modalities and technology have improved the physicians' ability to assess the SPN, using information such as margin patterns, size, growth rate and location. Diagnostic and treatment decisions depend on the physician's assessment of the probability of malignancy of each SPN.

Improving the diagnostic accuracy of SPN in individual patients is a challenge. Previous methods for estimating the likelihood of each diagnosis used either Bayesian statistics or multivariate logistic regression analysis.[6–8] A Bayesian statistical model estimates the likelihood ratios by measuring the probability of finding a given feature in a population with malignant SPNs and dividing that by the probability of finding the same feature in a population with benign SPNs.[6,7] A certain scepticism regarding the Bayesian approach exists, and these models are not often applied in Western countries.[9] After the advent of CT scanning, Swensen *et al.* reported a multivariate logistic model that incorporated patient age, smoking history, a history of extra-thoracic malignancy, diameter, upper lobe location of the SPN, and the presence or absence of spiculation.[8]

Serum tumour markers are associated with malignancy.[10] No study has analysed the role of biochemical markers in differentiating between malignant and benign SPNs, and whether it may improve the performance of a prediction model. The aim of this study was to develop a simple prediction model for the diagnosis of SPN, based on clinical characteristics and thin-section CT findings, and to evaluate its diagnostic accuracy compared with that of an experienced chest radiologist.

METHODS

Records of 452 patients with SPNs who underwent surgery because of suspicion of primary lung cancer between July 1998 and March 2004 were retrieved (the training set: 113 benign SPNs, 339 malignant SPNs). All resected SPNs were evaluated histologically and a diagnosis made. Inclusion criteria were an SPN which was a solitary, round or oval lesion ≤30 mm in diameter and without associated adenopathy, atelectasis or effusion.[11] Any SPN diagnosed as metastatic extrapulmonary cancer was excluded. Benign SPN was defined as a benign pathological diagnosis. Malignant SPN was defined as that caused by primary lung cancer, such as non-small-cell or small-cell lung cancer. This study was approved by the institutional review board after confirmation of informed consent by the patient to have records and images reviewed.

The CT scans were acquired on single helical or multidetector scanners (X-Vigor, or Aquilion, Toshiba Medical Systems, Tokyo, Japan). The helical technique in all patients consisted of 10.0-mm collimation for individual scans of the entire lung (120 kVp, 150–200 mA) and reconstruction by a standard algorithm prior to surgery. In all patients, additional thin-section CT images were obtained with 2.0-mm collimation, a 20-cm field of view, 120 kVp and 200 mA per rotation, 0.5- or 1.0-s gantry rotation, and a high-spatial-frequency reconstruction algorithm. Hard-copy images were photographed at window settings for the lung (centre: –600 HU and width: 2000 HU) and mediastinum (centre: 50 HU and width: 550 HU).

Development of the prediction model

The clinical characteristics of patients comprising the training set ($n = 452$) were collected from the medical records: gender, age, smoking status (current or former smoker), history of extra-thoracic malignancies (>5 years previously), blood count and blood chemistry, including WCC, CRP, and serum tumour markers: carcinoembryonic antigen (CEA), sialyl-specific embryonic antigen, cytokeratin 19 fragment, squamous-cell carcinoma antigen, carbohydrate antigen 19–9, neurone-specific enolase and progastrin-releasing peptide. Two observers (K.Y., S.T.), blind to the clinical or histopathological diagnoses, reviewed the thin-section CT images, and the final decisions on diagnosis were reached by consensus. Thin-section CT findings of SPN in the training set were reviewed as follows: lung side (right vs left), location (lobe), diameter (mm), the presence or absence of calcification, cavity, lobulation, spiculation and CT bronchus sign. The CT bronchus sign was defined as a visible bronchus or bronchiolus leading to the SPN.[12]

Categorical variables of clinical characteristics and thin-section CT findings were compared using the chi-square test. Continuous variables were compared using the *t*-test. The prediction model was identified from multivariate logistic regression analysis. A stepwise procedure was used to select independent variables from the statistically significant variables in univariate analysis. Continuous variables in serum tumour markers without complete data sets were excluded from the logistic regression analysis. Some functional forms for continuous variables were also explored in the logistic regression analysis. The Hosemer-Lemeshow test was performed to evaluate goodness-of-fit of the prediction model. The performance of the prediction model was evaluated by calculating the prediction accuracy, the receiver-operating characteristic (ROC) curve and area under the curve. The statistical analysis was performed with SAS version 8.2 (SAS Institute, Cary, NC, USA), and the significance level was set at 0.05 (two-sided).

Validation of the prediction model

The prediction model was validated by comparing its accuracy with that of an experienced chest radiologist in the validation set. The validation set comprised a consecutive series of 148 patients with undiagnosed SPNs, who underwent thin-section CT imaging prior to surgery, between May 2004 and April 2005. The resected SPN was diagnosed pathologically as benign or malignant.

The reviews of thin-section CT images were performed independently before surgery. An observer (U.T.) reviewed the radiological predictive factors for

Table 1 Comparison of the clinical and radiological characteristics of the patients diagnosed with a benign or malignant solitary pulmonary nodule in the training set

Classification	Benign (n = 113)	Malignant (n = 339)	Total (N = 452)	P-value*
Clinical characteristics				
Male (%)	54	56	55	0.70
Mean age (years)	58	64	62	0.002
Current or past smoker (%)	47	54	49	0.86
Mean pack-years (n)	19	24	23	0.59
Other cancer >5 years ago (%)	11	10	10	0.79
Mean WCC ($\times 10^9$/L)	6.27	6.33	6.31	0.41
Mean serum CRP (mg/L)	1.0	3.0	2.0	<0.001
Mean serum CEA (ng/mL)	2.7	5.6	4.9	<0.001
Radiological characteristics				
Mean diameter (mm)	17	20	19	0.02
Side (right) (%)	65	63	64	0.65
Location (%)				
LUL	11	17	15	0.14
Lingular segment	5	5	5	0.80
LLL	18	14	15	0.37
RUL	30	31	31	0.91
RML	13	9	10	0.21
RLL	22	22	22	0.99
Calcification (%)	12	2	5	<0.001
Cavitation (%)	10	8	8	0.49
Lobulation (%)	16	12	13	0.26
Spiculation (%)	2	84	63	<0.001
CT bronchus sign (%)	32	82	69	<0.001

*Univariate analysis: chi-squared test and t-tests were performed for proportional differences or mean differences in variables between benign and malignant SPNs.

CEA, carcinoembryonic antigen; LLL, left lower lobe; LUL, left upper lobe; RLL, right lower lobe; RML, right middle lobe; RUL, right upper lobe.

the prediction models in the thin-section CT images. An experienced chest radiologist (M.K.) reviewed the thin-section CT images with no knowledge of either the clinical data or the present prediction model. This experienced chest radiologist evaluated prediction of an SPN on a 5-point scale, scored as follows: 1, benign; 2, probably benign; 3, indeterminate; 4, probably malignant; and 5, malignant. He then determined a diagnosis (malignant or benign SPNs). Based on the pathological diagnosis, all SPNs were divided into benign or malignant for the purpose of evaluation of the validation set.

RESULTS

The clinical and the radiological characteristics of the 452 patients in the training set (113 benign SPNs, 339 malignant SPNs) are presented in Table 1.

Sixty-three per cent of the benign SPNs were granulomas, and the remainder included sclerosing haemangioma and hamartoma. The majority of the malignant SPNs were adenocarcinoma of the lung (80.2%). Other malignancies included squamous-cell carcinoma (14.5%), large-cell carcinoma (3.5%), adenosquamous carcinoma (0.3%), adenoid cystic carcinoma (0.3%), neuroepidermoid carcinoma (0.3%) and small-cell lung carcinoma (0.9%).

The SPN prediction model defined the probability of benign SPN as follows: Probability of benign SPN = $e^x/(1 + e^x)$, $x = 3.7009 + (3.0705 \times \text{calcification}) + (-1.3243 \times \text{CT bronchus sign}) + (-5.3399 \times \text{spiculation}) + (-1.16 \times \sqrt{\text{CEA}}) + (-1.4987 \times \sqrt{\text{CRP}})$, e is the base of natural logarithms. Calcification = 1 if calcification is present in the SPN (otherwise = 0); CT bronchus sign = 1 if CT bronchus sign is present (otherwise = 0); spiculation = 1 if spiculated appearance is present (otherwise = 0); CEA = serum CEA level (ng/mL); CRP = serum CRP level (mg/L). The area under ROC in the training set was 0.966 (Fig. 1). The goodness-of-fit statistic, as described by Hosmer-Lemeshow, for the derivation dataset was $\chi^2 = 11.608$, P-value = 0.170. If the biological parameters were not available, the alternative prediction model was as follows: Probability of benign SPN = $e^x/(1 + e^x)$, $x = 1.084 + (2.7851 \times \text{calcification}) + (-1.1795 \times \text{CT bronchus sign}) + (-5.4481 \times \text{spiculation})$. The area under ROC of the alternative prediction model in the training set was 0.94.

Following derivation of the prediction model, a further 148 patients with a newly discovered indeterminate SPN were identified between May 2004 and

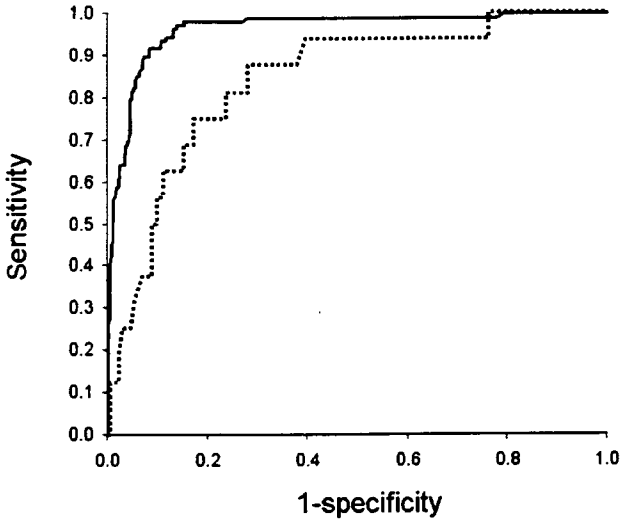

Figure 1 Receiver-operating characteristic curve for the prediction model for the diagnosis of a solitary pulmonary nodule (SPN) (solid line). The prediction model was developed based on data from 452 patients with a solitary pulmonary nodule. Receiver-operating characteristic curve for the prediction model when validated in 148 consecutive patients with SPNs (dotted line).

April 2005 (the validation set). The differences in the clinical and radiological characteristics of both sets are presented in Table 2.

Eighty-nine per cent of the validation set had malignant SPNs, and the reminder benign SPNs. Forty-seven per cent of the benign SPNs were granulomas; other diagnoses included sclerosing haemangioma, hamartoma, fibrosis, organizing pneumonia and inflammation. The majority of the malignant SPNs were adenocarcinoma of the lung (88%); other malignancies included squamous-cell carcinoma (7.4%), large-cell carcinoma (2.7%), pleomorphic carcinoma (0.7%), small-cell lung carcinoma (0.7%), and mucosa-associated lymphoid tissue (MALT) lymphoma of the lung (0.7%). The respective mean serum CEA and CRP levels in benign SPNs were 3.1 ng/mL (range 0.8–16.7) and 1.0 mg/L (range 1.0–3.0). The corresponding mean serum CEA and CRP levels in malignant SPNs were 5.1 ng/mL (range 0.5–58.4) and 3.0 mg/L (range 1.0–165). The mean diameters of the benign and malignant SPNs were 16 mm (range 10–23) and 17 mm (range 8–30), respectively.

In the validation set (n = 148), the area under ROC of the prediction model was 0.84 (Fig. 1), and the accuracies of the prediction model and the experienced chest radiologist were 0.858 and 0.905, respectively. ROC and accuracy of the alternative prediction model were 0.81 and 0.75. The experienced chest radiologist's prediction was inconsistent with 59% of benign SPNs, and 4% of malignant SPNs (3: adenocarcinoma; 1: pleomorphic carcinoma; 1: MALT of the lung).

DISCUSSION

This study has demonstrated a simple prediction model for the diagnosis of SPN based on two clinical variables and three radiological variables using thin-section CT findings. The accuracy of the prediction model was demonstrated to be high in a validation based on consecutive patients with undiagnosed SPNs, although not as high as the accuracy of the experienced chest radiologist's diagnostic prediction.

Statistical analysis indicated two biochemical markers for inclusion in the prediction model, but did not include the patient characteristics of age, cigarette-smoking status and past history of extra-pulmonary malignancy.[8] The frequency of a malignant SPN rises with increasing patient age. In the present study, age was significantly associated with a malignant SPN on univariate analysis. Several studies have reported the frequency of malignant SPN to be more than 50% higher in patients aged 50 years and over than in those aged less than 50 years.[13] Similarly, cigarette-smoking and past history of extra-pulmonary malignancy reportedly have strong associations with lung cancer or malignant SPNs. The training and validation sets included 9% and 6% of patients less than 50 years of age, respectively. Although the mean age and frequency of past history of extra-pulmonary malignancy in this study were similar to those in the previously reported prediction model, the frequency of cigarette smoking in this study was slightly lower than the 68% of patients with SPNs in the previous study.[9] According to data from Japan Tobacco Industry, Inc., the overall smoking prevalences in men and women were 57.5% and 14% in 1996 respectively, which are significantly different from the prevalence in the 1960s.[14] These patients' demographic characteristics varied from the frequencies in other SPN reports, possibly influenced by the role of the institution involved, the period or the patient population.

Serum markers based on tumour biology would not have been strongly influenced by patient population or institution. CEA is a marker for a wide range of malignancies, including lung cancer. Elevation of serum CEA was found in 16% to 27% of patients with clinical stage IA non-small-cell lung cancer.[15,16] Although the serum CEA level was associated with cigarette smoking and age,[17] the multivariate analysis used to develop the prediction model indicated the significance of serum CEA rather than these known confounding factors included in the previous prediction model.

CRP, an acute-phase protein synthesized in hepatocytes, is up-regulated by cytokines, such as IL-6 and tumour necrosis factor-α.[18,19] Increased serum CRP levels have also been recognized as part of a paraneoplastic syndrome for several malignant tumours, including lung cancer.[20] In addition, serum CRP was significantly higher in patients with lung cancer than in those with benign lung disease or the healthy population.[21–23] The laboratory variables in the prediction model were supported by these previous studies and our results; that is, the serum markers in patients with malignant SPNs were both higher than in those with benign SPNs. CT has been shown to be an effective means of morphologically differentiating benign from malignant SPNs. Numerous studies have described the radiographic characteristics of SPN,

Table 2 Comparison of the clinical and radiological characteristics in both training and validation patient sets

	Training set (n = 452)	Validation set (n = 148)	P-value*
Clinical characteristics			
Male (%)	56	57	0.69
Mean age (years)	62	63	0.82
Current or past smoker (%)	49	50	0.82
Other cancer >5 years ago (%)	10	11	0.60
Mean serum CEA (ng/mL)	4.9	4.9	0.99
Mean serum CRP (mg/L)	2.0	3.0	0.56
Radiological characteristics			
Mean diameter (mm)	19	17	<0.01
Side (right) (%)	64	61	0.63
Location (%)			
LUL	15	20	0.15
Lingular segment	5	2	0.13
LLL	15	16	0.73
RUL	31	32	0.74
RML	10	7	0.32
RLL	22	22	0.85
Calcification (%)	5	0	<0.01
Spiculation (%)	63	68	0.27
CT bronchus sign (%)	69	74	0.31

*Univariate analysis: chi-squared test and t-tests were performed for proportional differences or mean differences in variables between training and validation sets (two-sided).

CEA, carcinoembryonic antigen; LLL, left lower lobe; LUL, left upper lobe; RLL, right lower lobe; RML, right middle lobe; RUL, right upper lobe.

and each suggests, but does not guarantee, a diagnosis of benign versus malignant SPN. Spiculation, lobulation and vascular convergence are typically associated with malignancy.[23,24] Calcification, a well-defined margin, cavitation and CT bronchus signs including bronchus involvement show considerable overlap between benign and malignant SPNs.[24–28] The reported frequencies of these radiological findings have varied among studies. In the present study, the prediction model identified calcification, spiculation and CT bronchus sign as being useful for distinguishing benign from malignant SPNs. The validity of the present prediction model is supported by the previous prediction model and the typical radiographic features of malignant versus benign SPNs.

There were no differences in the laboratory or radiological results or in the pattern of care for SPN between the training set and the validation set. However, a limitation of the present study is the potential for selection bias, just as in previous studies. Selection bias has arisen because the model is based on a population with clinically suspicious lesions that required resection. It is unclear how this would affect the performance of the prediction model in a more general population. There were significant differences in two radiological variables (mean diameter and calcification) between the training set and the validation set. The prediction model would also need to be validated in other patient populations outside that of our single institution.

The best clinical management of a patient with an SPN requires evaluation of the probability of malignancy, which determines the most cost-effective diagnosis and treatment strategies. In several decision analyses, the probability of malignancy is divided into four categories: low (<10%), intermediate (10–60%), high (>60 to 90%) and very high (>90%).[29,30] The management pathway should take into consideration patient preference and the potential complications after assessment of the probability of malignancy. CT or CXR follow up is preferred in patients with a low probability of malignancy. When the probability is intermediate, additional testing would be required, including contrast-enhanced CT, fine-needle aspiration, bronchoscopy and PET/CT. Surgery is recommended in patients with a high or very high probability of malignancy. PET/CT has become important in differentiating benign from malignant nodules, and its estimated sensitivity for identifying a malignant process is 96%, and its specificity is 88%.[31] Currently, investigators have made efforts to develop a computer-aided diagnosis system for SPN. The computer-aided diagnosis system has the potential to improve diagnostic accuracy of SPN in future.[32]

Radiological examinations such as CT will detect a large number of SPNs. The physician can use the prediction model for differentiating between benign and malignant SPNs. Even if serum CEA and CRP cannot be measured, the alternative prediction model might assist in the evaluation of the probability of malignancy. The prediction model and particular observation of CT imaging for each patient would assist in determining the optimal management and in identifying the indications for invasive and

expensive procedures, such as surgery. Future studies should evaluate the accuracy of this prediction model in a larger cohort or non-Japanese cohort, and whether the clinical decisions made based on the prediction model actually improve clinical outcomes.

ACKNOWLEDGEMENTS

We thank Shingo Takano, Shun-ichi Watanabe, Kenji Suzuki, Yasuaki Arai, Noriyuki Moriyama and Ryosuke Tsuchiya (National Cancer Center Hospital, Tokyo) for their help in carrying out this study.

REFERENCES

1 Holin SN, Dwork RE, Glaser S, Rickli AE, Stocklen JB. Solitary pulmonary nodules found in a community-wide chest roentgenographic survey. *Am. Tuberc. Pulm. Dis.* 1959; **79**: 427–39.
2 Swensen SJ, Silverstein MD, Edell ES, Trastek VF, Aughenbaugh GL *et al.* Solitary pulmonary nodules: clinical prediction model versus physicians. *Mayo Clin. Proc.* 1999; **74**: 319–29.
3 Midthun DE, Swensen SJ, Jett JR. Approach to the solitary pulmonary nodule. *Mayo Clin. Proc.* 1993; **68**: 378–85.
4 Khouri NF, Meziane MA, Zerhouni EA, Fishman EK, Siegelman SS. The solitary pulmonary nodule: assessment, diagnosis, and management. *Chest* 1987; **91**: 128–33.
5 Swensen SJ, Jett JR, Payne WS, Viggiano RW, Pairolero PC *et al.* An integrated approach to evaluation of the solitary pulmonary nodule. *Mayo Clin. Proc.* 1990; **65**: 173–86.
6 Cummings SR, Lillington GA, Richard RJ. Estimating the probability of malignancy in solitary pulmonary nodules. A Bayesian approach. *Am. Rev. Respir. Dis.* 1986; **134**: 449–52.
7 Gurney JW. Determining the likelihood of malignancy in solitary pulmonary nodules with Bayesian analysis. *Radiology* 1993; **186**: 405–13.
8 Swensen SJ, Silverstein MD, Ilstrup DM, Schleck CD, Edell ES. The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules. *Arch. Intern. Med.* 1997; **157**: 849–55.
9 Baldwin DR, Birchall JD, Ganatra RH, Pointon KS. Evaluation of the solitary pulmonary nodule: clinical management, role of CT and nuclear medicine. *Imaging* 2004; **16**: 22–36.
10 Bates SE. Clinical application of serum tumor markers. *Ann. Intern. Med.* 1991; **115**: 623–85.
11 Ausitn JH, Muller NL, Friedmen PJ, Hansell DM, Naidich DP *et al.* Glossary of terms for CT of the lungs: recommendations of the Nomenclature Committee of the Fleischner Society. *Radiology* 1996; **200**: 327–31.
12 Seemann MD, Seemann O, Luboldt W, Bonel H, Sittek H *et al.* Differentiation of malignant from benign solitary pulmonary lesions using chest radiography, spiral CT and HRCT. *Lung Cancer* 2000; **29**: 105–24.
13 Trunk G, Gracey GR, Byrd RB. The management and evaluation of the solitary pulmonary nodule. *Chest* 1974; **66**: 236–9.
14 Health and Welfare Statistics Association. Tobacco. *J. Health Welfare Stat.* 1997; **44**: 101–4 (in Japanese).
15 Suzuki K, Nagai K, Yoshida J, Nishimura M, Nishiwaki Y. Predictors of lymph node and intrapulmonary metastasis in clinical stage IA non-small cell lung carcinoma. *Ann. Thorac. Surg.* 2001; **72**: 352–6.
16 Sawabata N, Ohta M, Takeda S, Hirano H, Okumura Y *et al.* Serum carinoembryonic antigen level in surgically resected clinical stage I patients with non-small cell lung cancer. *Ann. Thorac. Surg.* 2002; **74**: 174–9.
17 Alexander JC, Silverman NA, Chretien PB. Effect of age and cigarette smoking on carcinoembryonic antigen levels. *JAMA* 1976; **235**: 1975–9.
18 Castell JV, Gomez-Lechon MJ, David M, Fabra R, Trullenque R *et al.* Acute phase response of human hepatocytes: regulation of acute-phase protein synthesis by interleukin-6. *Hepatology* 1990; **12**: 1179–86.
19 Pepys MB, Hirschfield GM. C-reactive protein: a critical update. *J. Clin. Invest.* 2003; **111**: 1805–12.
20 Katsumata N, Eguchi K, Fukuda M, Yamamoto N, Ohe Y *et al.* Serum levels of cytokines in patients with untreated primary lung cancer. *Clin. Cancer Res.* 1996; **2**: 553–9.
21 Yanagawa H, Sone S, Takahashi Y, Haku T, Yano S *et al.* Serum levels of interleukin 6 in patients with lung cancer. *Br. J. Cancer* 1995; **71**: 1095–8.
22 Alexandrakis MG, Passam FH, Perisinakis K, Ganotakis E, Margantinis G *et al.* Serum proinflammatory cytokines and its relationship to clinical parameters in lung cancer patients with reactive thrombocytosis. *Respir. Med.* 2002; **96**: 553–8.
23 McKeown DJ, Brown DJF, Kelly A, Wallace AM, McMillan DC. The relationship between circulating concentrations of C-reactive protein, inflammatory cytokines and cytokine receptors in patients with non-small-cell lung cancer. *Br. J. Cancer* 2004; **91**: 1993–5.
24 Siegelman SS, Khouri NF, Leo FP, Fishman EK, Braverman RM *et al.* Solitary pulmonary nodules: CT assessment. *Radiology* 1986; **160**: 307–12.
25 Erasmus JJ, Connolly JE, McAdams HP, Roggli VL. Solitary pulmonary nodules: part I. Morphologic evaluation for differentiation of benign and malignant lesions. *Radiographics* 2000; **20**: 43–58.
26 Qiang JW, Zhou KR, Lu G, Wang Q, Ye XG *et al.* The relationship between solitary pulmonary nodules and bronchi: multi-slice CT-pathological correlation. *Clin. Radiol.* 2004; **59**: 1121–7.
27 Grewal RG, Austin JHM. CT demonstration of calcification in carcinoma of the lung. *J. Comput. Assist. Tomogr.* 1994; **18**: 867–71.
28 Takashima S, Sone S, Li F, Maruyama Y, Hasegawa M *et al.* Indeterminate solitary pulmonary nodules revealed at population based CT screening of the lung. *Am. J. Roentgenol.* 2003; **180**: 1255–63.
29 Decamp MM. The solitary pulmonary nodule: aggressive excisional strategy. *Semin. Thorac. Cardiovasc. Surg.* 1992; **14**: 292–6.
30 Gambhir SS, Shepherd JE, Shah BD, Hart E, Hoh CK *et al.* Analytical decision model for the cost effective management of solitary pulmonary nodules. *J. Clin. Oncol.* 1998; **16**: 2113–25.

31 Yi CA, Lee KS, Kim BT, Choi JY, Kwon OJ *et al.* Tissue characterization of solitary pulmonary nodule: comparative study between helical dynamic CT and integrated PET/CT. *J. Nucl. Med.* 2006; **47**: 443–50.

32 Li F, Aoyama M, Shiraishi J, Abe H, Li Q *et al.* Radiologists' performance for differentiating benign from malignant lung nodules on high-resolution CT using computer-estimated likelihood of malignancy. *Am. J. Roentgenol.* 2004; **183**: 1209–15.

Med Oncol (2007) 24:425–430
DOI 10.1007/s12032-007-0033-9

ORIGINAL PAPER

Prognostic factors for malignant pericardial effusion treated by pericardial drainage in solid-malignancy patients

Kan Yonemori · Hideo Kunitoh · Koji Tsuta · Tetsutaro Tamura · Yasuaki Arai · Yasuhiro Shimada · Yasuhiro Fujiwara · Yuko Sasajima · Hisao Asamura · Tomohide Tamura

Received: 20 April 2007 / Accepted: 1 May 2007 / Published online: 1 June 2007

Abstract

Purpose Malignant pericardial effusion is a frequent complication of advanced incurable malignancies and requires treatment. The purpose of this study was to identify prognostic factors for cytology-positive malignant pericardial effusion in patients treated by pericardial drainage.

Methods We retrospectively analyzed a series of consecutive patients diagnosed with cytologically positive malignant pericardial effusion who were treated by pericardial drainage at the National Cancer Center Hospital, Tokyo.

Results A total of 88 patients with pericardial effusion were treated by pericardial drainage, 60 patients were diagnosed with cytological positive malignant pericardial effusion including 32 with non-small cell lung cancer, 13 with breast cancer, 8 with gastrointestinal cancer, and 7 with miscellaneous cancers. Subxiphoid pericardiostomy was performed in 50 of the patients and percutaneous tube pericardiostomy in the other 10 patients. Malignant pericardial effusion recurred in 14 patients, and pericardial drainage was performed again in 9 of them. The median overall survival time was 6.1 months, and the 1-year survival rate was 28%. A multivariate analysis revealed the following significant negative prognostic factors: performance status, development of malignant pericardial effusion during chemotherapy, mediastinal lymph node enlargement, and cytologic type. ($P = 0.03$, 0.02, 0.01, 0.001, respectively).

Conclusion Patients with poor prognostic factors may be better to consider as indication of palliative therapy, even if oncologic emergency had been resolved rapidly by drainage.

Keywords Oncologic emergency · Malignant pericardial effusion · Drainage · Supportive care · Prognostic factor

K. Yonemori (✉) · H. Kunitoh · Y. Fujiwara · T. Tamura
Division of Medical Oncology, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan
e-mail: kyonemor@ncc.go.jp

K. Tsuta · Y. Sasajima
Division of Clinical laboratory, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan

T. Tamura
Division of Cardiology, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan

Y. Arai
Division of Diagnostic Radiology, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan

H. Asamura
Division of Thoracic Surgery, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan

Y. Shimada
Division of Gastrointesinal Oncology, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan

Introduction

Although metastatic involvement of the heart and pericardium is found at autopsy in 8.5–21% of patients who have died from a malignant tumor, it is detected in a much smaller percentage clinically [1, 2]. The most common cancer involving the pericardium are lung cancer, breast cancer, esophageal cancer, lymphoma, and leukemia [3, 4].

The majority of patients with pericardial effusion are asymptomatic and previous study showed that pericardial effusions less than 10 mm on an echocardiogram may be

an incidental finding and asymptomatic [2, 5]. The consequences of pericardial effusion mainly depend on the rate of exudation, compliance of the pericardium, and fluid volume. When malignant pericardial effusion (MPCE) occurs with symptoms of hemodynamic compromise either as an acute emergency or with insidious development of symptoms, prompt diagnosis, and treatment result in palliation and significant prolongation of survival time.

The overall survival time of MPCE patients has generally been short, ranging from 2 to 5.6 months [6–11]. The management and clinical course of MPCE depend on many factors, such as the underlying medical status of patient and the extent and type of the underlying malignant disease. To our knowledge there have been only a few reports describing prognostic factors in patients with MPCE [8–12]. The design of the studies may not have been entirely appropriate, however, because the analyses of prognostic factors included comparing patients with cytopathology-negative pericardial effusion or MPCE patients with patients with pericardial effusions of benign etiology [8–12]. Therefore, these study properly concluded patients with cancer-related or cytology positive MPCE was poor prognosis compared with patients with pericardial effusion of benign etiology [9, 11, 12].

In the present study we retrospectively analyzed the predictive factor for MPCE recurrence and the prognostic factors for cytology-positive MPCE in a series of consecutive solid-cancer patients treated with MPCE by pericardial drainage.

Patients and methods

Patients

A total of 88 patients with pericardial effusion were treated by pericardial drainage (subxiphoid pericardiostomy or percutaneous pericardiostomy) between February 1998 and November 2005 at the National Cancer Center Hospital, Tokyo. Of these, 60 patients were diagnosed with cytology-positive MPCE. The baseline investigation of the patients included a history and physical examination, chest computed tomography, and electrocardiography. Trans-thoracic echocardiography was performed by a cardiologist (T.T.).

Subxiphoid pericardiostomy was performed by making a small vertical incision extending 4–6 cm inferiorly from the xiphoid process. The anterior pericardium was identified, and a piece of pericardium 2–4 cm in diameter was removed. The pericardial sac was drained with a 20–24 Fr silicone tube. Percutaneous tube pericardiostomy was performed by inserting an 8 Fr tube percutaneously via a subxiphoid approach into the pericardial sac and then using the Seldinger technique. Both procedures were performed under local anesthesia. The tubes were positioned to allow drainage by gravity, and they were suctioned or flushed as necessary. The pericardial effusion obtained was routinely submitted for bacterial culture, the results were negative in every case. When the volume of drained pericardial fluid decreased to 25 ml or less per day, the drainage tube was removed. Sclerotherapy by pericardial instillation of a sclerosing agent was performed at the discretion of the attending physician.

Evaluation and statistical analysis

Recurrence of MPCE was defined as progression based on radiographic or echocardiographic examination after removing pericardial drainage tube. Overall survival time was measured from the first day of pericardial drainage until death.

We analyzed predictive factors for recurrence of MPCE by univariate analysis (Pearson chi-square test/Fisher exact test). The clinical variables before pericardial drainage were chosen by considering possible factors based on previous studies and our experience [8–12]. At the time of the pericardial drainage for MPCE, 12 categorical pretreatment variables were selected for statistical analysis: gender (male versus female), age (<60 versus 60 ≤), disease (lung cancer versus other), cytologic type (adenocarcinoma versus other), performance status (≥3 versus 3>), number of metastatic site (≥4 versus 4>), mediastinal lymph node enlargement (present versus absent), pleural effusion (present versus absent), MPCE at the time of diagnosis (present versus absent), malignancy recurred in MPCE (yes versus no), occurrence of MPCE during chemotherapy (yes versus no), and any prior history of chemotherapy (yes versus no). Median overall survival was estimated by the Kaplan–Meier method, and the survival analyses were assessed by the log-rank test. Cox's proportional hazards model according to a stepwise procedure was used to evaluate prognostic factors that were significantly related to survival. The statistical analyses were performed with SPSS 12.0 J software (SPSS Inc., Chicago, IL), and differences were considered significant at a P value <0.05 (two-sided).

Results

The background of the 60 patients is summarized in Table 1. The underlying disease was non-small cell lung cancer in 32 patients, breast cancer in 13 patients, esophageal cancer in 6 patients, small-cell lung cancer, and gynecological cancer in 2 patients each, and gastric cancer,

Table 1 Patient backgrounds

No. of Patients	60
Male/female	30/30
Median age (range)	58 (32–72)
Performance status	2 (1–4)
Disease	
Non-small cell lung cancer	32
Breast cancer	13
Esophageal cancer	6
Small cell lung cancer	2
Gynecological cancer	2
Gastric cancer	1
Colorectal cancer	1
Cholangiocellular cancer	1
Cancer, unknown primary	1
Melanoma	1
Cytologic type	
Adenocarcinoma	48
Squamous cell carcinoma	9
Small-cell carcinoma	2
Melanoma	1
Median number of organs involved (range)	4 (2–8)
Site of disease	
Mediastinal lymph node enlargement (subcarina)	48 (29)
Pleural effusion (bilateral)	40 (21)
Clinical manifestation at presentation	
Dyspnea	52
Fatigue	50
Cough	21
Hypotension	16
Edema	4
Chest pain	3
Abnormal ECG (low voltage, or tachycardia)	19
Median interval between the visceral and parietal pericardium (range)	
Echocardiography	22 mm (12–48)
Computed tomography	15 mm (10–25)

colorectal cancer, cholangiocellular cancer, melanoma, and primary unknown cancer in 1 patient each. About 14 patients had developed MPCE at the time of the diagnosis of their malignant disease, and all 14 of those patients were diagnosed with primary lung cancer. MPCE occurred in 7 patients as initial site of recurrence (5 patients with lung cancer, and 1 patient each with breast cancer and esophageal cancer), and the median interval until recurrence was 28.9 months (range 6.5–171 months). About 21 patients, including one patient treated with concurrent chemoradiation therapy had a prior history of chemotherapy. And 11 patients experienced disease progression in the form of MPCE during systemic chemotherapy, including during concurrent chemoradiation therapy in two patients. Seven patients developed MPCE during supportive care after their disease had recurred in other sites.

Around 50 patients underwent subxiphoid pericardiostomy, and percutaneous tube pericardiostomy was performed in the remaining 10 patients. No patients died as a direct result of either procedure. The median total volume of fluid drained was 953 ml (range 200–3970 ml). The median duration of drainage was 6 days (range, 1–33 days). There was no statistical difference in time to recurrence of MPCE or survival between patients undergoing tube pericardiostomy and subxiphoid pericardiostomy. ($P = 0.69$, $P = 0.42$, respectively) Among the 13 patients in whom sclerotherapy was performed, bleomycin was used as the agent in 10 patients, OK-432 in 2 patients, and minomycin in one patient. There was no statistical difference in time to recurrence of MPCE or survival in patients with or without sclerotherapy ($P = 0.71$, $P = 0.48$, respectively). Complications during pericardial drainage included atrial fibrillation in 5 patients, premature ventricular contractions in 3 patients, and chest pain in one patient.

MPCE recurred in 14 patients (6 with breast cancer, 6 with lung cancer, and one each with esophageal cancer and melanoma), including in 2 patients who underwent bleomycin sclerosis. The median interval until recurrence was 4.3 months (range 0.25–12 months), and pericardial drainage was performed in 9 of them (6 with breast cancer and 3 with lung cancer). Subxiphoid pericardiostomy was used again in 7 patients, and percutaneous tube pericardiostomy was used in the other 2 patients. Only one patient developed a second recurrence 7 weeks after removing the tube, and subxiphoid pericardiostomy was used to treat it. The median overall survival was 6.1 months, and the 1-year survival rate was 28% (range 0.25–48.4 months Fig. 1). The hospital mortality rate was 12% (7 of 60 patients), and the primary cause of death was the underlying malignancy in 6 patients and atrial perforation during subxiphoid pericardiostomy for recurrence of MPCE.

Table 2 shows the results of the univariate analysis for a relationship between the pretreatment variables, recurrence of the MPCE, and overall survival. There was no predictive factor for recurrence of MPCE. The Multivariate analysis yielded the following prognostic factors: performance status (hazard ratio, 2.1; 95%CI, 1.1 to 4.2; $P = 0.03$), occurrence of MPCE during chemotherapy (hazard ratio, 2.3; 95%CI, 1.1–4.7; $P = 0.02$), mediastinal lymph node enlargement (hazard ratio, 3.3; 95%CI, 1.3–8.1; $P = 0.011$), and cytologic type (hazard ratio, 3.1; 95%CI, 1.5–6.3; $P = 0.001$).

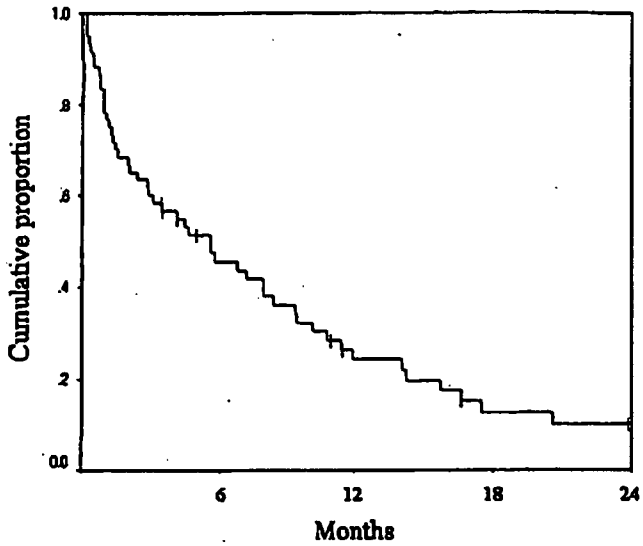

Fig. 1 Kaplan–Meier analysis of overall survival

Discussion

In the present study, we retrospectively analyzed cytology-positive MPCE cases, and the median survival time tended to be longer than in the previous studies. The prognostic value of pericardial fluid cytology was a matter of controversy in previous studies [8, 9, 12]. Since pericardial effusion is frequently caused by a cytopathology-negative cancer-related etiology, such as post-radiation therapy, post-bone marrow transplantation, and post-thoracic surgery [6, 12], the diagnosis based on cytopathologic confirmation is important in survival and prognostic analysis.

The pathways to the pericardium is considered: lymphatic, hematogeneous metastasis, direct tumor invasion, and retrograde lymphatic migration of tumor cells which involves mediastinal lymph nodes [13, 14]. Although it is difficult to distinguish direct invasion from retrograde lymphatic migration on CT scan images, the mediastinal lymph nodes, including the subcarina lymph node, were most common site of metastasis in this study. The presence of mediastinal lymph node enlargement was identified as a poor prognostic factor, and it may have contributed to obstruction of lymphatic drainage and retrograde tumor spread.

MPCE is uncommon in the absence of any other metastatic disease or as the primary manifestation of malignancy [2, 15, 16]. One fifth of the patients in our study had MPCE as the primary manifestation of malignancy, and all of them were diagnosed with primary lung cancer. The variables such as MPCE at the time of initial diagnosis or initial recurrence caused by MPCE, were not poor prognostic factors, patients with these rare manifestations have indications for systemic anticancer therapy for their cancer after the MPCE drainage procedures. On the other hand, MPCE during chemotherapy was found to be a poor prognostic factor in this study. This factor means MPCE developed with chemotherapy-resistant condition and it may be essentially severe visceral crisis, even if cardiac tamponade was rescued by drainage.

The efficacy of systemic anticancer therapy in controlling MPCE after the initial therapeutic pericardiocentesis, was 67.4% in the previous report (71.1% in breast cancer, 33.3% in other solid tumors, 100% in lymphomas) [16]. Performance status thought to be generally related with survival in patients treated with chemotherapy and present study indicated poor performance status was one of prognostic factors. Therefore, patients with favorable performance status may have a chance to achieve good condition with longer survival by receiving anticancer therapy after pericardial drainage.

Symptomatic MPCE can be effectively relieved by several procedures, but they need to be individualized based on consideration of both the underlying malignant disease and medical status of patient. Simple pericardiocentesis may be life-saving in emergency cases of cardiac tamponade. Although medical, interventional, or surgical treatment is recommended in almost all patients because of the high rate of recurrence, whether partial pericardiectomy, subxiphoid pericardiostomy, percutaneous balloon tube pericardiostomy, percutaneous tube pericardiostomy with or without sclerotherapy, radiotherapy, or systemic chemotherapy is the optimal method of management of MPCE has been a matter of controversy [17].

Several investigators retrospectively compared procedures [12, 18, 19]. Due to their short-life span, McDonald et al. [18] recommended percutaneous tube pericadiostomy for MPCE patients with positive-cytology. Although one patient in our study died, because of atrial perforation during subxiphoid pericardiostomy for MPCE recurrence, subxiphoid pericardiostomy was considered to have-high success rate with an acceptable risk [6, 9–11, 20].

Although the present study showed no statistical difference for clinical outcome in patient with or without sclerotherapy, the efficacy of sclerotherapy is unknown in patient with MPCE [21]. To obtain clear evidence of the effectiveness of sclerotherapy for the management of MPCE, we joined Japan Clinical Oncology Group (JCOG) 9812 trial, a randomized trial comparing instillation of bleomycin with no instillation of a sclerosing agent, and we hope that the results of the JCOG 9812 trial will be conclusive.

Conclusions

Pericardial drainage should be performed for immediate symptom relief in patients with MPCE, and it makes it

Table 2 Univariate analysis of pre-treatment variables for recurrence of MPCE and overall survival

Variables	No. of pts	No. of MPCE recurrence	*P* value*	Median survival time (months)	*P* value**
Gender					
Male	30	6	0.54	3.4	<0.01
Female	30	8		12.5	
Age (years)					
60 ≤	28	7	0.78	6.1	0.9
60>	32	7		4.9	
Disease					
Lung cancer	34	6	0.23	5.1	0.68
Other	26	8		6.1	
Cytologic type					
Adenocaricinoma	48	10	0.45	7.9	<0.01
Other	12	4		1.25	
Performance status					
3 ≤	15	3	0.99	1.4	0.21
3>	45	11		5.9	
No. of metastatic site					
≤	23	4	0.53	3.2	0.44
4>	37	10		8.6	
Mediastinal lymph node enlargement					
Present	48	10	0.45	3.4	<0.01
Absent	12	4		22.4	
Pleural effusion					
Present	40	9	0.99	4.2	0.06
Absent	20	5		8.1	
MPCE at the time of diagnosis					
Present	14	3	0.99	4.2	0.75
Absent	46	11		5.7	
MPCE was initial site of recurrence					
Yes	7	1	0.99	13	0.11
No	53	13		4.5	
MPCE occurred during chemotherapy					
Yes	11	1	0.43	1.3	<0.01
No	49	13		7.0	
After history of chemotherapy					
Yes	21	6	0.58	5.9	0.34
No	39	8		4.2	

* Statistical analyses was performed by Pearson chi-square test/Fisher exact test

** Statistical analysis was performed by log-rank test

possible to prolong survival by preventing sudden life-threatening hemodynamic failure. Patients without poor prognostic factors may be recommended to consider further indications for systemic anticancer therapy.

Acknowledgement Authors thank to Hajime Uno, (Division of Biostatistics, Kitasato University Graduate School), Shunichi Watanabe, Kenji Suzuki (Division of Thoracic Surgery, National Cancer Center Hospital), Tsutomu Kouno, Hiroshi Nokihara, Ikuo Sekine, Yuichiro Ohe (Division of Medical Oncology, National Cancer Center Hospital), and Yoshihiro Matsuno (Division of Clinical Laboratory, National Cancer Hospital) for their assistance.

References

1. Theologides A. Neoplastic cardiac tamponade. Semin Oncol 1978;5:181–2
2. Thurber D, Edwards J, Anchor R. Secondary tumors of the pericardium. Circulation 1962;26:228–41

3. Abraham KP, Reddy V, Gattuso P. Neoplasms metastatic to the hearts: review of 3314 consecutive autopsies. Am J Cardiovasc Pahtol 1990;3:195–8
4. Wilkes JD, Fidias P, Vaickus L, Prez RP. Malignancy-related pericardial effusion. 127 cases from the Rosewell Park Cancer Institute. Cancer 1995;76:1377–87
5. Savage DD, et al. Prevalence and correlates of posterior extra echocardiographic spaces in a free-living population based sample. Am J Cardiol 1998;51:1207–12
6. Shepherd FA, et al. Medical management of malignant pericardial effusion by tetracycline sclerosis. Am J Cardiol 1987;60:1161–6
7. Maher EA, Shepherd FA, Todd TJR. Pericardial sclerosis as the primary management of malignant effusion and cardiac tamponade. J Thoracic Cardiovasc Surg 1996;112:637–43
8. Wang PC, et al. Prognostic role of pericardial fluid cytology in cardiac tamponade associated with non-small cell lung cancer. Chest 2000;118:744–9
9. Cullinane CA, Paz IB, Smith D, Carter N, Grannis FW Jr. Prognostic factors in the surgical management of pericardial effusion in the patient with concurrent malignancy. Chest 2004;125:1328–34
10. Okamoto H, Shinkai T, Yamakido M, Saijo N. Cardiac tamponade caused by primary lung cancer and the management of pericardial effusion. Cancer 1993;71:93–8
11. Allen KB, Faber LP, Warren WH, Shaar CJ. Pericardial effusion: subxiphoid pericardiostomy versus percutaneous catheter drainage. Ann Thoracic Surg 1999;67:437–40
12. Gornik HL, Gerhard-Herman M, Beckman JA. Abnormal cytology predict poor prognosis in cancer patients with pericardial effusion. J Clin Oncol 2005;23:5211–6
13. Tamura A, et al. Cardiac metastasis of lung cancer: a study of metastatic pathways and clinical manifestations. Cancer 1992;70:437–42
14. Hanfling SM. Meatstatic cancer to the heart. Circulation 1960;22:474–83
15. Fraser RS, Viloria JB, Wang NS. Cardiac tamponade as a presentation of extracardiac malignancy. Cancer 1980;45:1697–704
16. Buck M, Ingle JN, Giuliani ER, Gordon JR, Therneau TM. Pericardial effusion in woman with breast cancer. Cancer 1987;60:263–9
17. Nguyen DM, Schrump DS. Malignant pericardial effusion. In: DeVita VT, Hellman S, Rosenberg SA, editors. Cancer, principles and practice of oncology. 7th ed. Philadelphia, PA: Lippincott Williams and Willkins; 2005. pp. 2387–2392
18. McDonald JM, et al. Comparison of open subxiphoid pericardial drainage with percutaneous catheter drainage for symptomatic pericardial effusion. Ann Thorac Surg 2003;76:811–5
19. Park JS, Rentschler R, Wilbur D. Surgical management of pericardial effusion in patients with malignancies. Comparison of subxiphoid window versus pericardiectomy. Cancer 1991;67:76–80
20. Vaitkus PT, Herrmann HC, LeWinter MM. Treatments of malignant pericardial effusion. JAMA 1994;272:59–64
21. Liu G, Crump M, Goss PE, Dancey J, Shepherd FA. Prospective comparison of the sclerosing agents doxycycline and bleomycin for the primary management of malignant pericardial effusion and cardiac tamponade. J Clin Oncol 1996;14:3141–7

Japanese Journal of Biometrics Vol. 28, No. 1, 37–46 (2007)

Original Article

The Q-Q Plot of p-values for Predicting Outcomes with the Gene Expression Data

Yoichi M. Ito[*1], Yasuhiro Fujiwara[*2] and Yasuo Ohashi[*1]

[*1]Department of Biostatistics/Epidemiology and Preventive Health Sciences, School of Health Sciences and Nursing, University of Tokyo

[*2]Breast and Medical Oncology Division, Department of Medical Oncology, National Cancer Center Hospital

ito@epistat.m.u-tokyo.ac.jp

Michiels et al. (2005) showed that a list of genes identified as predictors of prognosis via a non-repeated training – validation approach is unstable and advocate the validation by repeated random sampling. They considered that the genes which were selected as top 50 genes in more than half of their jackknife samples were stable for prediction. However, there is no rationale of the determination of the length of the gene list and the threshold of stability. Since evaluating an accumulation of low p-values in the repeated random sampling is essentially required for a stability assessment, it is better to compare the distribution of p-values of a gene observed with the distribution of p-values under the null hypothesis directly. In this study, the Quantile-Quantile plot (Q-Q plot) of p-values with null reference was proposed for this purpose. We applied the proposed method to a clinical data for primary breast cancer. The Q-Q plot approach can reveal that the genes with a similar p-value in the ordinary analysis have different p-value distributions in the repeated random sampling, and the gene with low p-values accumulated in the repeated random sampling could be evaluated according to the reference lines in the Q-Q plot.

Key words: bootstrap sampling, gene expression analysis, validation, Q-Q plots.

1. Introduction

Recently, cDNA microarray technology development produces an enormous amount of gene expression data. DNA microarrays are assays for quantifying the types and amounts of mRNA transcripts present in a collection of cells. The number of mRNA molecules derived from transcription of a given gene is an approximate estimate of the level of expression of that gene. In the proof of concept study, it is important to explore relationship between gene expressions in cancer cells and clinical response to the drug, because it reveals signal pathways affected by the drug in cancer cells.

When we construct the model for predicting clinical response, selection of the important

Received March 2007. Revised May 2007. Accepted June 2007.

genes for inclusion in the model must be needed because the sample size is usually limited in contrast to the huge number of genes. This is called feature selection, which is a common first step when constructing the model for predicting clinical response based on microarray data (Simon et al., 2003). It is generally reasonable to assume that only some subset from many measured genes contribute useful information for prediction. An approach to feature selection, therefore, is to select genes according to their statistical significance in the univariate regression analysis.

To estimate the accuracy of a feature selection based on the statistical significance, the standard strategy is via a training – validation approach, in which a training set is used to identify the candidate genes and a validation set is used to estimate the accuracy of prediction. Michiels et al. (2005) reanalyzed the data of published works about prediction of cancer outcome with microarrays by the delete-d jackknife sampling (Efron and Tibshirani, 1993). They revealed that the list of genes identified as predictors of prognosis was highly unstable; molecular signatures (based on one training set and the accuracy evaluated in one validation set) strongly depended on the selection of patients in the training sets. Therefore they advocate the use of validation by repeated random sampling.

Their validation approach was described as follows. First, the dataset (size N) with a binary outcome (dead or alive) was divided using the delete-d jackknife sampling into 500 training sets (size n) with $n/2$ patients having each outcome, and 500 associated validation sets (size $N-n$). Second, they identified a molecular signature for each training set and estimated the proportion of misclassifications for each associated validation sets. For a given training set, the molecular signature was identified as the 50 genes for which expression was most highly correlated with prognosis as shown by Pearson's correlation coefficient. They saw that the list of 50 genes that had the highest correlations with outcome was very unstable. For instance, with data by van't Veer and colleagues (van't Veer et al., 2002) and a training set of the same size as in original publication ($n = 78$), only 14 of 70 genes from the published signature were included in more than half of their 500 signatures. Also, ten genes not included in more than 250 of their signature.

They considered that the genes which were selected as top 50 genes in more than half of their jackknife samples were stable for prediction. However, there is no rationale of the determination of the length of the gene list and the threshold of stability. Since evaluating an accumulation of low p-values in the repeated random sampling is essentially required for a stability assessment, it is better to compare the distribution of p-values of a gene observed with the distribution of p-values under the null hypothesis directly.

The Q-Q plot is one of the useful methods for comparing the observed distribution of p-values to the null distribution. Wartenberg and Northridge (1991) proposed a method with the Q-Q plot to describe and summarize case-control data for exploratory analyses of disease-exposure relations. Holmgren (1995) discussed the properties of the probability-probability plot and the Q-Q plot as a description of treatment effect in a randomized controlled trials. Ito and Ohashi

(2001) proposed the theoretical Q-Q plot of the observed distribution of p-values for exploring the association between the gene expression and the pharmacokinetic parameter (Cmax) in the proof of concept study. Their theoretical Q-Q plot was, however, assumed the independence of p-values, observed p-values of regression with same Cmax variable are correlated with each other. Theoretical distribution based on the assumption of independence of the tests is inappropriate for the reference distribution of the observed distribution of p-values. In this study, we propose the visualization method of p-value distribution with the reference lines based on the normal random numbers, and apply to prediction of the clinical response with gene expression in cancer cell. In the next section, we present the proposed method. In Section 3, we describe the background of applied clinical data, and the proposed method is applied to the data. Finally, Section 4 provides some discussion.

2. The Q-Q plot of p-values with null reference

Wilk and Gnanadesikan (1968) proposed the empirical Q-Q plot for comparison of two one-dimensional samples. This is the useful method for visualizing the discrepancy of the observed distribution from the theoretical distribution. It is also useful for visualization of p-value distributions in the cross validation data sets.

Figure 1 is a diagram of the bootstrap sampling algorithm for Q-Q plot of p-values with null reference. First, we assumed that there is a dataset of size n which include one clinical response and g gene expressions. And the m sets of independent normal random numbers with size n are generated and merged to the original dataset. Second, b bootstrap samples are sampled from the merged dataset, and we consider the following model for the response:

$$g[E(y_i)] = \beta_{0j} + \beta_{1j} x_{ij}, \qquad i = 1,\ldots,n, \quad j = 1,\ldots,g, \tag{1}$$

where x_{ij} is the jth gene expression of ith subject, β_{0j} and β_{1j} are the intercept and slope parameter for jth gene respectively, and the function $g[\,]$ is the link function between the expectation of the response $E(y_i)$ and the linear predictor with the gene expression x_{ij}. The model (1) is applied to the b bootstrap samples, and b p-values for each of g genes about the slope parameter are estimated under the null hypothesis $\beta_{1j} = 0$. Third, we consider the following model for m sets of normal random numbers:

$$g[E(y_i)] = \gamma_{0k} + \gamma_{1k} z_{ik}, \qquad i = 1,\ldots,n, \quad k = 1,\ldots,m, \tag{2}$$

where $z_{ik} \sim N(0,\sigma^2)$, γ_{0k} and γ_{1k} are the intercept and slope parameter for kth random number respectively. The standard deviation of random numbers σ could be determined based on the distribution of gene expression in the applied data. The model (2) is applied to the b bootstrap samples, and b p-values for each of m sets of random numbers about the slope parameter are estimated under the null hypothesis $\gamma_{1k} = 0$. Even though two sets of random numbers are

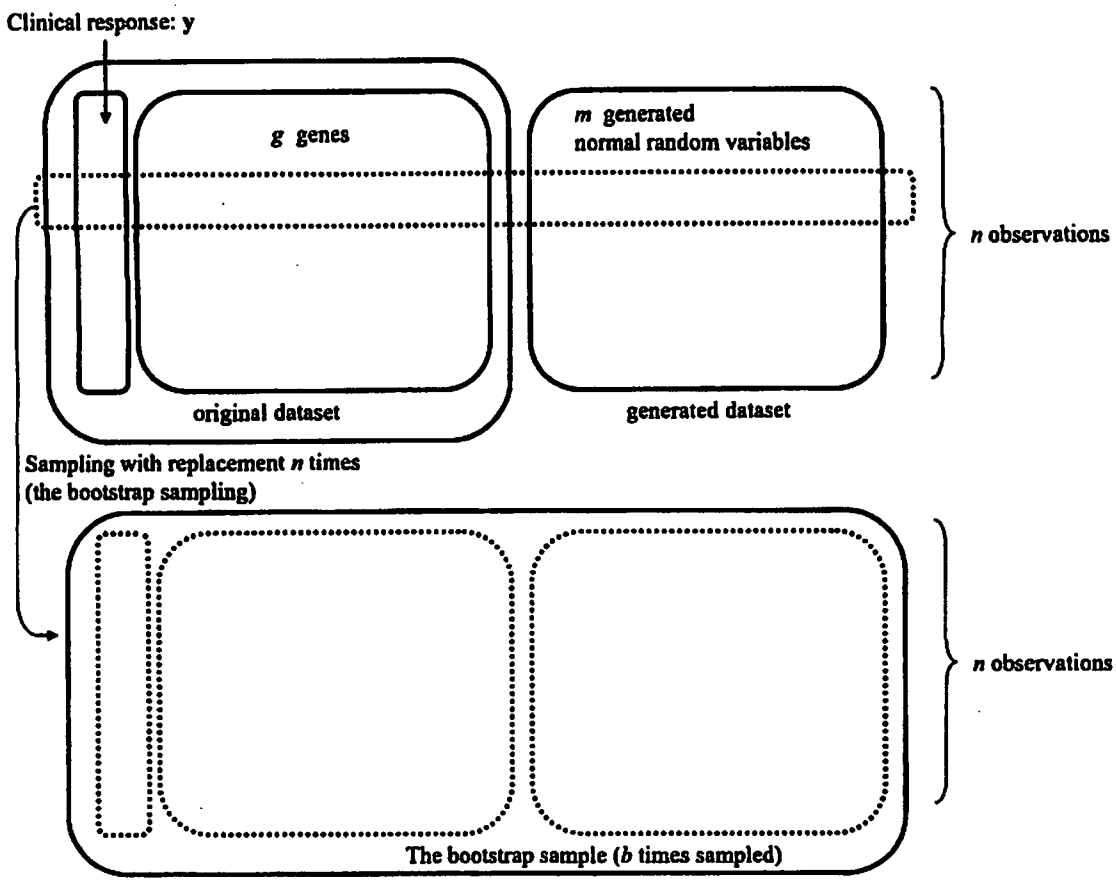

Fig. 1. The diagram of the bootstrap sampling algorithm for the Q-Q plot of p-values with null reference.

generated independently, calculated p-values for prediction are correlated among the bootstrap samples because the same outcome variable is predicted. To accommodate the correlation of p-values, m sets of random numbers must be fixed before the bootstrap sampling. Finally, we obtain g empirical distributions of p-values for the gene of size b and m empirical distributions of p-values for random numbers of size b. The negative logarithms of p-values are calculated for stressing on the region of low p-values.

When we sort the negative logarithms of p-values for each m sets of random numbers descendingly, we can obtain rth $(r = 1, \ldots, b)$ order statistic. There are m replicates of rth $(r = 1, \ldots, b)$ order statistic, then we can consider p percentile of rth $(r = 1, \ldots, b)$ order statistic. We plot p percentile of rth order statistic on y-axis and the median of rth order statistic on x-axis, and join these points. This Q-Q plot is called the p percentile of null reference in our approach.

Similarly, the negative logarithms of p-values for a gene are sorted in descending order, we can obtain rth $(r = 1, \ldots, b)$ order statistic for the gene. We plot the rth $(r = 1, \ldots, b)$ order statistic for the gene on y-axis and the median of rth $(r = 1, \ldots, b)$ order statistic for random numbers on x-axis. Comparing the Q-Q plot for the gene with each p percentile of null reference,

we can know whether the p-values of the gene are smaller than the expected values under the null hypothesis.

If the Q-Q plot is near to the upper left corner of the plot, there are more lower p-values than expected. Therefore we consider the area under the curve (AUC) as the index of the discrepancy from the null distribution. AUC is calculated by the trapezoidal method. If the AUC of a gene is larger than the AUC of 95th percentile of null reference, then the probability of obtaining such a large AUC can be found less than 5% under the null hypothesis.

The value of m and the bootstrap sample size b are tuning parameters for Q-Q plots. Large values of m produce more precise estimates of the p percentile of null reference and large values of b produce smoother Q-Q plots, but this further increases the computational burden because p-values must be calculated $b \times m$ times for the null reference and $b \times g$ times for the Q-Q plots for the genes. To assess the precision of the p percentile of null reference, we consider the confidence interval of p percentile (Altman et al., 2000). For the p percentile, first calculate the following quantities:

$$l_p = mq - Z_{1-\alpha/2} \times \sqrt{mq(1-q)} \quad \text{and} \quad u_p = 1 + mq + Z_{1-\alpha/2} \times \sqrt{mq(1-q)} \tag{3}$$

where $q = p/100$ and $Z_{1-\alpha/2}$ is the $100(1-\alpha/2)$ percentile of the standard normal distribution. Then round l_p and u_p to the nearest integers. The l_pth and u_pth observations in the sorted m replicates of rth $(r = 1, \ldots, b)$ order statistic are the $100(1-\alpha)\%$ confidence limits for the p percentile of the null reference. It is preferable that the confidence interval of a percentile dose not include the confidence limit of the adjacent percentile and the boundary of the observation list (1st and mth observations). For example, if we consider the 95% confidence interval of the 99th and 95th percentiles of null reference, then $l_{99} > u_{95}$ and $u_{99} \leq m$ are required. These yield $m \geq 298$ and $m \geq 476$ respectively. In practice, the value of m should be larger than these figures.

3. Applied Example

We applied the proposed method to phase II study data of neoadjvant 5FU/epirubicin/cyclophosphamide followed by weekly paclitaxel for primary breast cancer (Shimizu et al., 2005). The number of patients who is evaluated about pathological response and gene expression in both cancer cells and normal peripheral mononuclear cells are 34. The frequency of pathological response was as follows; grade 1a or 1b was 20, grade 2 was 5 and grade 3 was 9. Gene expression was measured by Custom ATLAS^{TM}Array (BD Biosciences Clontech) which can measure 988 genes at once.

We construct a prediction model for pathological response with gene expressions. Because pathological response is ordinal, we consider the following proportional odds model:

$$\log \frac{\pi_c}{1-\pi_c} = \beta_{0cj} + \beta_{1j} x_{ij}, \qquad i = 1, \ldots, 34, \quad j = 1, \ldots, 988, \quad c = 1, 2, 3 \tag{4}$$

Table 1. The top 20 genes with the lowest p-values according to the proportional odds model in the original data.

No	Gene ID	Parameter Estimate	Standard Error	Wald Chi Square	Wald p-value	Likelihood Ratio Chi Square	Likelihood Ratio p-value
1	Gene949	1.98	0.78	6.41	0.011	7.52	0.006
2	Gene218	2.63	1.16	5.16	0.023	5.99	0.014
3	Gene416	1.88	0.88	4.53	0.033	5.79	0.016
4	Gene26	1.57	0.68	5.33	0.021	5.67	0.017
5	Gene900	0.87	0.40	4.73	0.030	5.42	0.020
6	Gene449	1.46	0.65	4.97	0.026	5.37	0.021
7	Gene38	1.99	0.92	4.72	0.030	5.32	0.021
8	Gene592	1.19	0.59	4.01	0.045	4.78	0.029
9	Gene55	1.18	0.58	4.10	0.043	4.75	0.029
10	Gene437	1.63	0.81	3.99	0.046	4.72	0.030
11	Gene965	1.79	0.93	3.72	0.054	4.14	0.042
12	Gene782	-0.35	0.19	3.35	0.067	3.79	0.052
13	Gene75	1.18	0.63	3.51	0.061	3.76	0.053
14	Gene799	0.81	0.45	3.23	0.072	3.61	0.057
15	Gene860	1.78	0.97	3.35	0.067	3.46	0.063
16	Gene418	0.99	0.57	2.97	0.085	3.45	0.063
17	Gene859	0.67	0.39	2.84	0.092	3.44	0.064
18	Gene749	1.19	0.70	2.88	0.090	3.29	0.070
19	Gene883	0.58	0.38	2.32	0.128	3.28	0.070
20	Gene750	1.35	0.77	3.12	0.077	3.27	0.070

where $\pi_C = P(Y \leq c \mid x_{ij})$, c is the grade category of pathological response ($c = 1$ for grade 1a or 1b, 2 for grade 2 and 3 for grade 3).

Table 1 shows the top 20 genes with the lowest p-values according to the proportional odds model in the original data. The likelihood ratio p-values of 11 genes are less than 5%. Estimated odds ratio indicates a possibility of prediction for clinical response.

Figure 2 shows the null references and Q-Q plots of p-values of m sets of normal random numbers. The standard deviation of normal random numbers was determined at 0.3 based on the distribution of gene expression in the applied data. The value of m and the number of bootstrap sample b were both set at 1000. In Figure 2(a), the horizontal axis represents the median of rth ($r = 1, \ldots, b$) order statistic. The broken Q-Q plots represent the 1st, 5th, 10th, 25th, 50th, 75th, 90th, 95th, and 99th percentile of null reference and the values in parenthesis besides each curves are its AUC. When we consider Q-Q plots of each m sets of normal random numbers, we can obtain m Q-Q plots and AUCs. In Figure 2(b), the solid Q-Q plots represent Q-Q plots for the m sets of normal random numbers with corresponding percentile according to the AUC values. The AUC is given in parenthesis besides each curves. This figure shows the solid lines almost follow the broken lines, this implies the AUC is appropriate for the index of the discrepancy from the null distribution.

Figure 3 shows Q-Q plots of p-values with null reference for the top 3 genes according to the AUC values. The horizontal line in each Q-Q plot indicates the p-value in the original data. Left Q-Q plot with the highest AUC passes along between the 99th percentile and the 95th percentile of null reference, and then passes above the 99th percentile of null reference. This implies that

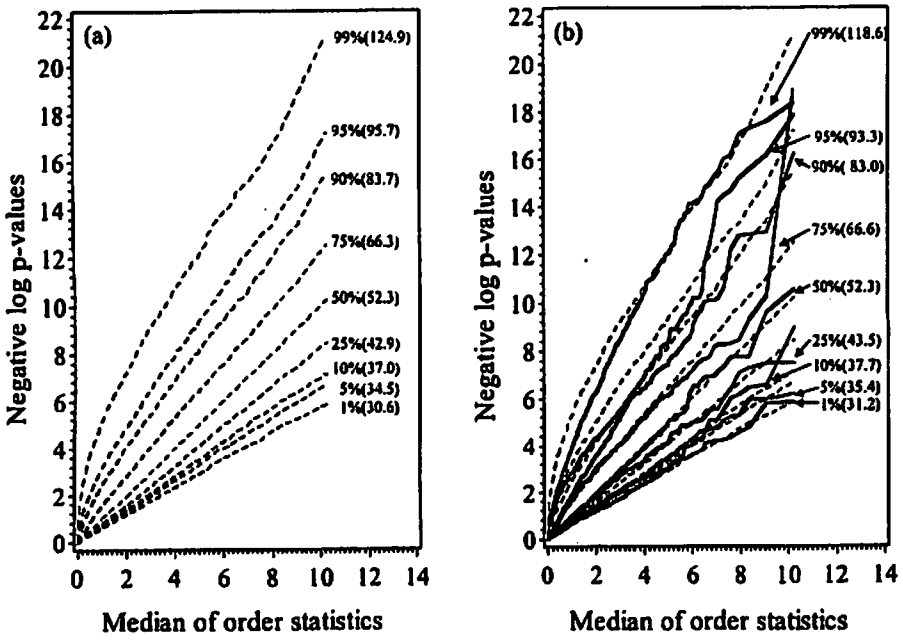

Fig. 2. The null references and Q-Q plots of p-values of m sets of normal random numbers. The values besides each curves indicate the corresponding percentile and AUC of the Q-Q plot in parenthesis.

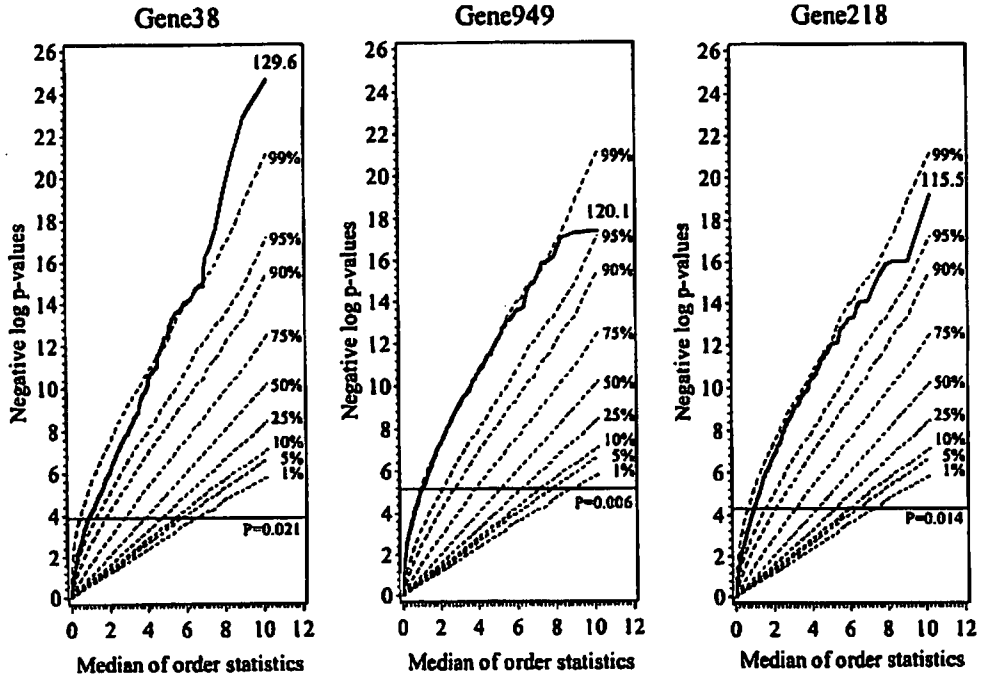

Fig. 3. The Q-Q plots of p-values with null reference for the top 3 genes according to the AUC. The horizontal line indicates the p-value in the original data. Dotted lines indicate the null references as shown in Figure 2(a).

the probability of obtaining the accumulation of low p-values is approximately less than 5% under the null hypothesis and there are some highly extreme low p-values. Although AUC of central Q-Q plot is smaller than AUC of left Q-Q plot, central Q-Q plot would be regarded as

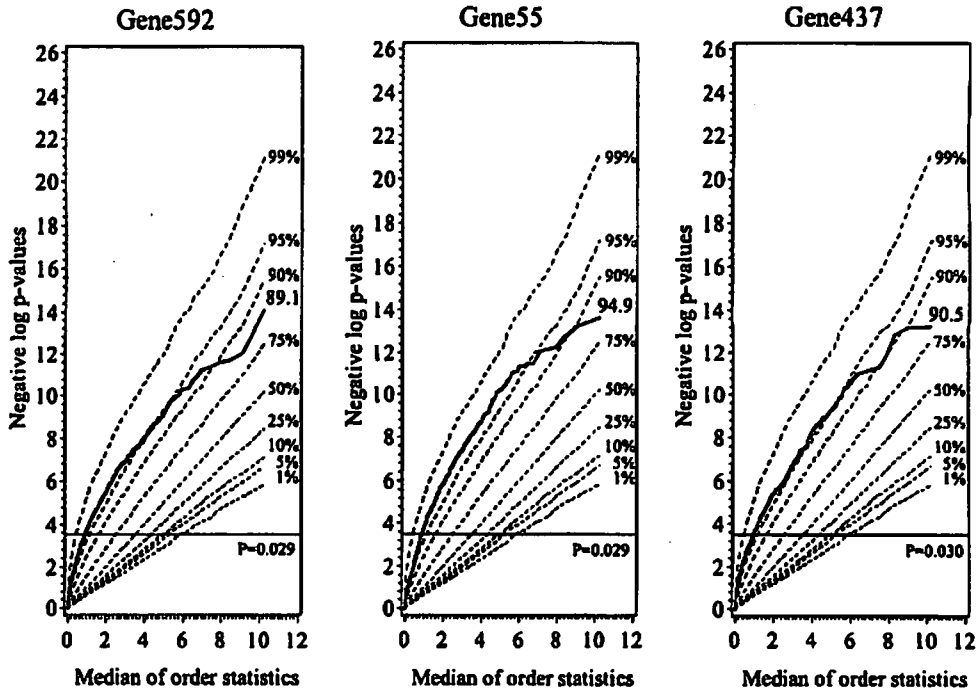

Fig. 4. The Q-Q plots of p-values with null reference of the genes whose p-value is around 3% in the original data. The horizontal line indicates the p-value in the original data. Dotted lines indicate the null references as shown in Figure 2(a).

more stable and significant than left Q-Q plot because central Q-Q plot almost follows the 99th percentile of null reference in the plot area where median of order statistic is less than 6. In fact, there are 98.6% of points for the Q-Q plot in this area. Thus we should pay attention to the shape of the Q-Q plot in this area.

Figure 4 shows the Q-Q plots of p-values with null reference of the genes whose p-value is around 3% in the original data. The curve of Gene55 almost lies between 99th percentile and 95th percentile of the null reference. On the other hand, the Q-Q plots for Gene592 and Gene437 follow the 95th percentile of the null reference. This imply that the Q-Q plot for Gene55 is more stable and significant than the other two Q-Q plots. Thus we could select Gene55 for the candidate gene.

4. Discussion

The Q-Q plot approach has an advantage for visualizing the whole distribution of the bootstrap p-values. As shown in the Figure 4, although the genes have a similar p-value in the ordinary analysis, their Q-Q plots make a difference. In addition, according to our approach, since there are the reference lines, we can evaluate the possibility that we obtain such a plot under the null hypothesis. If Q-Q plot of the gene observed lied above the 95th percentile of null reference, the probability of obtaining the bootstrap p-values was considered less than 5% under the null hypothesis.