PERGAMON

European Journal of Cancer 38 (2002) 2194–2199

European Journal of Cancer

www.ejconline.com

Assessment of the technology for functional imaging in cancer

G.R. Laking[a,*], P.M. Price[a], M.J. Sculpher[b]

[a]CRC PET Oncology Group, MRC Cyclotron Building, Hammersmith Hospital, Du Cane Rd, London W12 0NN, UK
[b]Centre for Health Economics, University of York, York, UK

Received 2 January 2002; accepted 10 January 2002

Abstract

Functional imaging can address hitherto irresolvable questions about cancer biology in both research and practice. In the clinic, by combining features of systemic and local disease markers into reference standards for the diagnosis of new disease entities functional imaging has the potential to literally redefine illness. Clinical assessments can be conducted at two levels: establishment of new reference standards, and evaluation of successor technologies that will substitute for a reference standard in practice. A union of functional imaging with anatomical criteria of disease has shown great promise in the management of numerous cancers. More work is required to use functional imaging to develop 'functional' approaches to diagnosis and therapy. Many methodologies exist for the acquisition of primary data on imaging technology efficacy. A form of economic cost-effectiveness modelling called iterative decision analysis can be used to set research and service priorities. Cancer clinicians need to take an increased role in functional imaging research, as they have primary expertise in the development and use of treatments modifying cell and tissue function.

Keywords: Neoplasms; Medical oncology; Diagnosis; Tomography, emission computed; Magnetic resonance spectroscopy; Radionuclide imaging; Technology assessment, biomedical; Cost effectiveness analysis; Decision support techniques; Monte Carlo simulation

1. Introduction

Functional imaging can address previously irresolvable questions about cancer biology in both research and practice. Its substantial expense mandates assessment of cost-effectiveness and cost-benefit. Such assessments can be made in several settings. The centres providing the images need evidence of sustained volumes of consumption in order to justify their capital outlay and ongoing costs [1]. The scientists and clinicians using the images must demonstrate added value on their own criteria. Functional imaging in basic research has enjoyed a supply of novel hypotheses which cannot be investigated by other means. For example, there is currently no other technology that can report directly on real-time drug metabolism *in vivo* within tissue volumes drawn from a cross-sectional image [2]. The ultimate value for research will depend on translation of these hypotheses into treatments of demonstrably increased efficacy and efficiency. The research applications of the technologies positron emission tomography (PET), single photon emission computed tomography (SPECT), functional magnetic resonance imaging (fMRI), and dynamic X-ray computed tomography (dCT) continue to evolve. For example, although PET ^{15}O-water scanning is presently a reference standard for quantitation of tumour perfusion, dCT may soon overtake this [3]. A fuller discussion of research developments is beyond the scope of this article. In the clinic, functional imaging has the potential to literally redefine illness, by combining features of systemic and local disease markers into new reference standards for the diagnosis of new disease entities. Assessments of diagnostic technology can be conducted at two levels: establishment of reference standards, and evaluation of successor technologies that will replace a reference standard in practice. The most rapidly advancing modality of clinical functional imaging has been PET scanning with ^{18}F-fluorodeoxyglucose (FDG). The combination of FDG-PET with X-ray computed tomography (CT) has been evaluated mainly in relation to pre-existing histopathological

* Corresponding author. Tel.: +44-208-383-3767; fax: +44-208-383-2029.
E-mail address: george.laking@csc.mrc.ac.uk (G.R. Laking).

0959-8049/02/$ - see front matter
PII: S0959-8049(02)00385-4

reference standards as a successor technology to conventional diagnostic methods. This wedding of functional images to pre-existing anatomical diagnostic criteria has shown great promise in the management of numerous cancers. To actually establish new reference standards using functional criteria will require some changes in the approach to technology assessment. Cancer clinicians need to take an increased role in functional imaging research, as they have primary expertise in the development and use of treatments modifying cell and tissue function.

2. Functional imaging and functional disease entities

The distinction between functional and anatomical criteria of disease is arbitrary, inasmuch as anatomical structure equates to enduring physiological function, and function to transient structure. Imaging can be called anatomical to the extent that it reports on macroscopic pathology, guides decisions based on disease stage, has tissue biopsy as its reference standard, and informs surgical decisions (Fig. 1). Imaging-directed surgical decisions can be related to predictable clinical outcomes based on the prognostic value of pathological stage and the image's power to predict biopsy results. This formula simplifies the execution and interpretation of large studies of anatomical diagnostic accuracy. Viewed in this way, when 'functional' technologies such as fluorine labelled deoxyglucose (FDG-PET) are used to stage cancer, they operate as anatomical imaging modalities. A contrasting example is when the anatomical technology of X-ray reveals abnormal tissue calcification attributable to the functional disease entity, parathyroid hormone-related peptide (PTH-RP) hypersecretion. Imaging can be called functional to the extent that it reports on pathophysiology, guides decisions based on disease grade, is not predicted by structural findings, and informs oncological decisions. Insofar as a test purports to identify new disease entities, the validity of these must be assessed as part of an entire package of medical management. Confirmation of a new diagnostic and therapeutic approach establishes the test as a reference standard in its own right. This opens the door to comparison between the reference standard and successor technologies, insofar as the basic concept of the disease entity does not change.

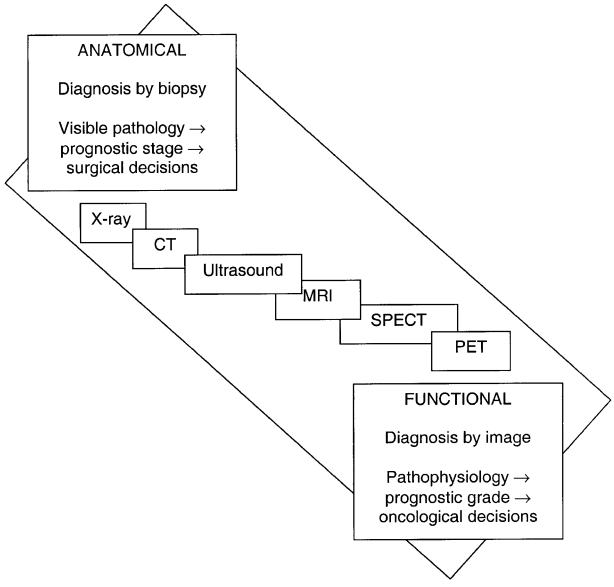

Fig. 1. A spectrum of anatomical and functional imaging modalities. CT, computed tomography; MRI, magnetic resonance imaging; SPECT, single photon emission computed tomography; PET, positron emission tomography.

3. The current status of assessment of functional imaging technology

3.1. Randomised controlled trials and head-to-head assessments

The evidence-based movement has identified the randomised controlled trial (RCT) as a key building block of proof of efficacy of medical interventions, although there has been debate about the proper extent of its use [4,5]. A key reason for undertaking RCTs of therapeutic interventions is to minimise bias in the assignation of patients to experimental or control groups. This reflects the fact that individual patients cannot independently undergo multiple treatments. By extension from their role in evaluation of treatments, RCTs are used for evaluation of reference standards defining new diagnostic entities. The role of RCTs in assessment of non-invasive diagnostic 'successor' technologies has been less clear, especially where the physiological impact may be so low as to enable 'head-to-head' comparisons in a cohort of patients who undergo multiple tests. Head-to-head data on diagnostic accuracy can be converted into predictions about patient outcomes by use of decision-modelling techniques. The use of head-to-head assessments has facilitated the expansion of FDG-PET into clinical practice, given the ubiquity of anatomical diagnostic reference standards. The large body of head-to-head studies has been extensively reviewed elsewhere [6,7].

3.2. Health technology assessment agencies and approaches

Many agencies have an interest in the evaluation of health technology within Europe. These include purchasing and funding authorities, industry sales divisions, medical publishing houses, and clinical expert associations. Organisations conducting evaluations at

national and international levels are listed in Table 1. Their output, of decisive influence in the use of diagnostic resources, is often not referenced in Medline. To this end, an electronic database for health technology assessment is maintained by the University of York [8]. The process of technology assessment typically starts with classification of primary data on efficacy. A useful approach was published by Fryback and Thornbury in the early 1990s [9]. Their 'hierarchical model' split the concept of diagnostic efficacy into a 'six tiered conceptual continuum', reproduced in Table 2. Anyone interested in justifiable allocation of health care resources needs assessments made at Level 6, 'societal efficacy'. The shortage of primary data at any level other than diagnostic efficacy is reflected in a recent call for research to demonstrate that ''using PET as a diagnostic technique will alter patient management'' [10]. Even this advance on basic diagnostic accuracy would still be two steps short of the goal of demonstrable cost-benefit (Table 2). The restriction of focus to diagnostic accuracy reflects functional imaging technology's use as a marker for histopathology. Because the reference standards in anatomical diagnosis are so secure, there is little need to go beyond the question 'how accurately does this test predict histopathology?' Competing tests can be regarded as interchangeable diagnostic modules of greater or lesser predictive power. Decision models make it easy to infer the impact on patient outcomes when one technology is substituted for another.

3.3. The need for accurate data in decision models

Despite its relative convenience and accuracy, the system for evaluating technology by use of head-to-head studies and decision models is not completely foolproof. A prime example is assessment of the utility of FDG-PET in staging non-small cell lung cancer (NSCLC) [11,12]. Decision models predicted substantial gains in cost effectiveness, but also that an extremely large and implicitly cost-ineffective randomised controlled trial (RCT) would be required to detect an improvement in survival. This influenced a school of thought that (RCTs) are not appropriate for imaging technology evaluation, especially in settings where more efficient diagnosis leads mainly to reduced costs as opposed to increased survival [13]. The argument appeared to be rebutted with publication of a RCT by Boyer and colleagues looking at the therapeutic efficacy of FDG-PET

Table 1
Organisations participating in health technology evaluation

International, EU and US federal organisations		
	International Network of Agencies for Health Technology Assessment (INAHTA)	http://www.inahta.org/
	International Society of Technology Assessment in Health Care (ISTAHC)	http://www.istahc.org
	WHO Technology Assessment and Quality Assurance	http://www.who.int/pht/technology_assessment/index.html
	The Cochrane Collaboration	http://www.cochrane.org
	The European Agency for the Evaluation of Medicinal Products (EMEA)	http://www.eudra.org/emea.html
	The EORTC Functional Imaging Group	http://www.eortc.be
	US Agency for Healthcare Research and Quality (AHRQ)	http://www.ahrq.gov
Organisations within Europe		
Norway	Norwegian Centre for Health Technology Assessment (SMM)	http://www.oslo.sintef.no/smm/
Finland	Finnish Office for Health Care Technology Assessment (FinOHTA)	http://www.stakes.fi/finohta/
Sweden	Swedish Council on Technology Assessment in Health Care (SBU)	http://www.sbu.se/
Denmark	Danish Institute for Health Technology Assessment (DIHTA)	http://www.dsi.dk/
Germany	German Scientific Working Group of Technology Assessment in Health Care	http://www.epi.mh-hannover.de/hta/index.htm
Netherlands	TNO'S HTA GROUP—The Netherlands Organization for Applied Scientific Research	http://www.health.tno.nl/en/about_tno/organisation/divisions/publichealth/health_technology_assessment.html
Great Britain	National Coordinating Centre for Health Technology Assessment	http://www.hta.nhsweb.nhs.uk/
France	L'Agence Nationale d'Accréditation et d'Evaluation en Santé (ANAES)	http://www.anaes.fr
Austria	ITA- Institute for Technology Assessment of the Austrian Academy of Sciences	http://www.oeaw.ac.at/~ita/welcome.htm
Switzerland	Swiss Science Council/Technology Assessment (SWISS/TA)	http://www.ta-swiss.ch/
Spain	Agencia de Evaluación de Tecnologías Sanitarias (AETS)	http://www.isciii.es/aets/

EU, European Union; US, United States; EORTC, European Organisation for Research and Treatment of Cancer.

in staging NSCLC [14]. All 179 patients had their mediastinum surgically staged, allowing confident estimation of their life expectancy [15]. Decision models had predicted that use of FDG-PET would give a 10% reduction in thoracotomy amongst patients with clinical Stage I or II NSCLC, leading to overall cost savings and some improvement in quality of life. The authors in a preliminary analysis found no alteration in thoracotomy rate or management of patients, and therefore no predicted cost savings. The chance of a type II (false-negative) statistical error in this study was calculated to have been 20%. Presuming that the findings are real, the most likely explanation is that the surgeons were 'biased' to operate at a lower threshold than had been foreseen. The majority of instances of management contrary to predictions were N2 stage disease, for which the published 2-year survival postresection is 40%, as opposed to 29% with radiotherapy [16]. Many surgeons and their patients faced with these odds would prefer to operate. The Boyer study demonstrates the utility of RCTs to screen for unsuspected sources of bias. The expedient of taking a surrogate endpoint enabled this small RCT, which showed up an erroneous modelling assumption about surgical practice. Even so, it could be argued that the need for a trial would have been less had surgeons been polled more closely on their management of N2 disease. In the end, decision models cannot be any better than the data that go into them. The strongest reason for retaining RCTs as a source of primary data on imaging technology is where the trial evaluates an entire management package associated with the genesis of new diagnostic entities.

3.4. Summary of current assessments

The realisation that FDG-PET does not have 100% specificity in relation to histopathology has reinforced its use as a surrogate marker for pre-existing 'anatomical' disease entities. Insofar as PET's function has been to detect unsuspected extensive disease, its clinical

Table 2
A hierarchical model of efficacy: typical measures of analyses[a]

Level 1:	Technical efficacy
	Resolution of line pairs
	Modulation transfer function charge
	Grey-scale range
	Amount of mottle
	Sharpness
Level 2:	Diagnostic accuracy efficacy
	Yield of abnormal diagnoses in a case series
	Diagnostic accuracy (percentage correct diagnoses in case series)
	Predictive value of positive or negative examination (in a case series)
	Sensitivity and specificity in a defined clinical problem setting
	Measures of area under the ROC curve
Level 3:	Diagnostic thinking efficacy
	Number (percentage) of cases in which image judged 'helpful' to making the diagnosis
	Entropy change in differential diagnosis probability estimation
	Difference in clinicians' subjectively estimated diagnosis probabilities pre- to post test information
	Empirical subjective log-likelihood ratio for test positive and negative in a case series
Level 4:	Therapeutic efficacy
	Number (percentage) of times image judged helpful in planning management of the patient in a case series
	Percentage of times medical procedure avoided due to image information
	Number or percentage of times therapy planned pre-test changed after the image information was obtained (retrospectively inferred from clinical records)
	Number or percentage of times clinicians' prospectively stated therapeutic choices changed after test information
Level 5:	Patient outcome efficacy
	Percentage of patients improved with test compared without test
	Morbidity (or procedures) avoided after having image information
	Change in quality-adjusted life expectancy
	Expected value of test information in quality-adjusted life years (QALYs)
	Cost per QALY saved with image information
Level 6:	Societal efficacy
	Benefit-cost analysis from societal viewpoint
	Cost-effectiveness analysis from societal viewpoint

ROC, receiver operating curve.
[a] Reproduced with permission from Ref. [10]

impact has been less on survival than on quality of life and costs. The data on accuracy in relation to predefined anatomical criteria have convinced the United States Medicare and the German national insurance scheme in the late 1990s to reimburse FDG-PET for detection and staging of a number of different cancers [17]. There can be problems with the generalisability of some of the decision models, in their reflection of primary data or patterns of practice [18]. Advances have been slower in the development of truly 'functional' imaging applications, and these are likely to require RCTs as a primary source of data.

4. Second order Monte Carlo simulations

A major problem with attempts to forge a systematic approach to imaging technology assessment has been the absence of an analytic framework combining research with practice. Although the idea of 'functional disease entities' may sound interesting, there has been no estimate of the benefit of diagnosing and treating these hypotheses. Additionally, there may already be enough evidence to justify the creation of new functional reference standards based on prognostic models drawing on pre-existing histological, systemic and imaging data, without recourse to new primary data from research. What has been needed is a method of quantitatively re-expressing diagnostic and therapeutic uncertainties in terms of the need to acquire further primary data. One such approach is the use of iterative decision analysis with Monte Carlo simulation [19]. This is a process that starts with the identification of a clinical problem and all feasible alternative management strategies (see Fig. 1). These may involve the use of new diagnostic equipment, identification of new disease entities, or new types of treatment. These strategies are set out in a decision model combining pooled data on the cost and efficacy of interventions with epidemiological features of the populations involved. A cost-effectiveness analysis then ranks each strategy across a range of 'λ', society's willingness-to-pay for health outcomes (usually given as € 50 000 per quality adjusted life year [12]). The '*a priori* strategy' is that which is most cost-effective for a given λ. This is what would be implemented in the absence of further information. The next step is the use of probabilistic methods to evaluate the model's uncertainty. Each model parameter is redefined as its statistical distribution from the literature. The model is then successively re-evaluated in a 'second-order Monte Carlo simulation'. This takes random samples from each distribution, creating a simulated cohort that approaches all possible combinations of patients, interventions and efficacies. Some of these combinations will have differing optimal strategies from the *a priori*. The next step is to compare the overall advantage, if the *optimal* strategy were to be used in every instance, as opposed to the *a priori*. This is called the 'Expected Value of Perfect Information' (EVPI), since if we had 'perfect information' we would always know the optimal strategy *a priori*. The EVPI is typically expressed in units of currency per patient. This, when multiplied by the predicted number of patients over a relevant timeframe, approaches the maximum value of any further data acquisition. The EVPI can be compared with the cost of a given programme of research: research costing more than the EVPI is not worth doing. Specific research questions can be examined. For example, in the case of staging NSCLC, it would be possible to value research into improved therapies for N2 disease, in comparison with improved detection of N2 disease. In its final step, the process can evaluate the worth of a clinical trial by calculating the 'expected value of sample information' (EVSI). EVSI approaches EVPI as the sample size increases. In summary, decision analysis with second-order Monte Carlo simulation permits quantitative statements on the value of research such as head-to-head studies and RCTs, with respect to the value of improved health outcomes. It therefore provides a quantitative framework for technology analysis that seamlessly combines research and practice.

5. Future directions for research

Functional imaging offers an approach to clinical research at the opposite end of a spectrum from the

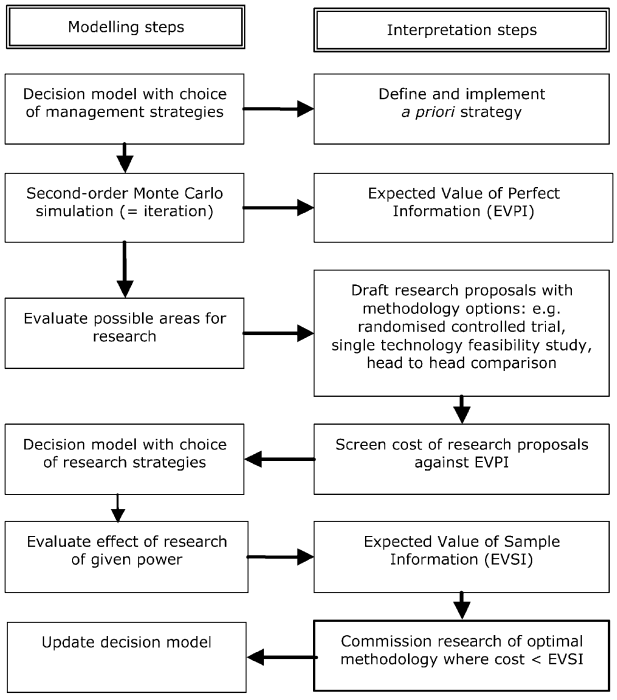

Fig. 2. Iterative decision analysis.

so-called 'megatrials'. Megatrials were conceived as a means of detecting small clinical effects from treatments applied to large 'lumped' sample populations [20]. The diagnostic 'splitting' enabled by the introduction of functional criteria promises substantial therapeutic advantages. An example of the benefits of diagnostic splitting is given by studies on acute myeloid leukaemia (AML), in which very large studies were able to confirm small incremental survival benefits with successive refinements of cytotoxic chemotherapy. Recognition of the acute promyelocytic diagnostic subtype, and its sensitivity to all-*trans* retinoic acid (ATRA), led to substantial prognostic improvement for this group of patients [21]. Similarly, selection for most consolidative and adjuvant therapies in oncology could be improved by recognition of 'at-risk' patients using functional imaging. For example, the use of neoadjuvant chemotherapy in NSCLC is currently being investigated in large RCTs comparing its use in all patients of a given stage group to no patients in that group. An approach using functional imaging would be to offer neoadjuvant chemotherapy to 'high-risk' patients defined on a prognostic index based on the metabolic rate of glucose calculated from a FDG-PET scan. Insofar as FDG-PET identified a subset of patients not predicted by conventional methods, it would be recognising a new disease entity. The functional approach could be validated in a RCT setting in comparison to the superior arm from conventional NSCLC studies.

6. Conclusions

Functional imaging has too much promise in cancer medicine to be left as purely a surrogate marker for histopathology (Fig. 2). Clinical research and practice with functional imaging is most likely to prosper insofar as it relates to functional models of disease. New functional diagnoses will first need to be assessed as part of an entire package of diagnosis and treatment. The key tool for acquisition of primary data on efficacy at this stage will be the RCT. Successive generations of technology can then be compared with baseline reference standards in head-to-head studies. Deployment of RCTs as an evaluative tool will primarily depend on their cost-benefits, which can be predicted using decision-analytical models with Monte Carlo simulation. The touchstone for all assessments of functional imaging in Oncology must be evidence of societal efficacy. The ultimate extent of service provision will be largely determined by economic assessments, which will inevitably reflect the available primary research. Cancer clinicians need to take a leading role in functional imaging research, as they have primary expertise in the development and use of treatments modifying cell and tissue function.

Acknowledgements

This work was funded by a Cancer Research Campaign Grant to P.M.P.

References

1. Keppler JS, Conti PS. A cost analysis of positron emission tomography. *Am J Roentgenol* 2001, **177**, 31–40.
2. Aboagye EO, Price PM, Jones T. In vivo pharmacokinetics and pharmacodynamics in drug development using positron-emission tomography. *Drug Discov Today* 2001, **6**, 293–302.
3. Miles KA, Kelley BB. CT measurements of capillary permeability within nodal masses: a potential technique for assessing the activity of lymphoma. *Br J Radiol* 1997, **70**, 74–79.
4. http://cebm.jr2.ox.ac.uk/docs/levels.html [accessed 21 September 2001].
5. Knapp MS, Gordon S, Lambert M, *et al*. Using clinical evidence. *Br Med J* 2001, **323**, 165.
6. Bar-Shalom R, Valdivia AY, Blaufox MD. PET imaging in oncology. *Sem Nucl Med* 2000, **30**, 150–185.
7. Laking G, Price P. 18-Fluorodeoxyglucose positron emission tomography (FDG-PET) and the staging of early lung cancer. *Thorax* 2001, **56**(Suppl. II), ii38–ii44.
8. http://agatha.york.ac.uk/welcome.htm [accessed 21 September 2001].
9. Thornbury JR, Fryback DG. Technology assessment—an American view. *Eur J Radiol* 1992, **14**, 147–156.
10. Robert G, Milne R. Positron emission tomography: establishing priorities for health technology assessment. *Health Technol Assess* 1999, **3**, 1–54.
11. Gambhir SS, Hoh CK, Phelps ME, *et al*. Decision tree sensitivity analysis for cost-effectiveness of FDG-PET in the staging and management of non-small-cell lung carcinoma. *J Nucl Med* 1996, **37**, 1428–1436.
12. Dietlein M, Weber K, Gandjour A, *et al*. Cost-effectiveness of FDG-PET for the management of potentially operable non-small cell lung cancer: priority for a PET-based strategy after nodal-negative CT results. *Eur J Nucl Med* 2000, **27**, 1598–1609.
13. Valk PE. Randomized trails are not appropriate for imaging technology evaluation. *J Nucl Med* 2000, **41**, 1125–1126.
14. Boyer MJ, Viney R, Fulham M, *et al*. A randomised trial of conventional staging (CS) with or without positron emission tomography (PET) in patients (Pts) with stage 1 or 2 non-small cell lung cancer. *Proc Am Soc Clin Oncol* 2001, **20** (abstr 1233).
15. Mountain CF. Revisions in the international system for staging lung cancer. *Chest* 1997, **111**, 1710–1717.
16. Saunders M, Dische S, Barrett A, Harvey A, Griffiths G, Palmar M. Continuous, hyperfractionated, accelerated radiotherapy (CHART) versus conventional radiotherapy in non-small cell lung cancer: mature data from the randomised multicentre trial. CHART Steering committee. *Radiother Oncol* 1999, **52**, 137–148.
17. http://149.142.143.206/newsletter/May_2001.pdf [accessed 21 September 2001].
18. Buxton MJ, Drummond MF, Van Hout BA, *et al*. Modelling in economic evaluation: an unavoidable fact of life. *Health Econ* 1997, **6**, 217–227.
19. Fenwick E, Claxton K, Sculpher M, Briggs A. *Improving the Efficiency and Relevance of Health Technology Assessment: The role of Iterative Decision Modelling. Discussion Paper 179. Centre for Health Economics Discussion Paper Series*. York, University of York, 2000.
20. Barer D. Clinical trials [Letter]. *Br Med J* 1999, **318**, 1138.
21. Clarkson B. Retinoic acid in acute promyelocytic leukemia: the promise and the paradox. *Cancer Cells* 1991, **3**, 211–220.