMENTS

We thank Dr S Lovell, Dr L Garske, Dr M Kelly, Dr C O'Dochartaigh, Dr D McAuley, Dr R Donnelly and the secretarial staff of the participating consultants for their assistance in identifying study cases. We also thank Dr J Lawson and Dr J Foster, consultant radiologists. We would like to thank Dr Gordon Cran for a helpful contribution in deriving the Random Effects Model.

FK had the original idea for the article, TO and FK did the literature search, and TO, FK and RL wrote the article. FK and TO act as guarantors for the study.

Authors' affiliations

T Owen, F Kee, R Leatham, Department of Epidemiology and Public Health, The Queen's University of Belfast, Belfast, UK

Funding: The study was funded by the Research and Development Office of the Northern Ireland Health and Personal Social Services. The authors' work was independent of the funders.

Competing interests: None.

The study was approved by the Research Ethics Committee of Queen's University, Belfast, UK.

The participant members of the Northern Ireland Regional Lung Cancer Team: Dr R Eakin, Professor S Elborn, Dr I Gleadhill, Mr A Graham, Dr S Guy, Dr L Heaney, Dr J Kidney, Mr J McGuigan, Dr J MacMahon, Mr K McManus, Dr AM Nugent, Dr A Patterson, Dr M Riley, Dr R Shepherd and Dr S Stranex.

REFERENCES

- 1 Expert Advisory Group Report on Cancer. A policy framework for commissioning cancer services: a report by the Expert Advisory Group on cancer to the Chief Medical Officers of England and Wales. London: Department of Health, 1995.
- 2 National Health Service Executive. Guidance on commissioning cancer services: improving outcomes in lung cancer. London: NHSE, 1998.
- 3 Detsky AS, Stricker SC, Mulley AG, et al. Prognosis, survival and the expenditure of hospital resources for patients in an intensive care unit. *N Engl J Med* 1981;**305**:667–72.
- 4 Muers MF, Shevlin P, Brown J, on behalf of the participating members of the Thoracic Group of the Yorkshire Cancer Organisation. *Thorax* 1996;**51**:894–902.
- 5 Christakis NA, Lamont EB. Extent and determinants of error in doctors' prognoses in terminally ill patients: prospective cohort study. *BMJ* 2000;**320**:469–73.
- 6 Yates JF. Scoring rules for forecasts. Ann Arbor, MI: University of Michigan, 1981. Cognitive Science Technical Report Series #14.
- 7 Arkes HR, Dawson NV, Speroff T, et al. The covariance decomposition of the probability score and its use in evaluating prognostic estimates. *Med Decis Making* 1995;**15**:120–31.
- 8 Laird N, Mosteller F. Some statistical methods for combining experimental results. *Int J Technol Assess Health Care* 1990;**6**:5–30.
- 9 Poses RM, Bekes C, Winkler RL, et al. Are two (inexperienced) heads better than one (experienced) head? Averaging house officers' prognostic judgments for critically ill patients. *Arch Intern Med* 1990;**150**:1874–8.
- 10 Northern Ireland Cancer Registry. <http://www.qub.ac.uk/nicr> (accessed 1 Feb 2007).
- 11 Haward R, Mair Z, Borrell C, et al. Breast cancer teams: the impact of constitution, new cancer workload, and methods of operation on their effectiveness. *Br J Cancer* 2003;**89**:15–22.
- 12 Gregor A, Thomson CS, Brewster DH, et al. Management and survival of patients with lung cancer in Scotland diagnosed in 1995: results of a national population based study. *Thorax* 2001;**56**:212–17.
- 13 Blazeby JM, Wilson K, Metcalfe C, et al. Analysis of clinical decision-making in multi-disciplinary cancer teams. *Ann Oncol* 2006;**17**:457–60.
- 14 Wachter R, Luce J, Hearst N, et al. Decisions about resuscitation: inequities among patients with different disease but similar prognoses. *Ann Intern Med* 1989;**111**:525–32.
- 15 Kee F, Owen T, Leatham R. Decision making in a multidisciplinary cancer team: does team discussion result in better quality decisions? *Med Decis Making* 2004;**24**:602–13.
- 16 Dolan J, Bordley D, Mushlin A. An evaluation of clinician's subjective probability estimates. *Med Decis Making* 1986;**6**:216–23.
- 17 Poses R, Cebul RD, Centor RM. Evaluating physicians' probabilistic judgments. *Med Decis Making* 1988;**8**:233–40.
- 18 Jones P, Roelofsma HMP. The potential for social, contextual and group biases in team decision making: biases, conditions and psychological mechanisms. *Ergonomics* 2002;**43**:1129–52.
- 19 Fischer GW. When oracles fail: a comparison of four procedures for aggregating subjective probability forecasts. *Organ Behav Hum Decis Process* 1981;**28**:96–110.
- 20 Myers DG, Lamm H. The group polarisation phenomenon. *Psychol Bull* 1976;**83**:602–27.

- 21 **Tierney WM**, Fitzgerald J, McHenry R, *et al*. Physicians estimates of the probability of myocardial infarction in emergency room patients with chest pain. *Med Decis Making* 1986;**6**:12–17.
- 22 **McNutt RA**, Selker HP. How did the acute ischaemic heart disease predictive instrument reduce coronary care unit admissions? *Med Decis Making* 1988;**8**:90–4.
- 23 **Christensen C**, Larson J, Abbott A, *et al*. Decision making of clinical teams: communication patterns and diagnostic error. *Med Decis Making* 2000;**20**:45–50.
- 24 **Stasser G**, Taylor LA, Hanna C. Information sampling in structured discussion among three and six person groups. *J Pers Soc Psychol* 1989;**57**:67–78.
- 25 **Stasser G**, Titus W. Pooling of unshared information in group decision making: biased information sampling during discussion. *J Pers Soc Psychol* 1985;**48**:1467–78.
- 26 **Parks C**, Kerr N. Twenty five years of social decision scheme theory. *Organ Behav Hum Decis Process* 1999;**80**:1–2.
- 27 **Hinsz V**. Group decision making with responses of a quantitative nature: the theory of social decision schemes for quantities. *Organ Behav Hum Decis Process* 1999;**80**:28–49.
- 28 **Roethlisberger FJ**, Dickson WJ. *Management and the worker*. Cambridge: Harvard University Press, 1939.
- 29 **Patel V**. Individual to collaborative cognition: a paradigm shift? *Artif Intell Med* 1998;**12**:93–6.