

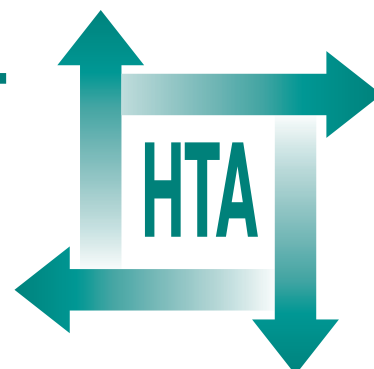
Cost-effectiveness of using prognostic information to select women with breast cancer for adjuvant systemic therapy

C Williams, S Brunskill, D Altman, A Briggs, H Campbell, M Clarke, J Glanville, A Gray, A Harris, K Johnston and M Lodge



September 2006

**Health Technology Assessment
NHS R&D HTA Programme**





INAHTA

How to obtain copies of this and other HTA Programme reports.

An electronic version of this publication, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website (<http://www.hta.ac.uk>). A fully searchable CD-ROM is also available (see below).

Printed copies of HTA monographs cost £20 each (post and packing free in the UK) to both public **and** private sector purchasers from our Despatch Agents.

Non-UK purchasers will have to pay a small fee for post and packing. For European countries the cost is £2 per monograph and for the rest of the world £3 per monograph.

You can order HTA monographs from our Despatch Agents:

- fax (with **credit card** or **official purchase order**)
- post (with **credit card** or **official purchase order** or **cheque**)
- phone during office hours (**credit card** only).

Additionally the HTA website allows you **either** to pay securely by credit card **or** to print out your order and then post or fax it.

Contact details are as follows:

HTA Despatch
c/o Direct Mail Works Ltd
4 Oakwood Business Centre
Downley, HAVANT PO9 2NP, UK

Email: orders@hta.ac.uk
Tel: 02392 492 000
Fax: 02392 478 555
Fax from outside the UK: +44 2392 478 555

NHS libraries can subscribe free of charge. Public libraries can subscribe at a very reduced cost of £100 for each volume (normally comprising 30–40 titles). The commercial subscription rate is £300 per volume. Please see our website for details. Subscriptions can only be purchased for the current or forthcoming volume.

Payment methods

Paying by cheque

If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *Direct Mail Works Ltd* and drawn on a bank with a UK address.

Paying by credit card

The following cards are accepted by phone, fax, post or via the website ordering pages: Delta, Eurocard, Mastercard, Solo, Switch and Visa. We advise against sending credit card details in a plain email.

Paying by official purchase order

You can post or fax these, but they must be from public bodies (i.e. NHS or universities) within the UK. We cannot at present accept purchase orders from commercial companies or from outside the UK.

How do I get a copy of HTA on CD?

Please use the form on the HTA website (www.hta.ac.uk/htacd.htm). Or contact Direct Mail Works (see contact details above) by email, post, fax or phone. *HTA on CD* is currently free of charge worldwide.

The website also provides information about the HTA Programme and lists the membership of the various committees.

Cost-effectiveness of using prognostic information to select women with breast cancer for adjuvant systemic therapy

C Williams,^{1*} S Brunskill,² D Altman,³ A Briggs,⁴
H Campbell,⁵ M Clarke,⁶ J Glanville,⁷ A Gray,⁵
A Harris,⁸ K Johnston⁹ and M Lodge¹⁰

¹ Bristol Haematology and Oncology Centre, UK

² National Blood Service, John Radcliffe Hospital, Oxford, UK

³ Centre for Statistics in Medicine, Wolfson College, Oxford, UK

⁴ Section of Public Health and Health Policy, University of Glasgow, UK

⁵ Department of Public Health, University of Oxford, UK

⁶ Clinical Trial Service Unit and Epidemiological Studies Unit, Richard Doll Building, Oxford, UK

⁷ Centre for Reviews and Dissemination, University of York, UK

⁸ Cancer Research UK, Medical Oncology Unit, Churchill Hospital, Oxford, UK

⁹ Economics and Statistics Division, Scottish Executive Environment and Rural Affairs Department, Edinburgh, UK

¹⁰ Cochrane Cancer Network, Wolfson College, Oxford, UK

* Corresponding author

Declared competing interests of authors: none

Published September 2006

This report should be referenced as follows:

Williams C, Brunskill S, Altman D, Briggs A, Campbell H, Clarke M, *et al.*
Cost-effectiveness of using prognostic information to select women with breast
cancer for adjuvant systemic therapy. *Health Technol Assess* 2006;**10**(34).

Health Technology Assessment is indexed and abstracted in *Index Medicus/MEDLINE*,
Excerpta Medica/EMBASE and *Science Citation Index Expanded (SciSearch®)* and
Current Contents®/Clinical Medicine.

NHS R&D HTA Programme

The research findings from the NHS R&D Health Technology Assessment (HTA) Programme directly influence key decision-making bodies such as the National Institute for Health and Clinical Excellence (NICE) and the National Screening Committee (NSC) who rely on HTA outputs to help raise standards of care. HTA findings also help to improve the quality of the service in the NHS indirectly in that they form a key component of the 'National Knowledge Service' that is being developed to improve the evidence of clinical practice throughout the NHS.

The HTA Programme was set up in 1993. Its role is to ensure that high-quality research information on the costs, effectiveness and broader impact of health technologies is produced in the most efficient way for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined to include all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care, rather than settings of care.

The HTA Programme commissions research only on topics where it has identified key gaps in the evidence needed by the NHS. Suggestions for topics are actively sought from people working in the NHS, the public, service-users groups and professional bodies such as Royal Colleges and NHS Trusts. Research suggestions are carefully considered by panels of independent experts (including service users) whose advice results in a ranked list of recommended research priorities. The HTA Programme then commissions the research team best suited to undertake the work, in the manner most appropriate to find the relevant answers. Some projects may take only months, others need several years to answer the research questions adequately. They may involve synthesising existing evidence or conducting a trial to produce new evidence where none currently exists.

Additionally, through its Technology Assessment Report (TAR) call-off contract, the HTA Programme is able to commission bespoke reports, principally for NICE, but also for other policy customers, such as a National Clinical Director. TARs bring together evidence on key aspects of the use of specific technologies and usually have to be completed within a short time period.

Criteria for inclusion in the HTA monograph series

Reports are published in the HTA monograph series if (1) they have resulted from work commissioned for the HTA Programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

The research reported in this monograph was commissioned by the HTA Programme as project number 98/36/02. The contractual start date was in May 2001. The draft report began editorial review in November 2002 and was accepted for publication in November 2005. As the funder, by devising a commissioning brief, the HTA Programme specified the research question and study design. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the referees for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

The views expressed in this publication are those of the authors and not necessarily those of the HTA Programme or the Department of Health.

Editor-in-Chief: Professor Tom Walley
Series Editors: Dr Aileen Clarke, Dr Peter Davidson, Dr Chris Hyde,
Dr John Powell, Dr Rob Riemsma and Dr Ken Stein
Managing Editors: Sally Bailey and Sarah Llewellyn Lloyd

ISSN 1366-5278

© Queen's Printer and Controller of HMSO 2006

This monograph may be freely reproduced for the purposes of private research and study and may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising.

Applications for commercial reproduction should be addressed to NCCHTA, Mailpoint 728, Boldrewood, University of Southampton, Southampton, SO16 7PX, UK.

Published by Gray Publishing, Tunbridge Wells, Kent, on behalf of NCCHTA.
Printed on acid-free paper in the UK by St Edmundsbury Press Ltd, Bury St Edmunds, Suffolk.



Abstract

Cost-effectiveness of using prognostic information to select women with breast cancer for adjuvant systemic therapy

C Williams,^{1*} S Brunskill,² D Altman,³ A Briggs,⁴ H Campbell,⁵ M Clarke,⁶ J Glanville,⁷ A Gray,⁵ A Harris,⁸ K Johnston⁹ and M Lodge¹⁰

¹ Bristol Haematology and Oncology Centre, UK

² National Blood Service, John Radcliffe Hospital, Oxford, UK

³ Centre for Statistics in Medicine, Wolfson College, Oxford, UK

⁴ Section of Public Health and Health Policy, University of Glasgow, UK

⁵ Department of Public Health, University of Oxford, UK

⁶ Clinical Trial Service Unit and Epidemiological Studies Unit, Richard Doll Building, Oxford, UK

⁷ Centre for Reviews and Dissemination, University of York, UK

⁸ Cancer Research UK, Medical Oncology Unit, Churchill Hospital, Oxford, UK

⁹ Economics and Statistics Division, Scottish Executive Environment and Rural Affairs Department, Edinburgh, UK

¹⁰ Cochrane Cancer Network, Wolfson College, Oxford, UK

* Corresponding author

Objectives: To investigate the cost-effectiveness of using prognostic information to identify patients with breast cancer who should receive adjuvant therapy.

Data sources: Electronic databases from 1980 through to February 2002. A survey of clinical practice in UK cancer centres and units. Large retrospective dataset containing data on prognostic factors, treatments and outcomes for women with early breast cancer treated in Oxford.

Review methods: Between six and nine databases were searched by an information expert. Evidence-based methods were used to review and select those studies and the quality of each included paper was assessed using standard assessment tools reported in the literature or piloted and developed for this study. A survey of clinical practice in UK cancer centres and units was carried out to ensure that conclusions drawn from the report could be implemented. These data, along with the information gathered in the systematic reviews, informed the methodological approach adopted for the health economic modelling. An illustrative framework was developed for incorporating patient-level prediction within a health economic decision model. This framework was applied to a large retrospective dataset containing data on prognostic factors, treatments and outcomes for women with early breast cancer treated in Oxford. The data were used to estimate directly a parametric regression-based

risk equation, from which a prognostic index was developed, and prognosis-specific estimates of the baseline breast cancer hazard could be observed. Published estimates of treatment effects, health service treatment costs and utilities were used to construct a decision analytic framework around this risk equation, thus enabling simulation of the effectiveness and cost-effectiveness of adjuvant therapy for all possible combinations of prognostic factors included in the model.

Results: The lack of good-quality systematic reviews and well-conducted studies of prognostic factors in breast cancer is a striking finding. There are no registers of studies of prognostic factors or of reviews of prognostic studies. Many of the reviews used weak methods, primary studies are similar with poor methodology and reporting of results. In addition, there is much variation in patient populations, assay methods, analysis of results, definitions used and reporting of results. Most studies appear to be retrospective and some use inappropriate methods likely to inflate outcomes such as optimising cut points and failing to test the results in an independent population. Very few reviews used meta-analysis to conduct a pooled analysis and to provide an estimate of the average size of any association. Instead, most reviews relied on vote counting. Although many prognostic models for breast cancer have been published, remarkably few have been

re-examined by independent groups in independent settings. The few validation studies have been carried out on ill-defined samples, sometimes of smaller size and short follow-up, and sometimes using different patient outcomes when validating a model. The evidence from the validation studies shows support for the prognostic value of the Nottingham Prognostic Index (NPI). No new prognostic factors have been shown to add substantially to those identified in the 1980s. Improvement of this index depends on finding factors that are as important as, but independent of, lymph node, stage and pathological grade. The NPI remains a useful clinical tool, although additional factors may enhance its use. We accepted that hormone receptor status (ER) for hormonal therapy such as tamoxifen and prediction of response to trastuzumab by HER2 did not require systematic review, as the mechanism of action of these drugs requires intact receptors. There was no clear evidence that other factors were useful predictors of response and survival. The survey confirmed pathological nodal status, tumour grade, tumour size and ER status as the most clinically important factors for consideration when selecting women with early breast cancer for adjuvant systemic therapy in the UK. The protocols revealed that although UK cancer centres appear to be using the same prognostic and predictive factors when selecting women to receive adjuvant therapy, much variation in clinical practice exists. Some centres use protocols based upon the NPI whereas others do not use a single index score. Within NPI and non-NPI users, between-centre variability exists in guidelines for women for whom the benefits are uncertain. Consensus amongst units appears to be greatest when selecting women for adjuvant hormone therapy with the decision based primarily upon ER or progesterone receptor status rather than combinations of a number of factors. Guidelines as to who should receive adjuvant chemotherapy, however, were found to be much less

uniform. Searches of the literature revealed only five published papers that had previously examined the cost-effectiveness of using prognostic information for clinical decision-making. These studies were of varying quality and highlight the fact that economic evaluation in this area appears still to be in its infancy. By combining methodologies used in determining prognosis with those used in health economic evaluation, it was possible to illustrate an approach for simulating the effectiveness (survival and quality-adjusted survival) and the cost-effectiveness associated with the decision to treat individual women or groups of women with different prognostic characteristics. The model showed that effectiveness and cost-effectiveness of adjuvant systemic therapy have the potential to vary substantially depending upon prognosis. For some women therapy may prove very effective and cost-effective, whereas for others it may actually prove detrimental (i.e. the reductions in health-related quality of life outweigh any survival benefit).

Conclusions: Outputs from the framework constructed using the methods described here have the potential to be useful for clinicians, attempting to determine whether net benefits can be obtained from administering adjuvant therapy for any presenting woman; and also for policy makers, who must be able to determine the total costs and outcomes associated with different prognosis based treatment protocols as compared with more conventional treat all or treat none policies. A risk table format enabling clinicians to look up a patient's prognostic factors to determine the likely benefits (survival and quality-adjusted survival) from administering therapy may be helpful. For policy makers, it was demonstrated that the model's output could be used to evaluate the cost-effectiveness of different treatment protocols based upon prognostic information. The framework should also be valuable in evaluating the likely impact and cost-effectiveness of new potential prognostic factors and adjuvant therapies.



Contents

Glossary and list of abbreviations	vii	Class of prognostic factor: metastases	28
Executive summary	ix	Class of prognostic factor: standard pathology	32
I Background and objectives	1	Other prognostic factors	34
Background	1	Discussion	34
UK/USA, 1950–99: recent decrease in breast cancer mortality at ages 20–69 years	1	5 Prognostic models in breast cancer	37
Absolute risk reduction during the first 10 years, subdivided by tamoxifen duration and by nodal status [after exclusion of women with oestrogen receptor (ER)-poor disease]	2	Introduction	37
Objectives	3	Developing reliable prognostic models	37
Objective of the project	4	Quality assessment	42
2 Systematic review methods	5	Searches	43
Search strategy	5	Data extraction	43
Search questions A–E	5	Published prognostic models – review of methodology	44
Study question F	6	Published prognostic models – review of findings	50
Data extraction strategy	6	Validation of prognostic models	52
3 Assessing the methodological quality of prognostic studies	7	Discussion	58
Introduction	7	6 Predictive factors in breast cancer	63
Assessing methodological quality – study design	7	Quality and type of studies used to identify useful predictive factors	64
Assessing methodological quality – analysis	9	Oestrogen and progesterone receptors (ER and PR)	64
Assessing methodological quality – overall assessment	12	HER2 and trastuzumab (Herceptin)	64
Reporting prognostic studies	12	Cathepsin D	65
Predictive markers	13	HER2 (HER2/neu or c-erb-B2) as a predictive factor for adjuvant chemotherapy and hormone therapy	66
Discussion	13	p53 as a predictive factor	68
4 Systematic review of reviews of studies of prognostic factors	15	7 Use of prognostic and predictive factors to select women with early breast cancer for adjuvant systemic therapy – survey and review of treatment protocols from UK breast cancer centres and units (2001–2)	69
Introduction	15	Introduction	69
Searching and selection	15	The survey – questionnaires	69
Individual factors	16	The survey – study participants	70
Class of prognostic factor: oestrogen receptor pathways	16	The survey – questionnaire administration	71
Class of prognostic factor: growth factor pathways	18	The survey – results	72
Class of prognostic factor: cell proliferation markers	22	Clinicians' questionnaire – findings	72
Class of prognostic factor: genetic instability and checkpoints	24	Results by 'regional centre' and 'smaller units'	73
Class of prognostic factor: apoptosis	27	Histopathologists' questionnaire – findings	75
		The survey – review of unit protocols	76

Guidelines for adjuvant hormone therapy (45 protocols)	77	Systematic review of reviews of studies of prognostic factors	123
Guidelines for adjuvant chemotherapy (45 protocols)	79	Overall conclusions	129
Comparison of protocols using and not using the NPI to aid with the selection of patients for adjuvant chemotherapy	83	Prognostic models in breast cancer	130
Conclusions and discussion	88	Reviewing current models	131
8 Cost-effectiveness of prognostic models ...	91	Assessment of published prognostic models	131
Introduction	91	Predictive factors in breast cancer	133
Search results	92	Survey of UK cancer centres and units	133
Informing of model parameters (search 1)	92	Guidelines for prescribing adjuvant therapy	135
Cost-effectiveness of using prognostic information (search 2)	92	Health economic evaluation	138
Cost-effectiveness models	93	Acknowledgements	141
Conclusions	102	References	143
9 Modelling the cost-effectiveness of using prognostic information in the adjuvant treatment of breast cancer	103	Appendix 1 Search strategies	155
Introduction	103	Appendix 2 Forms used to assess the eligibility of studies identified using searches for Topics A–E	179
The patient data set	104	Appendix 3 Forms used to extract data from studies assessed as eligible for Topics A, B and C	183
Estimating a survival model: the OPI	108	Appendix 4 Questionnaires used to survey clinicians and histopathologists for Topic E	197
A model for the cost and (quality-adjusted) survival of breast cancer patients	112	Health Technology Assessment reports published to date	205
Illustrative predictions from the cost- effectiveness model	113	Health Technology Assessment Programme	219
Presentation of results – the risk table approach	115		
Further research	118		
Conclusions	119		
Conference presentations	119		
10 Summary, discussion and conclusions	121		
Approach to the problem	122		
Assessing the methodological quality of prognostic studies	122		



Glossary and list of abbreviations

Technical terms and abbreviations are used throughout this report. The meaning is usually clear from the context, but a glossary is provided for the non-specialist reader. In some cases, usage differs in the literature, but the term has a constant meaning throughout this review.

Glossary

Predictive factor or predictive marker

Any patient or tumour characteristic that is predictive of the patient's response to a specified treatment. Response is usually measured in terms of overall survival, disease-free survival and/or death.

Prognostic factor or prognostic marker

Any patient or tumour characteristic that is predictive of the patient's outcome. Outcome is usually measured in terms of overall survival, disease-free survival and/or death.

Predictive model A statistical combination of at least two predictive factors to predict response to a specified treatment.

Prognostic index Quantitative set of values based on results of a prognostic model.

Prognostic model A statistical combination of at least two separate prognostic variables to predict patient outcome.

List of abbreviations

ANN	artificial neural network	EBCTCG	Early Breast Cancer Trialists' Collaborative Group
ASCO	American Society of Clinical Oncology	ECCG	electrocardiogram
CART	classification and regression trees	EGFR	epidermal growth factor receptor
CCTR	Cochrane Controlled Trials Register	ELISA	enzyme-linked immunosorbent assay
CEA	carcinoembryonic antigen	EORTC–RBG	European Organisation for Research and Treatment of Cancer–Receptor and Biomarker Group
CI	confidence interval	EPV	event prognostic variable (value)
CT	controlled trial	ER	oestrogen receptor
DDFS	distant disease-free survival		
DRE	digital rectal examination		

continued

List of abbreviations continued

ERB-B2	see HER2	PR	progesterone receptor
HEED	Health Economics Evaluation Database	PSA	prostate-specific antigen
HER	human epidermal growth factor receptor	QALY	quality-adjusted life-year
HER2	human epidermal growth factor receptor family 2	RCT	randomised control trial
HR	hazard ratio	RFS	recurrence-free survival
ICER	incremental cost-effectiveness ratio	RR	relative risk
KPI	Kalmar Prognostic Index	SD	standard deviation
LYG	life-year gained	SE	standard error
NHS EED	NHS Economic Evaluation Database	SU	smaller unit
NPI	Nottingham Prognostic Index	uPA	urokinase-type plasminogen activator
OPI	Oxford Prognostic Index	uPAR	urokinase-type plasminogen activator receptor
PI	prognostic index	VEGF	vascular endothelial growth factor

All abbreviations that have been used in this report are listed here unless the abbreviation is well known (e.g. NHS), or it has been used only once, or it is a non-standard abbreviation used only in figures/tables/appendices in which case the abbreviation is defined in the figure legend or at the end of the table.



Executive summary

Background

During the second half of the twentieth century, researchers came to understand that breast cancer could spread to other parts of the body at an early stage in the development of the disease. This led to a large number of randomised trials testing the utility of adjuvant hormone and cytotoxic therapy. These trials have shown that adjuvant therapy reduces the risk of recurrence and death from breast cancer, such that combinations of modern hormonal and cytotoxic therapy might halve the risk of a woman dying of breast cancer in the first 10 years after diagnosis. However, these improvements in the outlook for women with breast cancer have – in the main – been achieved by research that has required the treatment of all patients, including those destined not to relapse and those who relapse despite adjuvant therapy. Because of this, researchers have sought to find prognostic and predictive factors which would allow patients most likely to benefit from adjuvant therapy to be identified. Using combinations of factors to develop prognostic models has further refined their use.

Objectives

The principal objective of this project was to investigate the cost-effectiveness of using prognostic information to identify patients with breast cancer who should receive adjuvant therapy.

Methods

This report systematically reviewed the literature on the prognostic and predictive factors in breast cancer. Health economic decision analytic modelling was then used to draw conclusions on the most effective and efficient use of these factors in selecting women with early breast cancer for adjuvant systemic therapy.

The size of the literature meant that it was not possible to review systematically all primary publications in the area. A series of systematic reviews and a survey were undertaken on the following topics:

- quality assessment of prognostic studies (not necessarily cancer)

- reviews of prognostic information in breast cancer
- prognostic models in breast cancer
- predictive factors in breast cancer
- the clinical use of prognostic information in breast cancer in the UK (survey)
- quality of life, cost and cost-effectiveness studies relevant to modelling.

Between six and nine databases were searched by an information expert. Evidence-based methods were used to review the abstracts, select those suitable for inclusion and extract the data using piloted data extraction forms for each of the systematic reviews. The quality of each included paper was assessed using standard assessment tools reported in the literature or piloted and developed for this study.

It was not possible to carry out a quantitative analysis of the data for any of the systematic reviews. Instead, narrative summaries of the evidence were prepared with commentaries on the strengths and weaknesses of the conclusions drawn.

A survey of clinical practice in UK cancer centres and units was carried out to ensure that conclusions drawn from the report could be implemented. These data, along with the information gathered in the systematic reviews, informed the methodological approach adopted for the health economic modelling. Estimation of a definitive model was not considered feasible based on the current published literature. Rather, given the obvious benefits to be gained by establishing prognosis and treatment effectiveness and cost-effectiveness for individual patients or groups of patients, a pragmatic decision was made to develop and report an illustrative framework for incorporating patient-level prediction within a health economic decision model. This framework was applied to a large retrospective dataset containing data on prognostic factors, treatments and outcomes for women with early breast cancer treated in Oxford. The data were used to estimate directly a parametric regression-based risk equation, from which a prognostic index was developed, and prognosis-specific estimates of the baseline breast cancer hazard could be observed. Published estimates of treatment effects, health

service treatment costs and utilities were used to construct a decision analytic framework around this risk equation, thus enabling simulation of the effectiveness and cost-effectiveness of adjuvant therapy for all possible combinations of prognostic factors included in the model. Various ways of using the outputs from this framework were explored.

Results

Methodological quality of prognostic studies

There was a lack of empirical evidence to support the importance of particular study features affecting the reliability of findings and the avoidance of bias. However, there is much evidence that prognostic research in cancer tends to be of poor quality, contributing to the fact that prognostic markers often remain under investigation for years without good evidence that they are useful. Multiple small, separate, uncoordinated and often unvalidated studies often delay the process of defining the role of particular prognostic markers. Cooperation between research groups could lead to clear results emerging more rapidly, especially if such efforts are put into prospective studies or retrospective studies based on individual data from carefully assembled databases and/or tissue banks.

Systematic review of studies of prognostic factors

There is a plethora of evidence relating to possible prognostic factors for breast cancer. It was only possible to review those reviews that appeared to use systematic methods. There is a lack of high-quality, well-reported evidence in areas where it is taken for granted that factors have prognostic value, such as node status and age, and we have not reviewed these, accepting the commonly assumed value of such factors. A small number of eligible reviews (from 1 to –6 per factor) were found for each of 18 different factors. The lack of good-quality systematic reviews and well-conducted studies of prognostic factors in breast cancer was striking. In only five instances was the evidence strong enough to conclude that there is clear evidence of a relationship between the factor and survival (tumour size, proliferation indices, p53, cathepsin D and urokinase and its receptors).

Prognostic models

Although many prognostic models for breast cancer have been published, remarkably few have

been re-examined by independent groups in independent settings. The few validation studies have been carried out on ill-defined samples, sometimes of smaller size and short follow-up, and sometimes using different patient outcomes when validating a model.

The evidence from the validation studies shows support for the prognostic value of the Nottingham Prognostic Index (NPI). No new prognostic factors have been shown to add substantially to those identified in the 1980s. Improvement of this index depends on finding factors that are as important as, but independent of, lymph node, stage and pathological grade. The NPI remains a useful clinical tool, although additional factors may enhance its use.

Predictive factors

We accepted that hormone receptor status (ER) for hormonal therapy such as tamoxifen and prediction of response to trastuzumab by HER2 did not require systematic review, as the mechanism of action of these drugs requires intact receptors. There was no clear evidence that other factors were useful predictors of response and survival.

Survey of UK practice when selecting women for adjuvant therapy

The survey confirmed pathological nodal status, tumour grade, tumour size and ER status as the most clinically important factors for consideration when selecting women with early breast cancer for adjuvant systemic therapy in the UK. The protocols revealed that although UK cancer centres appear to be using the same prognostic and predictive factors when selecting women to receive adjuvant therapy, much variation in clinical practice exists. Some centres use protocols based upon the NPI whereas others do not use a single index score. Within NPI and non-NPI users, between-centre variability exists in guidelines for women for whom the benefits are uncertain. Consensus amongst units appears to be greatest when selecting women for adjuvant hormone therapy with the decision based primarily upon ER or progesterone receptor (PR) status rather than combinations of a number of factors. Guidelines as to who should receive adjuvant chemotherapy, however, were found to be much less uniform.

Cost-effectiveness of prognostic models

Searches of the literature revealed only five published papers that had previously examined

the cost-effectiveness of using prognostic information for clinical decision-making. These studies were of varying quality and highlight the fact that economic evaluation in this area appears still to be in its infancy.

By combining methodologies used in determining prognosis with those used in health economic evaluation, it was possible to illustrate an approach for simulating the effectiveness (survival and quality-adjusted survival) and the cost-effectiveness associated with the decision to treat individual women or groups of women with different prognostic characteristics.

The model showed that effectiveness and cost-effectiveness of adjuvant systemic therapy have the potential to vary substantially depending upon prognosis. For some women therapy may prove very effective and cost-effective, whereas for others it may actually prove detrimental (i.e. the reductions in health-related quality of life outweigh any survival benefit).

Conclusions and further research

Outputs from the framework constructed using the methods described here have the potential to be useful for clinicians, attempting to determine whether net benefits can be obtained from administering adjuvant therapy for any presenting woman; and also for policy makers, who must be able to determine the total costs and outcomes associated with different prognosis-based treatment protocols as compared with more conventional treat all or treat none policies. A risk table format enabling clinicians to look up a patient's prognostic factors to determine the likely benefits (survival and quality-adjusted survival) from administering therapy may be helpful. For policy makers, it was demonstrated that the model's output could be used to evaluate the cost-effectiveness of different treatment protocols based upon prognostic information. The framework should also be valuable in evaluating the likely impact and cost-effectiveness of new potential prognostic factors and adjuvant therapies.

Chapter I

Background and objectives

Background

Although breast cancer incidence and mortality internationally have been stable or declining over the past decade for the first time since data were collected,¹ breast cancer remains the most common cancer occurring in women worldwide and it is the leading cancer-related cause of death for women in Europe.^{2,3} The lifetime risk of a woman developing breast cancer is one in eight in the USA and one in nine in England and Wales.^{2,4}

Although there is increasing evidence on some of the major risk factors for breast cancer (*Table 1*), few are amenable to easy change⁵ so that early diagnosis and therapy remain crucial to improving survival rates.

Despite this, England and Wales (along with the USA and some other developed countries) have seen a decline in mortality from breast cancer in the last decade⁶ (*Figure 1*).

UK/USA, 1950–99: recent decrease in breast cancer mortality at ages 20–69 years

Part of this improved survival has been due to increasing use of adjuvant systemic therapies.⁵ The key tools in developing such an approach have been the randomised controlled trial (RCT) and systematic review with meta-analysis. Controlled trials (CTs) of various forms of systemic adjuvant therapy started as early as the 1950s, but it was not until the advent of systematic reviews and meta-analysis that a true estimate of the effectiveness of these approaches was understood. Meta-analyses based on individual patient data have proved to be a powerful tool for understanding the effectiveness of therapy and of studying groups of patients who may respond differently to therapy and the best ways of delivering these treatments.

Every 5 years since 1984–5, the Early Breast Cancer Trialists' Collaborative Group (EBCTCG)

TABLE 1 Established and probable risk factors for breast cancer

Factor	Relative risk	High-risk group
Geographical location	5	Developed country
Age	> 10	Elderly
Age at menarche	3	Menarche before age 11 years
Age at menopause	2	Menopause after age 54 years
Age at first full pregnancy	3	First child in early 40s
Family history	≥ 2	Breast cancer in first-degree relative when young
Previous benign disease	4–5	Atypical hyperplasia
Cancer in other breast	> 4	
Socioeconomic group	2	Groups I and II
Diet	1.5	High intake of saturated fat
Body weight:		
Premenopausal	0.7	Body mass index > 35
Postmenopausal	2	Body mass index > 35
Alcohol consumption	1.3	Excessive intake
Exposure to ionising radiation	3	Abnormal exposure in young females after age 10
Taking exogenous hormones:		
Oral contraceptives	1.24	Current use
Hormone replacement therapy	1.35	Use for ≥ 10 years
Diethylstilbestrol	2	Use during pregnancy

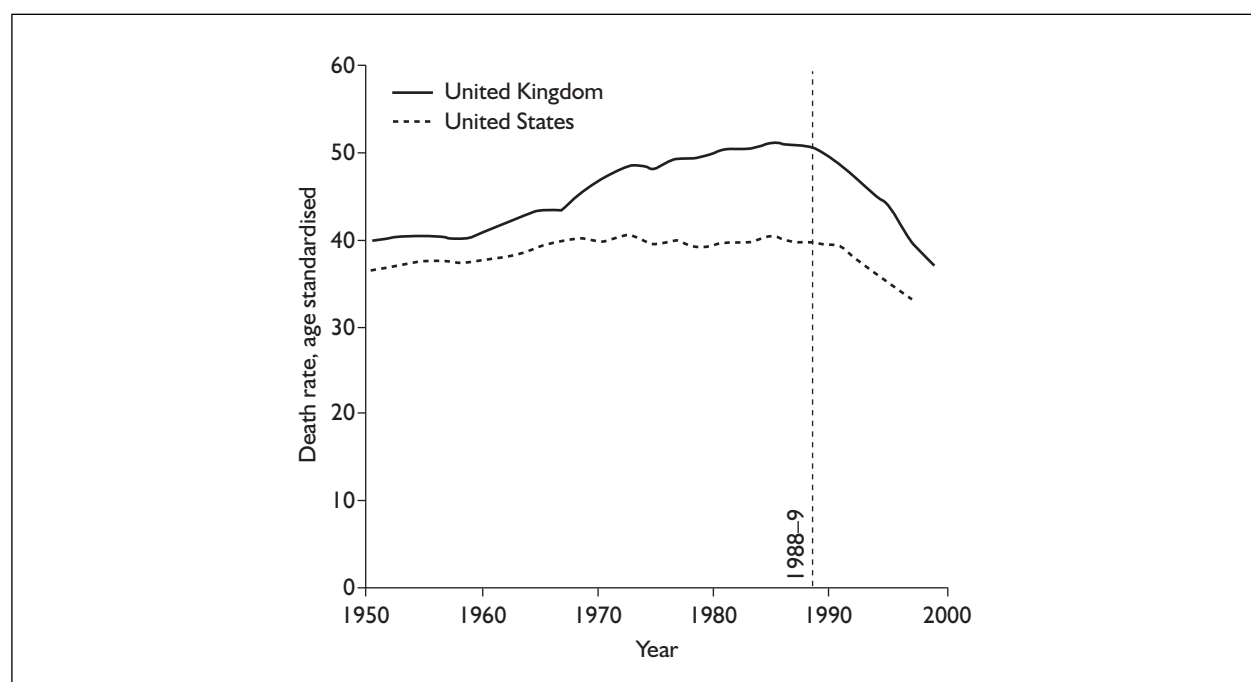


FIGURE 1 UK/USA 1950–99. Recent decrease in breast cancer mortality at ages 20–69 years: mean of annual rates per 100 000 women in component 5-year age groups.

has undertaken systematic overviews of all randomised trials of any aspect of the treatment of early breast cancer.^{7–14} These have given a major insight into how and when to use adjuvant therapy in women with early breast cancer.

In women with ‘early’ breast cancer, all detectable cancer is, by definition, restricted to the breast (and, in the case of node-positive patients, the local axillary lymph nodes) and can be removed surgically. However, undetected micro-metastatic deposits of the disease may remain and, perhaps after a delay of several years, develop into a clinically detectable recurrence that eventually causes death. The aim of adjuvant systemic therapy is to eradicate these micro-metastases. It has been shown by the EBCTCG that

- The use of adjuvant tamoxifen, usually for 5 years, significantly improves the 15-year survival for such women.^{7–9,14}
- Ovarian ablation significantly improved 15-year survival.¹¹
- Some months of adjuvant chemotherapy with two or more cytotoxic drugs (polychemotherapy) significantly improves 15-year survival.^{7–9,14}

However, the absolute survival gain with each of these systemic therapies is modest (5–10%) and each has its own pattern of mild to severe toxicity.

Since 90–95% of women do not benefit from adjuvant systemic therapy [some relapsing despite adjuvant therapy and others never destined to relapse (see example of benefits of tamoxifen – *Figure 2*)] uncertainty has remained about who should be treated and with which therapy or combination of therapies. Women with breast cancer and their clinicians need better information on which patients are most likely to benefit from adjuvant therapy and which are very unlikely to benefit.

Absolute risk reduction during the first 10 years, subdivided by tamoxifen duration and by nodal status [after exclusion of women with oestrogen receptor (ER)-poor disease]¹³

Attempts to select patients for therapy according to prognostic factors is not new; indeed, many of the ‘battles’ of Haagensen and Stout¹⁵ and Crile¹⁶ regarding the extent of surgery in the middle of the last century revolved around patient selection. Over the past several decades, many studies of prognostic factors have been reported, prognostic models have been developed and reviews of this research have been published. Each of these has attempted to identify the degree of risk of relapse

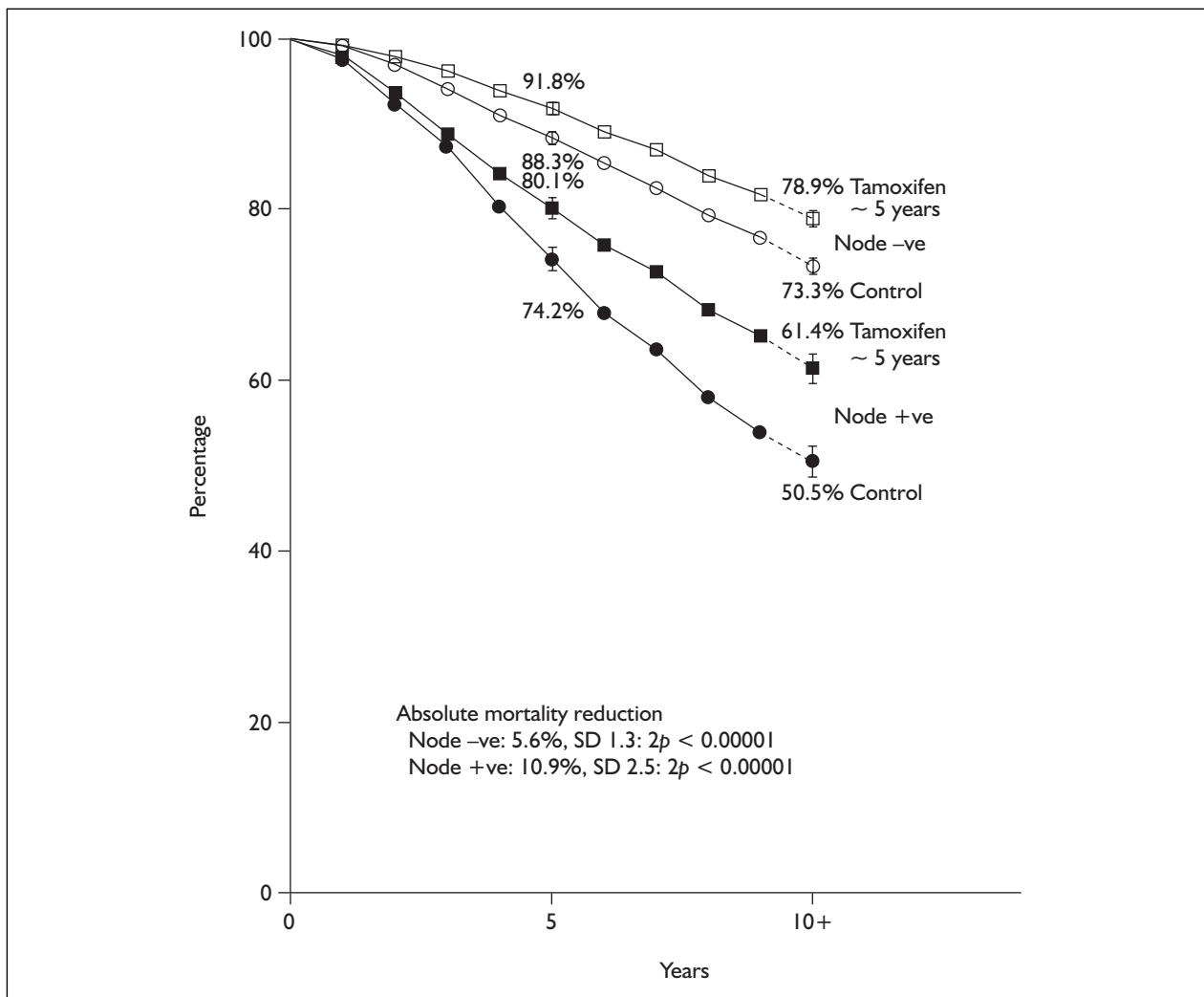


FIGURE 2 Mortality (death from any cause). Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer: an overview of randomised trials. *Lancet* 1998;**351**:1451-67.

and death from breast cancer for individual patients. Latterly there have also been attempts to develop and use predictive factors designed to select patients who are more likely to respond to a specific therapy, for example hormone receptors [ERs and progesterone receptors (PRs)] when using tamoxifen or other hormonal agents.

Policies on how and when to use systemic adjuvant therapy in early breast cancer have been developed internationally and nationally by a series of consensus conferences, often based on the EBCTCG reports.¹⁷⁻²² However, though the results from RCTs and meta-analyses are generally reliable, studies of prognostic factors seem to be of poorer methodological quality and less reliable. Saeki¹⁷ found thousands of reports of varying prognostic factors in the literature, many of which were of poor quality and had not been validated.

Objectives

Prognostic and predictive factors may be used to indicate the presence, status, future behaviour and likelihood of response of women with breast cancer to various therapies. New factors and models are frequently introduced into clinical practice without proper assessment, on the assumption that clinicians will be able to use them to benefit patients. However, inappropriate use of prognostic and predictive factors may produce a worse outcome for patients than if decisions had been made in the absence of this information. Although there have been some systematic attempts to establish guidelines for the use of prognostic and predictive information in breast cancer,²² more work is needed. Moreover, none of these guidelines have examined the cost-effectiveness of basing adjuvant systemic therapy on such information.

Objective of the project

The principal objective of this project was to investigate the cost-effectiveness of using prognostic information to identify patients who should receive adjuvant therapy. This was achieved by developing a decision analytic model to estimate an incremental cost per life-year gained and per quality life-year gained for prognostic groups. This was done by

- identifying prognostic models through systematic review methodology that reliably distinguish clinically important variation in prognosis among groups of women with newly diagnosed breast cancer
- identifying predictive factors through systematic review methodology that reliably predict clinically important variation in response (overall survival, disease-free survival, mortality) to adjuvant therapy amongst groups of breast cancer patients
- surveying the current use of prognostic information and use of adjuvant therapy in the UK
- developing a decision analytic model to integrate the above information with data on costs and quality of life, in order to estimate incremental cost per life-year or per quality adjusted life-year gained.

Six systematic reviews were undertaken to meet the objective criteria outlined above. The titles and a brief description of these reviews are outlined below.

Topic A: Prognostic models in breast cancer

A prognostic model is one that reliably distinguishes clinically important variation (prediction of time to death or recurrence) in prognosis among groups of women with newly diagnosed breast cancer. For the purposes of this question, a model is defined as a combination of two or more factors (clinical, pathological, demographic, molecular), created using statistical methodology. Both primary and validity studies of prognostic models were sought.

Topic B: Predictive factors in breast cancer

A study that investigates patient/tumour factors as possible predictors of response (overall survival, disease-free survival, death) to treatment is the focus of this question. Both individual studies reporting such predictive factors in breast cancer and reviews of these were searched for. The treatment focus for this project was systemic

adjuvant therapy, of which there are only two types: hormonal therapy or chemotherapy.

Topic C: Reviews of prognostic information in breast cancer

There are a large number of papers on cancer examining the prognostic value of individual putative prognostic factors. Currently, however, very few variables are widely accepted as genuinely prognostic and fewer as clinically useful.²² In this question, no attempt was made to identify all prognostic studies in breast cancer. Apart from the enormous volume of literature that this would have created, this method would not have directly answered the search question. This is because what is clinically relevant from studies of new prognostic factors is whether the data provided add to existing clinical information. For this project, reliable information about prognosis needed to include information from multiple studies of one prognostic factor. For this reason, only review papers were sought. Prognostic studies were defined within this project as ones that investigate patient/tumour factors as possible predictors of time to death or recurrence.

Topic D: Quality assessment of prognostic studies (not necessarily cancer)

Underlying the assessment of published studies and reviews of prognostic factors are serious concerns about the quality of prognostic studies. Therefore, the focus of this question was papers that have attempted to develop assessment or scoring schemes to consider the quality of studies of prognosis.

Topic E: The clinical use of prognostic information in breast and other cancers

The focus of this question was the application of prognostic information within a clinical treatment setting. The search sought to find papers that present attempts to use prognostic evidence/information within a clinical setting, be it through audit, unit projects or protocol development. The survey formed part of the response to this research question.

Topic F: Quality of life, cost and cost-effectiveness studies relevant to modelling

This question relates directly to the decision analytic model. Papers were sought that focused on cost, quality of life and cost-effectiveness of adjuvant systemic therapy for breast cancer.

Chapter 2

Systematic review methods

Search strategy

Search strategies for all six questions were created by an information specialist: Julie Glanville, NHS Centre for Reviews and Dissemination, University of York. Members of the project team (who have a wide range of clinical and methodological knowledge and expertise) supplied her with relevant keywords and phrases that were applicable to each of the search questions. The first iterations were constructed from these. Three iterations were required for all search questions before a final search strategy for each question was devised. The search strategies were initially developed on CancerLit (OVID version).

The following electronic databases were searched for relevant published literature for all search questions (details of the search strategy for each of the six research questions are given in Appendix 1):

- MEDLINE (OVID): 1980 to August 2001 for questions A–D, 1990 to August 2001 for question E and 1966 to December 2000 for question F. This search was completed in November 2001.
- CancerLit (OVID: 1995 to March 2001; Dialog: 1980 to 1995): 1980 to March 2001 for questions A–D, 1990 to March 2001 for question E and 1963 to September 2000 for question F. This search was completed in October 2001.
- EMBASE (winSPIRS): 1980 to 2001. This search was completed in January 2002.
- BIOSIS (Biological Abstracts) (winSPIRS): 1980 to 2001. This search was completed in February 2002.
- The Cochrane Library: Issue 1, 2002. This search was completed in February 2002.
- Cochrane Cancer Network's Controlled Trials Register (CCN CTR). This was searched in February 2002.

In addition, specific databases for health economic questions were searched:

- Central Cochrane Controlled Trials Register (CCTR): Issue 1, 2001. This search was completed in August 2001.

- NHS Economic Evaluation Database (NHS EED: Centre Research Dissemination administration database): 1995 to August 2001. This search was completed in August 2001.
- Health Economics Evaluation Database (HEED): searched to April 2001. This search was completed in August 2001.

Finally, the bibliographies of a number of key retrieved articles were checked and the journal *Breast Cancer Research and Treatment* was hand-searched from its inception for further studies relevant to any of the research questions.

The Information Specialist did the searching of all databases for the Health Economics question and the searching of CancerLit and MEDLINE for each question. The searching of EMBASE, BIOSIS, The Cochrane Library and the Cochrane Cancer Network's Controlled Trials Register was done by a member of the project team.

Inclusion and exclusion criteria

Study eligibility forms were developed for the first five of the six research questions (Appendix 2). Eligibility criteria common to all questions were:

- Studies involving humans. All animal, cell line and *in vitro* studies were excluded.
- Case studies, case reports, letters and comments were excluded.
- Questions A–C, E and F were limited to cancer. Questions A–C were limited to early-stage breast cancer (primary, operable cancer, non-metastatic/non-advanced disease, requiring adjuvant therapy only). Question D was not restricted to cancer.
- Studies exclusive to male breast cancer were excluded.
- Search strategies were not restricted to the English language.

Search questions A–E

Initially, one reviewer screened all identified titles and abstracts to assess relevance to the search question. Abstracts were identified as 'relevant', 'maybe relevant' and 'not relevant'. Abstracts that

were completely irrelevant to the search question were removed.

All abstracts that were deemed to be relevant were independently assessed by two other reviewers. Full paper manuscripts for abstracts that were rated as relevant by each of the three reviewers were obtained for data extraction. Abstracts rated as 'maybe relevant' by the initial reviewer were also independently assessed by two other reviewers and full paper manuscripts of those deemed to be relevant were obtained; abstracts deemed to be irrelevant were rejected. Full paper manuscripts of abstracts that were rated as maybe relevant by the two other reviewers were obtained for further assessment. Any disagreements were resolved by consensus.

Study question F

Abstracts of papers identified during searching were assessed by one of the research team and a decision was made on whether to retrieve or reject each paper. Where the potential usefulness of a

paper was unclear, a second member of the team reviewed the abstract and a decision was reached by consensus on whether to accept or reject the paper.

Data extraction strategy

For questions A–E, independent data extraction was done where possible to record information on study methodology and study findings. The procedures followed are described in each chapter but, for all, validated data extraction forms were used (Appendix 2) and disagreements were resolved by discussion. The data extraction forms were developed by the project team members responsible for each specific question. The forms were piloted on a small number of relevant papers before use within the study. Piloting was repeated until a version of each form was created that was judged applicable to the specific question. The forms followed a question and answer structure, with forced choice responses. Free text responses were restricted in order to reduce bias and confusion at the data analysis stage.

Chapter 3

Assessing the methodological quality of prognostic studies

Introduction

A total of 5897 abstracts were initially identified for question D from across four databases [Cochrane Library and CCTR were not searched for question D]: BIOSIS, 2832; CancerLit, 2; MEDLINE, 274; EMBASE, 2789.

Searching for relevant publications was very difficult, as the natural keywords such as quality and prognosis are not discriminating. Early searches produced tens of thousands of references of which almost none was relevant. Searches were refined to reduce the number of hits, but the specificity was still very poor. Of the 274 references identified from MEDLINE only five appeared possibly relevant, of which three had already been identified, one had no abstract and the other was in German. Scanning of titles and abstracts of the first few hundred of the 2832 references from BIOSIS showed that these were almost entirely substantive prognostic studies in various medical fields and did not yield any useful studies. Hence it was not felt valuable to examine all 5000 abstracts.

Seventeen papers were identified from those picked up from the searches for other parts of the project.^{23–39} Several of these were systematic reviews of prognostic studies in various medical fields. Other papers identified in an *ad hoc* way before and after the searching and from examining publications cited in other papers have also been used to prepare this chapter.

Few articles explicitly discussed the desirable methodological attributes of prognostic studies. Several articles draw heavily on the same references, such as Simon and Altman⁴⁰ and McGuire,⁴¹ which have also been used as important sources of the recommendations in this chapter. Because of the nature of the literature, this chapter cannot be regarded as a systematic review in the customary sense.

Assessing methodological quality – study design

There are no widely agreed quality criteria for assessing prognostic studies. Further, there is as yet very little empirical evidence to support the importance of particular study features affecting the reliability of study findings, including the avoidance of bias. Nevertheless, theoretical considerations and common sense point to several methodological aspects that are likely to be important. We consider first generic criteria, likely to apply in all circumstances, and then study-specific criteria, of particular relevance to tumour makers and/or breast cancer.

Generic criteria

Table 2 shows a list of methodological features that are likely to be important for internal validity,^{42,43} which draws on previous suggestions.^{28,29,44–46} The items in *Table 2* are not phrased as questions but rather as domains of likely importance. Most authors have presented their checklists as questions. For example, Laupacis and colleagues²⁸ included the question, “Was there a representative and well-defined sample of patients at a similar point in the course of the disease?”, which includes three elements from *Table 2*. Their checklist is widely quoted, for example in a guide for clinicians,⁴⁷ but it omits several of the items in *Table 2*. Some authors have published checklists for looking at prognostic studies in cancer, such as Melnikow and colleagues,³³ Marras and colleagues³² and Levine and colleagues.²⁹

It is generally agreed that to be reliable (and clinically interpretable), a prognostic study requires a well-defined cohort of patients at the same stage of their disease. Some authors suggest that the sample should be an ‘inception’ cohort of patients very early in the course of the disease (perhaps at diagnosis).⁴⁴ This is just one example of a more general requirement that the cohort can be clearly described, which is necessary for the study to have external validity.

TABLE 2 A framework for assessing the internal validity of articles dealing with prognosis^{43,48}

Study feature	Qualities sought
Sample of patients	Inclusion criteria defined Sample selection explained Adequate description of diagnostic criteria Clinical and demographic characteristics fully described Representative Assembled at a common (usually early) point in the course of their disease Complete
Follow-up of patients	Sufficiently long
Outcome	Objective Unbiased (e.g. assessment blinded to prognostic information) Fully defined Appropriate Known for all or a high proportion of patients
Prognostic variable	Fully defined, including details of method of measurement if relevant Precisely measured Available for all or a high proportion of patients
Analysis	Continuous predictor variable analysed appropriately Statistical adjustment for all important prognostic factors
Treatment subsequent to inclusion in cohort	Fully described Treatment standardised or randomised

Not all prognostic studies relate to patients with overt disease. Both case-control and cross-sectional studies may be used to examine risk factors, but these designs are much weaker than cohort studies. Case-control designs have been shown to yield optimistic results for evaluations of diagnostic tests,⁴⁹ a result which is likely to be relevant to prognostic studies. In cross-sectional studies, it may be very difficult to determine whether the exposure or outcome came first, for example in studies examining the association between oral contraceptive use and HIV infection.

Most authors of checklists have not considered the issue of subsequent treatment. If the treatment received varies in relation to prognostic variables, then the study cannot deliver an unbiased and meaningful assessment of prognostic ability unless the different treatments are equally effective (in which case why vary the treatment?). Such variation in treatment may be common once there is some evidence that a variable is prognostic. Ideally, therefore, prognostic variables should be evaluated in a cohort of patients treated in the same way or in a randomised trial.^{45,50} Such studies are relatively rare; in practice, it is likely that prognostic factors do influence treatment. For this reason, including type of treatment in the model may not have much impact when the important prognostic variables are also included.

The important methodological dimensions will vary to some extent according to circumstances. For example, in some prognostic studies the reliability of the measurements may be of particular importance. Many biochemical markers can be measured by a variety of methods (such as assays), and studies comparing these often show that the agreement is not especially good. It is desirable, therefore, that the method of measurement is stated and that the same method was used throughout a study; this information is often not given explicitly.

Adequate sample size is equally as important for prognostic factor studies as for clinical trials. Sample size has received little attention in most such studies, however, because they are usually performed on already available data sets. Also, sample size calculations with survival data are complex and need to consider, among other things, the length of follow-up and the number of expected events. Sample size calculations can be simplified by considering the power to detect a specified outcome difference at a fixed point in time such as the 2-year survival rate. There are some simple approaches to sample size estimation.^{51,52}

Bentzen⁵³ considered the low power of small studies. He found that three-quarters of 47 papers reporting prognostic studies in osteosarcoma had fewer than 100 cases. He used bootstrapping to

illustrate how small studies would be likely to miss important prognostic effects. Small studies are liable to yield unreliable results, and those showing a large prognostic impact of some favoured tumour marker are more likely to be published than those that do not. Such publication bias is well recognised in randomised trials,⁵⁴ and it seems certain that this bias afflicts prognostic studies to a great extent. As Simon⁵⁵ wrote, "... the literature is probably cluttered with false-positive studies that would not have been submitted or published if the results had come out differently." There is as yet little clear evidence of publication bias in prognostic studies, but it has been clearly shown in studies of Barrett's oesophagus as a risk factor for cancer⁵⁶ and has been strongly suspected in other reviews of prognostic studies, such as that of Popat and colleagues.⁵⁷

Context-specific criteria

The inclusion of context-specific in addition to generic aspects of methodological quality is sometimes desirable. For example, Marx and Marx⁴⁶ included two questions on the nature of the end-points, reflecting particular problems encountered in their review of prognosis of idiopathic membranous nephropathy, where many studies used ill-defined surrogate end-points.

Brocklehurst and French⁵⁸ considered whether there was an adequate description of the maternal stage of disease.

There are various aspects of prognostic studies that are specific to breast cancer, such as whether the study included screen-detected cancers and the specific assay used. Such factors impact on clinical and statistical heterogeneity; we do not consider that any are relevant to an assessment of methodological quality.

Assessing methodological quality – analysis

The criteria in *Table 2* include two items relating to difficult aspects of data analysis – adjustment for other variables and the analysis of continuous prognostic variables. Here we consider in some detail these important issues, which have a major influence on whether any meta-analysis might be possible. *Table 3* shows recommendations of Altman and Lyman⁵⁹ regarding analysis of prognostic factor studies. *Table 4* shows the more detailed recommendations of Riley and colleagues;⁶⁰ these cover various aspects of analysis although labelled as reporting guidelines.

TABLE 3 Recommendations for the analysis and reporting of prognostic factor studies⁵⁹

Action	Recommendations
Analysis	<ol style="list-style-type: none"> 1. Base analysis, including any hypothesis testing, on the primary and major secondary outcomes specified prior to the study 2. Consider possible bias due to missing data 3. Consider the issue of multiple comparisons when evaluating many prognostic factors or cut-points and adjust tests of significance accordingly 4. Beware of the problems associated with the interpretation of stepwise multiple regression models, including model instability and likely exaggeration of coefficient estimates and their associated <i>p</i>-values 5. Adjust the effect of new prognostic factors for existing prognostic factors of recognised and accepted importance 6. Outcome differences between subgroups should be assessed by testing the interaction between the prognostic factor and the variable defining the subgroups rather than by separate analyses within subgroups 7. Interpret with caution apparent outcome or prognostic marker differences between subgroups (many such differences arise from multiple testing or small sample size within subgroups) 8. Analysis of subgroups defined only during or after completion of the study should be acknowledged as exploratory
Reporting	<ol style="list-style-type: none"> 9. Clearly state the study design: exploratory/confirmatory, prospective/retrospective, treatment (e.g. randomised or standardised), blinding, main outcomes, etc. 10. Report the number of patients excluded because of missing data 11. Specify study duration including criteria for study termination (if relevant) 12. Report methods of measurement of prognostic markers, if possible with information about reproducibility 13. Define clearly all study end-points 14. Summarise outcomes as quantitative estimates and confidence intervals 15. Emphasise the outcome differences observed for all patients more than those found among subgroups 16. Discuss any weaknesses of the study, especially related to subgroup analyses and multiple comparisons

TABLE 4 Guidelines for the reporting of the results of prognostic factor studies⁶⁰

Results of all the marker analyses should be presented – both significant and non-significant – and we recommend the following:

Essential to present:

1. The **hazard ratio** and its **confidence interval**, or the $\ln(\text{hazard ratio})$ and its variance. Markers that have a continuous function should be modelled as a continuous variable using appropriate methods. If there is a justifiable reason for using a cutoff level for a continuous marker it should be specified **at the start** of the study and be clearly reported.
2. The **number of patients** and **number of events** in total. For binary markers (and continuous markers if a cutoff level is used) also report the numbers within each group.
3. Both **unadjusted** and **adjusted** results for each marker. For adjusted results, clearly state what variables have been adjusted for. Ideally, a consistency in the set of adjustment factors used across studies should be sought through **collaborative groups** working toward prospectively planned pooled analyses. Otherwise, (i) always present results adjusted for age and stage of disease, and (ii) consider using the same set of adjustment factors as in important earlier studies.
4. **Individual patient data** in the paper or on the Internet, or make available with details clearly indicated within the paper. Data on markers that were not analysed should be included. Subject to any restrictions imposed by data protection laws and guidelines, include:
 - exact initial marker level and how marker was measured
 - time of disease recurrence (if appropriate)
 - follow-up time
 - final disease status
 - levels of other existing prognostic markers of recognised and accepted importance for current clinical practice
 - patient subgroup information, e.g. age, stage of disease, type of treatment received
 - details of inclusion/exclusion criteria would also be beneficial

Highly desirable to present:

5. **Exact p -values.** Reporting of results as ‘significant’ or ‘not significant’ is insufficient. Very small p -values can be given as $p < x$ (e.g. $p < 0.0001$), but in this case the exact χ^2 -statistic is also needed.
6. **Survival curves** showing the difference in survival over time between the groups, with clear **step** and **censoring points**; also the initial numbers in each group, and the number of events and remaining numbers at various time points during follow-up are needed.
7. **% survival at n years** with a confidence interval using Kaplan–Meier or other methods that allow for censoring, together with the number of patients at risk at that time in each group.

Adjustment for covariates

It is important to adjust for other prognostic variables to obtain a valid picture of the relative prognosis for different values of the primary prognostic variable. This procedure is often referred to as control of confounding. It is necessary because patients with different values of the covariate of primary interest are likely to differ with respect to other prognostic variables. Also, in contexts where much is known about prognosis, such as breast cancer, it is important to know whether the variable of primary interest (such as a new tumour marker) offers prognostic value over and above that which can be achieved with previously identified prognostic variables. It follows that prognostic studies require multiple regression analysis and, as such studies have outcomes that are times to a specific event, survival analysis methods are necessary. Cox proportional hazards regression models are almost universally used in this context, although other methods exist – they are discussed in Chapter 5.

Many studies seek parsimonious prediction models by retaining only the ‘most important’ prognostic factors, most commonly selected using multiple regression analysis with stepwise variable selection. Unfortunately, this method may be likely to be misleading.⁴⁰ In the context of exploring a particular prognostic factor, such methods are not appropriate. Rather, recognised (‘standard’) prognostic factors should be adjusted for, regardless of whether they reach specified levels of significance.⁵⁹ The model with the marker and the standard variables provides an estimate of its independent effect and a test of whether it contributes additional prognostic information over and above the standard factors.

When comparing multiple studies, as in a systematic review, it is customary to find that different researchers use a variety of statistical approaches to adjustment, and that they adjust for different selections of variables. Such variation severely hinders simple interpretation. Some authors have argued that this situation indicates

TABLE 5 Problems with systematic reviews of reports of prognostic studies^{43,48}

Difficulty of identifying all studies
Negative (non-significant) results may not be reported (publication bias)
Inadequate reporting of methods
Variation in study design
Most studies are retrospective
Variation in inclusion criteria
Lack of recognised criteria for quality assessment
Different assays/measurement techniques
Variation in methods of analysis
Differing methods of handling of continuous variables (some data dependent)
Different statistical methods of adjustment
Adjustment for different sets of variables
Inadequate reporting of quantitative information on outcome
Variation in presentation of results (e.g. survival at different time points)

that one should extract, and possibly meta-analyse, only unadjusted estimates.⁶¹ Although such estimates may be more comparable, unadjusted estimates are not useful, so it seems preferable to extract, and possibly combine, estimates where some attempt was made to adjust for known factors, however imperfectly this was done. That said, it is likely that estimates from studies that had used some data-dependent methods, especially the choice of cut-points, will yield inflated estimates of association with outcome. With enough studies, such bias might be detected in a comparison of studies using different methods.

Table 5 gives a list of possible difficulties encountered when carrying out a systematic review of prognostic studies. Many of the difficulties are illustrated in the ironically titled review of the evidence about many prognostic factors in non-small cell lung cancer:⁶²

“... for many factors, the strength of the independent association of that factor with survival outcomes is also quite variable. There are many potential reasons for this observed heterogeneity. Some studies are clearly statistically underpowered, given that the median number of patients enrolled per study was only 120 (range, 31 to 1281 patients). In addition, variation in the case mix of the study populations, variation in the other explanatory variables included in each analysis, and variation in the methods used to define and quantify prognostic factors will be reflected in the heterogeneity of the apparent strength of association between the prognostic factor and the relevant outcome ... These inter-study differences have resulted in some controversy regarding the clinical value of many prognostic factors, and our review illustrates that this controversy exists for many factors.”

Poor reporting of primary studies is a difficulty commonly encountered in systematic reviews. It is a greater problem for prognostic studies than randomised trials, especially when survival times are analysed, for many of which the results are presented only graphically (unadjusted effects). Methods exist to estimate the logarithm of the hazard ratio (lnHR) [and standard error (SE)] from Kaplan–Meier graphs,⁶³ but these methods cannot always be used and they make unverifiable assumptions (especially when estimating the SE). Also, the methods provide unadjusted estimates, and are therefore much more suited to randomised trials than observational studies.

Riley and colleagues⁶⁰ reviewed prognostic markers for neuroblastoma. There were 575 reports of markers in 211 papers, for which the lnHR and its SE were directly obtainable for just three markers, all reported in a single paper. Even using 10 different methods of extracting data, estimates of the lnHR and its SE were obtained from just 204 of the 575 reports (35%).

Handling continuous predictor variables

Most putative prognostic markers in cancer are continuous measurements. If such a variable were truly prognostic, the risk of an event would usually be expected to increase or decrease systematically as the level increases. Nonetheless, many researchers prefer to categorise patients into high- and low-risk groups based on a threshold or cut-point. This type of analysis discards potentially important quantitative information and considerably reduces the power to detect a real association with outcome.^{64,65} Reasonable approaches to choosing a cut-point include using a cut-point reported in another study, one based on the reference interval in healthy individuals or the

median (or other prespecified centile) from the present study.

Cut-points should not be determined by a data-dependent process. Some investigators compute the statistical significance level for all possible cut-points and then select the cut-point giving the smallest p -value. There are several serious problems associated with this so-called 'optimal' cut-point approach.^{59,66} In particular, the p -values and regression coefficients resulting from these analyses are biased and, in general, the prognostic value of the variable of interest will be overestimated. The bias cannot be adjusted for in any simple manner, and it is carried across into subsequent multiple regression analyses. Misleading results from individual studies are bad enough, but when such studies are included in a meta-analysis they may well distort the results.

Keeping variables continuous in the analysis has the considerable advantages of retaining all the information and avoiding arbitrary cut-points. It may also greatly simplify any subsequent meta-analysis. Many researchers, however, are unwilling to assume that the relationship of marker with outcome is log-linear, that is, that the risk (expressed as lnHR) either increases or decreases linearly as the variable increases. The assumption of linearity may well be more reasonable than the assumptions that go with dichotomising, namely constant risk either side of the cut-point. Investigations of non-linear (curved) relationships are uncommon. An example is the modelling of Breslow thickness in melanoma.⁶⁷ Some studies have addressed this issue in breast cancer – examples are Hilsenbeck and colleagues⁶⁸ and Sauerbrei and colleagues.⁶⁹

Using a small number of groups, say four, offers a good compromise between dichotomising and treating the data as continuous, which requires assumptions about the shape of the relation with the probability of the event. This approach is common in epidemiology. However, it may lead to problems for the systematic reviewer, because it is rare that different studies use the same groupings. For example, Buettner and colleagues⁷⁰ summarised 14 studies that had examined the prognostic importance of tumour thickness in primary cutaneous melanoma. The number of cut-points used varied between studies and no two studies had used exactly the same cut-points. Further, several studies had used the 'optimised' approach that, as noted above, is inherently overoptimistic.

Assessing methodological quality – overall assessment

One common approach to assessing quality is to derive a quality score, based on a checklist of items that are sought from each paper. For example, Steels and colleagues⁷¹ developed a scoring system with up to 10 points given in each of four domains: scientific design, laboratory methodology, generalisability and analysis. Scores were rescaled as a percentage of the maximum. In a review of 74 prognostic studies, they found a median score of 55%.

Quality scores are problematic, however, as they generally combine information that is important to credibility of the results (in the sense of being associated with the risk of bias) with other information.^{72,73} In general, they mix quality of methodology with quality of reporting and perhaps also aspects of external validity. Many of the items listed by Steels and colleagues⁷¹ are things one would wish to find in a paper, but their absence would not of itself suggest that the study results were untrustworthy. Examples include description of positive and negative control procedures and summary of the length of follow-up. Unfortunately, it is the norm that the quality of research (of all types) is hard, if not impossible, to assess properly because of incomplete reporting.

For these reasons, attempts to investigate the possible association between methodological quality and study outcomes have turned more towards investigation of specific methodological features using subgroup analysis, for one variable at a time, or meta-regression, for multiple variables simultaneously.⁷³ In the current context, this approach would indicate looking directly at aspects of study design, including laboratory methods and choice of cut-points.

Reporting prognostic studies

As noted already, the reporting of prognostic studies has frequently been found to be inadequate. In addition to patient outcomes, discussed above, essential information about study methods and patient characteristics is frequently reported inadequately.⁷⁴ As Andersen⁷⁵ noted: "It is about as informative to say 'that a multifactorial analysis was performed using the proportional hazards method proposed by Cox and the variables were entered in a stepwise fashion' as it is to say that 'the manuscript was typed using an

electric typewriter’.” Several authors have made suggestions regarding the information that should be reported.^{59,60,74}

Table 3 shows reporting recommendations of Altman and Lyman.⁵⁹ Riley and colleagues⁶⁰ gave detailed suggestions for the presentation of the results of such studies, with one aim being the facilitation of possible subsequent meta-analysis (*Table 4*). Forthcoming reporting guidelines aimed at prognostic studies of tumour markers outline the elements of good reporting.⁷⁶ Support for and implementation of these recommendations by cancer journals would be a valuable step to improving the quality of the published prognostic literature.

Predictive markers

In the preceding discussion, we have made no distinction between prognostic and predictive markers. All the concerns expressed so far apply to both types of study. However, there are reasons to be even more concerned about the reliability of published information regarding predictive markers.

Showing that a marker predicts response to treatment effectively requires a test of marker by treatment interaction. As such, the power available to detect such an effect is markedly lower than for detecting a prognostic effect – roughly four times the sample size is necessary to have equivalent power. Analyses based on randomised trials are not protected from the dangers.

A major concern is that few such analyses were prespecified, but rather based on exploratory analyses and possibly also selected from among several such analyses. A test of interaction is exactly equivalent to a comparison of subgroups (here, those with or without a raised marker level);

much has been written about the dangers of their interpretation.⁷⁷⁻⁷⁹ In general, such findings are of little value unless demonstrated in a study in which the particular analysis was prespecified. Otherwise, such studies are exploratory and require confirmation in further studies – an example is given by Royston and colleagues.⁸⁰

Discussion

As a consequence of the poor quality of research, prognostic markers may remain under investigation for many years after initial studies without any resolution of the uncertainty. Multiple separate and uncoordinated studies may actually delay the process of defining the role of prognostic markers.

Systematic reviews of published studies generally find a confusing picture, and that many studies are poorly done and poorly reported. They can usually not answer questions but rather draw attention to the paucity of good quality evidence and thus help to improve the quality of future research. A typical summary is that of Schmitz-Dräger and colleagues:⁸¹

“From this analysis it becomes evident that further retrospective investigations will not contribute to the solution of the problem and thus are obsolete. There is an obvious need for standardization of the assay procedure and the assessment of the specimens as well as for the initiation of a prospective multicenter trial to provide definite answers.”

Cooperation from the outset between different research groups could lead to clear results emerging more rapidly than is commonly the case, especially if such efforts are put into prospective studies or retrospective studies based on individual data from carefully assembled databases and/or tissue banks.⁸²

Chapter 4

Systematic review of reviews of studies of prognostic factors

Introduction

The aim of this systematic review of reviews was to identify good-quality systematic reviews of studies of factors that might be prognostic for women with operable breast cancer, with a focus on those that were found to be routinely available in the survey of current practice (see Chapter 7). Given the plethora of studies of prognostic factors, it was not feasible, within the confines of this project, to seek to identify, summarise and act on the findings of individual studies.

This chapter is structured in such a way that the process for identifying relevant reviews is described and then each factor for which at least one relevant review was found is described in detail. For each factor, information is provided on the prognostic factor, the scope and quality of the identified review(s), the conclusions and opinion of the strength of these conclusions. This has been done so as to allow readers who wish to concentrate on specific prognostic factors to do so easily by taking the relevant section as a standalone piece of text. Some of the studies included in these reviews and hence, some of the systematic reviews included here – either through design or because of the limitations of the studies available – focused on a ‘target’ prognostic factor within a subset of women as categorised by another prognostic factor (e.g. women who had node-negative breast cancer). Where this is the case, it is highlighted below.

Our systematic review of factors that might be predictive of response to treatment is the subject of Chapter 6. That chapter considers both previous reviews of predictive factors and also, unlike this chapter, individual studies of such factors.

Searching and selection

A description of the searching that was done to identify reviews of prognostic factors in breast cancer is given in Chapter 2. The electronic searches returned a total of 996 records. Of these, 113 were from MEDLINE, 99 from EMBASE, 220

from BIOSIS, 504 from CancerLit and 60 from the Cochrane Central Register of Controlled Trials (CENTRAL) in The Cochrane Library. Following careful review of the titles and abstracts of each of these records by at least two of three members of the research team (SB, MC and CW), a total of 128 full articles were obtained. Again, at least two of these three then assessed each of these articles for relevance and extracted data using an agreed and piloted data extraction form (Appendix 2). This resulted in the identification of a core set of 25 articles containing systematic reviews of studies of prognostic factors in women with operable breast cancer. These articles included a total of 49 separate reviews (43 of which are discussed in detail in this chapter), with some of the articles reporting reviews of more than one factor. For example, the article by Mirza and colleagues included relevant systematic reviews for each of 11 separate factors.⁸³ Many non-systematic reviews were found in which a description of a prognostic factor was given along with examples of studies that had looked for an association with, for example, disease-free or overall survival. Such non-systematic reviews have been included within this chapter only if a systematic review of the relevant prognostic factor had not been identified.

The following domains were assessed in judging the quality of each review:

- extent of searching (including databases and any language or time period restrictions)
- description of the eligibility criteria for the review
- comparability of the included studies
- assessment of publication bias
- assessment of heterogeneity
- conduct of sensitivity analyses.

These assessments were done by at least two of three members of the research team (SB, MC and CW). Disagreements were uncommon and were resolved by discussion between the two people who had done the assessment, with recourse to the third person in the rare circumstances where agreement could not be reached between the two. It was always possible to reach a consensus among

the three. The assessment, along with the data extraction, formed the basis for deciding on the inclusion of reviews in this chapter. Information from the assessment and data extraction were supplemented by use of the original articles in the drafting of this chapter by the member of the research team responsible for this chapter (MC).

As with all forms of systematic review, it is possible that this review of reviews is subject to publication bias – in that the findings of the reviews that have been published might be systematically different to the results of any reviews that have been conducted but not published. Unfortunately, however, it is not possible to investigate this in depth since there is no equivalent to a prospective register of randomised trials for reviews of prognostic studies – just as there is no prospective register of the prognostic studies themselves. The availability of reviews for this systematic review depends, therefore, on both the willingness of the original reviewers to write up their work and of a journal to publish it.

It is hoped that the results of the reviews that have been carried out will not have been too strong a determinant in either decision and, because of the extent of the searching, the authors tried to maximise the likelihood of finding relevant reviews from a wide range of healthcare publications. In addition, because the aim of this systematic review of reviews was to identify good-quality systematic reviews of studies of factors that might be prognostic for women with operable breast cancer, extensive searching for unpublished reviews was not undertaken because it is unlikely to have been cost-effective. This is because of the additional time and resources that would have been required to search for unpublished research of this type, the difficulties likely to be encountered in assessing adequately any unpublished reviews that might have been found and the lack of any intention on our part to try to produce a numerical estimate of an ‘average result’ from the reviews included.

Individual factors

The most commonly used, and currently the most effective, prognostic factors in breast cancer have been little studied in systematic reviews. This is perhaps akin to the lack of systematic reviews, or even randomised trials, for many interventions whose effectiveness is striking enough to be beyond doubt. The prognostic factors to which this applies include nodal status and age at diagnosis.

Rather, the reviews that were identified are of factors for which the evidence is, in general, less robust. In many of these reviews, the importance given to the above two factors is apparent, in that the focus of the reviews might be of a factor within a group of women whose breast cancer has already been categorised using one of the above. This most commonly related to nodal status, with several reviews seeking to investigate additional prognostic factors for women with node-negative breast cancer.

Eligible reviews were identified for the following prognostic factors. As noted above, some of the identified reports included reviews of more than one factor.

- oestrogen receptor (ER) status (1 review)
- pS2 (1 review)
- epidermal growth factor receptor (EGFR) (2 reviews)
- human EGFR family 2 (HER2, CerbB2) (6 reviews)
- C-MYC amplification (2 reviews)
- cell proliferation markers (10 reviews)
- aneuploidy or DNA ploidy (2 reviews)
- p53 (5 reviews)
- p21 (2 reviews)
- bcl-2 (1 review)
- cathepsin-D (3 reviews)
- urokinase-type plasminogen activator (uPA) (7 reviews)
- nm23 (1 review)
- bone marrow micrometastases and minimal residual disease (2 reviews)
- tumour size (1 review)
- tumour grade (1 review)
- vascular invasion (1 review)
- body size (1 review).

The following structure is used to describe the reviews found for each prognostic factor:

- brief description of a ‘class’ of prognostic factor
- brief description of the reviews identified for each of the prognostic factors in this ‘class’
- scope of the reviews
- quality of the reviews
- conclusions of the reviews and the strength of the conclusions.

Class of prognostic factor: oestrogen receptor pathways

The ER is of crucial importance in breast cancer, both as a cause of growth of breast cancer when

stimulated by oestrogen and as a therapeutic target to inhibit growth. It is a nuclear steroid receptor and can be measured by several methods, the older ones being ligand binding, more recently immunoassays and now immunochemistry, which can be performed on routine histological sections of tumour.

When activated, the ER switches on the PR and many other genes, including a secreted factor pS2. Thus, the latter two genes act as markers for active ER signalling. If the ER is too low to detect, the PR may still be detectable and measurement of this can provide evidence of an active oestrogen responsive element signalling pathway, warranting antioestrogen therapy.

Prognostic factor: oestrogen receptor status

One review⁸³ of studies of the prognostic nature of ER status was identified. This was within a broader review of prognostic factors in node-negative breast cancer.

Scope of the review

The review only included studies in women with node-negative breast cancer. It was restricted to studies with at least 200 women and 5 years of follow-up. The review includes seven studies (4192 women). The largest of these involved 1157 women. Studies were identified for the review through a MEDLINE search for articles published during 1996–2000. However, studies published before then were also included, but no information is given on how these were selected, and there were two such studies in this review of ER status, including the largest. The review focused on the relationship between ER-negative status and disease-free survival and overall survival.

Quality of the review

Although the eligibility criteria for the review were stated, no details were given of how these criteria were applied. The review relied on the published results of the studies identified and no additional information appears to have been sought from the original researchers.

There is no discussion of the quality of the studies that were brought together in the review and the possible impact of publication bias is not discussed. Similarly, there is no direct discussion of heterogeneity but details are given of the techniques used to assess ER status and of the use of systemic therapy, which may be of particular relevance given the relationship between ER status and tamoxifen. In six of the seven studies, no

systemic therapy was used. In the seventh, about two-thirds of patients received systemic therapy but none of these women received tamoxifen.

Conclusions of the review and the strength of the conclusions

Four of the seven studies conducted a univariate analysis of the relationship between ER status and disease-free or overall survival. Three of these showed a significant relationship, but one did not. Four studies used a multivariate analysis. ER-negative status was significantly associated with worse survival in two studies and was not significantly associated in two others.

On the basis of these findings, Mirza and colleagues⁸³ conclude that ER status is one of the prognostic factors for which there are 'mixed results'. This conclusion, that an association between ER status and prognosis in women with node-negative breast cancer has not been confirmed or refuted by the studies in the review, is justified.

Prognostic factor: ps2

One review⁸⁴ of studies of the prognostic nature of ps2 was identified. This was within a review of the prognostic relevance of biological markers in general.

Scope of the review

The review contains two studies, both published in the early 1990s. The total number of breast cancer patients in these studies is not clearly reported. It is unclear if any eligibility criteria were set or applied by the reviewers.

Quality of the review

The search strategy, eligibility criteria and the use of any quality assessment for the included studies are not reported for the identified review. It relies on the published results of the two studies identified.

Conclusions of the review and the strength of the conclusions

The review reported that both studies showed that ps2 was associated with poorer prognosis, in terms of disease-free survival and overall survival. Given the lack of information on the methods of this review, it is not possible to assess the quality of the review itself or whether the two studies identified are a complete sample of the worldwide evidence as it stood in the early 1990s. The findings of the review must be treated with caution and additional research would be needed to confirm or refute any association between ps2 and prognosis.

Class of prognostic factor: growth factor pathways

About one-third of breast cancers do not express the ER status and are regulated by other growth factor pathways. Many of these signal via transmembrane receptors, the extracellular domain binding the growth factor, the intracellular domain activating a signalling cascade. One receptor commonly expressed is the EGFR. Another in the same family is HER2. Both tend to be reciprocally expressed with ER, although all combinations of expression occur. These growth factor receptors can be inhibited by drugs or antibodies and are an important target for therapy in breast cancer. The gene coding HER2 is often amplified (i.e. multiple copies in the cancer cells) and this genetic change is strong evidence for a major role in the tumours that express it. Both receptors are associated with stimulation of growth and invasion in breast cancer cell lines, hence their investigation for a direct role in the behaviour of breast cancer.

Up-regulation of transcription factors and nuclear proteins that regulate the cell cycle is another mechanism of transformation and the gene for C-MYC is commonly amplified in breast cancer. In general, it is considered that there are at least five genetic events involved in transforming a normal cell into a cancer cell, so most tumours will have many pathways abnormally regulated.

Prognostic factor: epidermal growth factor receptor

Two reviews^{85,86} were identified of studies of the prognostic nature of EGFR. One of these was part of the presentation of the detailed results for a series of patients⁸⁵ and the other was a standalone review of EGFR.⁸⁶

Scope of the reviews

The Klijn review⁸⁶ examined the association of EGFR with other prognostic factors and with clinical outcomes. It identified 40 relevant series of patients, including a total of 5232 women with breast cancer. Among these, 2500 of the tumours (48%) were positive for EGFR. However, of the 40 studies, only 9 research groups (1728 patients) had looked at relationship between EGFR and survival. These studies ranged in size from 55 to 376 patients and in EGFR positivity from 14 to 55% and had a median follow-up of 12–66 months. The studies included both node-positive and node-negative patients.

The Fox review⁸⁵ is reported in the context of a new series of 370 patients. It cites the Klijn review,⁸⁶ uses many of the same references, but includes a total of 16 series (3009 patients) with follow-up data and survival analysis. The reports in the Fox review⁸⁵ (including the reviewers' own study) that were not in the Klijn review⁸⁶ were all published after 1990. All studies in the Klijn review⁸⁶ were published in that year or earlier. Both reviews examined EGFR in relation to recurrence- or disease-free survival and overall survival.

No details of the searching are provided for the Fox review,⁸⁵ but the reviewers do report that they used the largest series with the longest follow-up where more than one report had been published for the same series of patients. The searching for the Klijn review⁸⁶ is reported to have involved a search of the MEDLINE/EBSCO database but no details are given of the search terms used or of any limitation by language or publication year. The reviewers wrote that they used the "most representative series of patients reported in the most recent paper from each group containing the largest series of patients".

Quality of the reviews

The eligibility criteria for including studies were not clearly stated in either review. Both reviews appear to have relied exclusively on the published results of the studies they identified but the Fox review⁸⁵ did, of course, have access to the raw data from their own study. Neither discussed the quality of the studies that were brought together, or the possible impact of publication bias on their findings.

Conclusions of the reviews and the strength of the conclusions

Much of the report of the Klijn review,⁸⁶ and most of the 40 included studies, are related to correlations between EGFR and other prognostic factors. The strongest association appears to be between EGFR and steroid receptor levels, with a negative association between EGFR and ER or PR levels. From the nine studies that reported on the relationship between EGFR and survival, five found a significant relationship between EGFR and overall survival or recurrence-free survival (RFS) using univariate analyses, and two of three that investigated this, found the same for multivariate analyses. However, the reviewers note that the effect appeared strongest in those studies with shortest follow-up and they conclude, "There is little agreement on the prognostic value of EGFR, with most studies indicating a tendency or weak association between EGFR and RFS [relapse-free survival] or OS [overall survival]".

In the Fox review,⁸⁵ 12 of the 16 studies that undertook a univariate analyses found a significant association between EGFR and disease-free or overall survival. However, the reviewers also point out the problems caused by short follow-up and note that the greater effect on relapse than on survival might be because of this. Ten and five studies conducted multivariate analyses of RFS and overall survival, with significant associations with EGFR being found in six and one of these, respectively. The reviewers conclude that many of the studies were too small, with only six having more than 200 patients and three more than 300 patients. They report that “firm conclusions cannot be drawn” about the association between EGFR and survival.

The Klijn review⁸⁶ is a comprehensive account of EGFR and its associations with other prognostic factors and survival. However, there is insufficient information on searching and the eligibility and assessment of studies to assess its quality. The Fox review⁸⁵ contains less information on the review process, but more relevant data. However, the caution expressed in both reviews about the association between EGFR and survival is justified and requires further study.

Prognostic factor: human EGF receptor family 2 (HER2)

Six reviews were identified of studies of the prognostic nature of HER2. These included three standalone reviews^{87–89} and one within a general discussion of prognostic and predictive factors.⁹⁰ A fifth review investigated HER2 within a wide-ranging review of prognostic factors in node-negative breast cancer.⁸³ The Porter-Jordan review⁸⁴ of biological markers in general also contained a section on HER2 but this has been superseded by the more recent reviews and is not discussed further here.

In addition, although the Nunes review⁸⁹ is the most recent and includes some studies that were published in the 2000s (i.e. after the other reviews were published), it is not considered further here. It provides little or no information on how studies were sought, assessed for eligibility or quality, and included. As is shown below, some of the other reviews found dozens of studies of HER2 and prognosis but the Nunes review includes only eight studies with a total of 1871 patients, all of whom received some form of systemic therapy. All these studies show a significant association between HER2 and shorter disease-free or overall survival, in either univariate or multivariate analyses.

Scope of the reviews

The Henderson review⁹⁰ includes a table showing 19 studies, with a total of 4653 patients, which were reported in articles published between 1989 and 1998. No details of the searching are provided for this review.

The Révillion review⁸⁷ reports that the CancerLit database was searched for relevant studies, but does not give details of the search terms used or any restrictions based on, for example, language or year of publication. Ninety-seven studies, involving 22,616 women, were identified, with the relationship between HER2 and prognosis being described in 34 of these studies. However, the way in which the review is reported makes it difficult to identify how many women were involved in total and how many of these studies contributed to the review’s assessment of the association with prognosis. Where a study was reported more than once, the most recent paper was used. The review discusses node-negative and node-positive women separately and provides details on the relevant case mix for the included studies.

The Ross review⁸⁸ includes 52 studies with a total of 16,975 patients assessing the association between HER2/neu gene amplification or HER2 protein over-expression and disease outcome. These studies were published between 1987 and 1998. No details of the searching are provided for this review.

The Mirza review⁸³ only included studies in women with node-negative breast cancer. It was restricted to studies with at least 200 women and 5 years of follow-up. The review includes 13 studies (4996 women) of the association of HER2-increased expression with disease-free survival and overall survival. Studies were sought for the review through a MEDLINE search for articles published during 1996–2000. However, studies published before then were also included although no information is given on how these were selected. Nine of the 13 included studies were published before this period, meaning that the publication years for the included studies span 1991 to 2000.

Quality of the reviews

No details of the eligibility criteria or assessment process are given for the Henderson,⁹⁰ Ross⁸⁸ or Révillion⁸⁷ reviews.

The Ross review⁸⁸ does not contain a direct discussion of heterogeneity but, in discussing the findings of the studies in relation to the significance of the association between HER2 and prognosis,

the reviewers draw attention to details of the laboratory measurement of HER2, including the methods used to store the samples. The Révillion review⁸⁷ does not discuss heterogeneity in detail but, as shown below, it does present separate results for patients who were node negative and node positive. It also comments on the wide variations in HER2 positivity in the included studies.

Although the eligibility criteria for the Mirza review⁸³ were stated briefly, no details were given of how these criteria were applied. There is no discussion of the quality of the studies that were brought together in the review and the possible impact of publication bias is not discussed. Similarly, there is no direct discussion of heterogeneity but mention is made of the use of systemic therapy in each study and the different ways of measuring HER2 expression.

All four of the reviews discussed in detail here seem to have relied on the published results of the studies identified, and no additional information appears to have been sought from the original researchers. This is also true for the other two reviews of HER2 and prognosis that are not dealt with in detail.^{84,89}

Conclusions of the reviews and the strength of the conclusions

The table of studies in the Henderson review⁹⁰ shows the univariate *p*-value for disease-free survival and/or overall survival for each study. Most of the studies showed a statistically significant association between worse outcome and HER2 level and the reviewers report “there is now wide acceptance that c-erbB-2 amplification and over-expression are associated with worse disease-free and overall survival”. They add “whether it is prognostic for node negative patients is controversial”.

The Ross review⁸⁸ adopts a ‘vote-counting’ approach to assessing the strength of any association between HER2 and prognosis. It reports that six of the 52 studies (1222 of the 16,975 patients) showed no correlation between HER2 status and patient outcome. Thirteen of the studies (3884 patients) reported prognostic significance on univariate analyses only (seven of which did not conduct a multivariate analyses) and 33 studies (11,869 patients) reported independent prognostic significance on multivariate analyses. The reviewer concludes that “the preponderance of evidence indicates that HER-2/neu gene amplification and protein over-expression are associated with an adverse outcome in breast

cancer”, but does not differentiate between node-negative and node-positive patients in this regard.

The Révillion⁸⁷ review also uses a ‘vote-counting’ approach and reports the number of studies that found a significant or non-significant relationship between HER2 and prognosis in node-negative and node-positive patients, dealing with RFS and overall survival separately. For node-negative patients, most of the 13 studies (3750 patients) that described an analysis for RFS survival found no significant relationship with HER2. Two studies (628 patients) found a significant association using a univariate analysis, as did one of these and one other study (totalling 608 patients) using a multivariate analyses. Ten studies of a total of 3763 node-positive patients were included. Seven of these (2777 patients) showed a significant relationship with RFS by univariate analysis, six studies (2625 patients) showed this in multivariate analyses and three studies (1018 patients) showed no significant association.

The results for overall survival in the Révillion review⁸⁷ were similar to those for recurrence-free survival, but more studies assessed this outcome. There were 16 studies (4694 patients) of node-negative patients. Five studies (1519 patients) showed a significant association using univariate analyses, which was confirmed by multivariate analysis in one of these studies. The other studies did not find a significant association, apart from some instances, within specific subgroups of patients. There were 13 studies of node-positive patients, including a total of 4797 women. Ten of these studies (3811 patients) showed a significant association between HER2 and overall survival, with four (2329) confirming this in multivariate analyses and six (2083 patients) finding no significant association in this type of analyses.

Révillion and colleagues⁸⁷ conclude that the associations they describe between HER2 and prognosis are “somewhat controversial”. They point out that, for node-positive patients, the status of HER2 as an independent prognostic factor was seen in some, but not all, studies, and that the studies with longer follow-up were less likely to find an association. For node-negative patients, their conclusion points out that “most studies did not find ERBB2 to be a prognostic indicator”.

In the Mirza review,⁸³ all 13 studies reported a univariate analysis of the relationship between HER2 and disease-free survival and/or overall survival. Eight of these showed a significant

association and five did not. Six studies reported a multivariate analysis: HER2 was significant for disease-free survival in two but not significant for both disease-free and overall survival in the other four. On the basis of their findings, Mirza and colleagues⁸³ conclude that there is only “limited association” between HER2 and prognosis.

The widely different number of studies in these reviews may be a reflection of the different eligibility or assessment criteria adopted or of the extent of searching involved. However, because these features are so poorly reported in each review, it is difficult to assess which factors are most important. In addition, comparing the sizes and the references for the studies in the different reviews reveals that there appear to be some studies included in one or a few of the reviews but not in all of them.

The suggestion of a likely association between HER2 and poorer prognosis in women with node-positive breast cancer appears to be justified based on the weight of evidence in these reviews. However, since all the reviews relied on some form of vote counting rather than an attempt to pool the results of the studies using appropriate weighting, the size and strength of this association warrant further investigation in a formal meta-analysis. There is less evidence for an association among node-negative patients and this, too, would benefit from a meta-analysis of the large number of studies that have investigated this issue over the last 15 years.

Prognostic factor: C-MYC amplification

Two reviews were identified of studies of the prognostic nature of C-MYC amplification. One of these⁹¹ was a standalone review of C-MYC amplification in women with breast cancer. The other⁸⁴ investigated C-MYC amplification within a broader review of biological markers in general. The Porter-Jordan review⁸⁴ does not contain details relating to search strategy or quality assessment and mentions only four relevant studies, all of which are included in the review by Deming and colleagues.⁹¹ In addition, there are several other studies in the latter, which were published before the Porter-Jordan review.⁸⁴ The Deming review⁹¹ is more comprehensive, up-to-date and of higher quality than the Porter-Jordan review⁸⁴ and, therefore, the remainder of this section concentrates on the Deming review.⁹¹

Scope of the review

The Deming review⁹¹ identified 29 relevant reports, relating to 26 studies. There were a total

of 3797 women in these studies, with the largest containing 1052 and the smallest, 30. The proportion of C-MYC amplified tumours ranged from 1 to 50%, with a pooled average of 15.7% [95% confidence interval (CI): 12.5 to 18.8%], but with substantial heterogeneity among the values from each study. Many of the studies did not report whether or not women with metastatic disease had been excluded, so it is possible that this review includes some data from such women. The outcome measures considered by the reviewers were disease-free and overall survival.

Quality of the review

The Deming review⁹¹ contains a detailed description of their literature search, including the keywords used. The reviewers searched MEDLINE, Current Contents and PubMed, but the period covered by the search, and whether any language restrictions were applied, are not reported. The references in retrieved articles and reviews were used as a means of identifying additional studies. The eligibility criteria for the review are stated and each study was assessed for key aspects of design, including sources of tissue and amplification controls, specification of the threshold used and a description of the study population. Where this was possible from the included studies, the reviewers investigated possible associations between C-MYC amplification and seven other prognostic factors: nodal status, ER status, PR status, age, menopausal status, tumour size and tumour grade.

The data used in the review were derived from the published reports only. No information is given on whether the authors of the original studies were contacted for missing data or clarifications. The relationship between C-MYC amplification and disease-free and/or overall survival was examined through a meta-analysis of the HR from each study. The relevant HRs were obtained directly from the original reports (three studies), from an analysis of raw data within these reports (one study), or were derived by calculating the smallest HRs that could be detected with 80% power at the *p*-values reported by the original studies (two studies). The review included tests for heterogeneity as part of the pooling of results in the meta-analyses, and also the use of a funnel plot (published in the article) to assess the possibility of publication bias.

Conclusions of the review and the strength of the conclusions

The review calculated a combined HR for disease-free survival of 2.05 (95% CI: 1.51 to 2.78) and of

1.74 (95% CI: 1.27 to 2.39) from the six studies that provided sufficient data for these calculations. The majority of the data in these meta-analyses come from the large study by Berns and colleagues⁹² (1052 patients), but the estimates from the other studies are consistent with this and there was no significant statistical heterogeneity among the studies. The review concludes that C-MYC amplification may have significant value as an independent prognostic factor for survival of women with breast cancer. It also concludes that there were significant, but weak, associations with tumour grade, node-positive disease, PR-negative status and postmenopausal status.

This review by Deming and colleagues⁹¹ was well conducted and is well reported. The searching was extensive and is well described. The reviewers' reliance on published studies, and upon data from within these publications, leaves the review open to publication bias. They investigated this explicitly using a funnel plot of study size versus prevalence of C-MYC amplified tumours, and concluded that a significant publication bias is unlikely. However, they do not appear to have assessed whether reporting of the association between C-MYC amplification and the disease-free or overall survival might be subject to publication (or selective reporting) bias. It is possible that the inclusion of these outcomes in some reports, but not others, might be related to the size or significance of any association found in the original studies. Hence, the reviewers' conclusion that more rigorous studies, with consistent methodology, are needed to verify the association is justified. However, based on the evidence in their review, it does appear that the presence of C-MYC amplification is an indicator for poor prognosis.

Class of prognostic factor: cell proliferation markers

There are many other growth factors and receptors involved in breast cancer, so a final common pathway, proliferation, is often measured. This can be assessed by the number of cells dividing (the mitotic index) or the expression of key molecules involved in regulating the cell cycle. S-phase fraction is a measure of the percentage of cells in the tumour that are making new DNA. Ki-67 is a nuclear protein that is up-regulated during the cell cycle and provides another marker for the number of proliferating cells. It can be scored on tumour histology sections.

Prognostic factors: cell proliferation markers

Five reviews were identified of studies of the prognostic nature of cell proliferation markers. These included one that was a standalone review of S-phase in women with breast cancer⁹³ and one that examined four proliferation indices:⁹⁴ thymidine-labelling index (TLI-BrdULI), S-phase fraction, mitotic count and Ki-67-MB-1. A third review⁸³ investigated S-phase fraction, mitotic index and Ki-67 within a review of prognostic factors in node-negative breast cancer. The Porter-Jordan review⁸⁴ of biological markers in general also contained a section on Ki-67 and the Sunderland review⁹⁵ contained a section on thymidine-labelling index, but both of these are superseded by the more recent reviews and are not discussed further here.

Scope of the reviews

The Daidone review⁹⁴ sought to include case series that investigated the possible association between any proliferation indices and clinical outcomes – principally, overall survival and disease-, event- or relapse-free survival. Patients who were node negative and node positive were eligible for the review.⁹⁴

The Wenger review⁹³ focused on studies of S-phase factor. The clinical outcome measures in these studies were overall survival, disease-free survival and response to treatment. Like the Daidone review,⁹⁴ it was restricted to studies with at least 100 patients. However, the Wenger review⁹³ contains more studies of S-phase fraction than the Daidone review⁹⁴ (even though it was conducted earlier). Limitations on our resources mean that it was not possible to identify the extent to which this is due to the requirement of Daidone and Silvestrini⁹⁴ that eligible studies had to have had at least 4 years of follow-up.

The Mirza review⁸³ only included studies in women with node-negative breast cancer. It was restricted to studies with at least 200 women and 5 years of follow-up. The review includes five studies (2369 women) of S-phase fraction, four studies (1350 women) of mitotic index and five studies (1960 women) of Ki-67. Studies were identified for the review through a MEDLINE search for articles published during 1996–2000. However, studies published before then were also included although no information is given on how these were selected. There were three such studies in the review of S-phase fraction and two in the Ki-67 review. The reviews focused on the relationship between these cell proliferation

markers and disease-free survival and overall survival.

Quality of the reviews

As noted above, the Wenger review⁹³ was restricted to studies with at least 100 patients. The review was based on a MEDLINE search for articles published between 1987 and 1997 using 'breast cancer' and 'S-phase fraction' as the keywords. Seventy-eight studies remained following the elimination of reports that were not in English, that did not report new data or dealt solely with methodology. Thirty-six studies used frozen tissue, 41 used paraffin-embedded tissue and one did not specify the type of tissue used. The data used in the review were derived from published reports only. No information is given on whether the authors of the original studies were contacted for missing data or clarifications. The reviewers identified heterogeneity among the studies and explained this in terms of both case mix and methodology. They also briefly discussed the possibility of publication bias and concluded that it was "probable" that their review was affected by this, although they wrote that "it should be noted that several of the articles did include negative results with respect to correlations with other prognostic factors and/or clinical outcomes".

The Daidone review⁹⁴ was based on a PubMed search for articles published up to July 2000 using 'breast cancer' and the name of each of the proliferation indices as keywords. Studies were restricted to those published in English with at least 100 patients and a minimum of 4 years of follow-up. A total of 120 papers were identified and these were divided into the different indices with the results presented for each study by giving the *p*-value from a multivariate analysis of disease-free survival or overall survival, but no information is given on how the eligibility criteria were applied. The data used in the review appear to have been derived from the published reports only. No information is given on whether the authors of the original studies were contacted for missing data or clarifications. There is no discussion of heterogeneity or the possibility of publication bias.

Although the eligibility criteria for the Mirza review⁸³ were stated, no details were given of how these criteria were applied. The review relied on the published results of the studies identified, and no additional information appears to have been sought from the original researchers. There is no discussion of the quality of the studies that were brought together in the review and the possible

impact of publication bias is not discussed. Similarly, there is no direct discussion of heterogeneity but details are given of the use of systemic therapy in each included study.

Conclusions of the reviews and the strength of the conclusions

The Wenger review⁹³ found 20 studies that investigated the relationship between S-phase fraction and overall survival using univariate analysis, of which 18 found that high S-phase fraction was associated with decreased overall survival (the other two found no significant relationship). Nineteen studies used multivariate analyses and S-phase fraction remained a significant prognostic factor in 14 and was not significant in the other five. The reviewers did not combine the results of the studies in a meta-analysis. They conclude that S-phase fraction "does have clinical utility for patients with breast cancer" and that high S-phase fraction is associated with poor prognosis.

The Daidone review⁹⁴ found eight studies (3364 women, half of whom were in one study) which investigated the relationship between thymidine-labelling index (TLI-BrdU) and survival using multivariate analysis. Five of these studies found that high thymidine-labelling index was significantly associated with worse disease-free or overall survival. There were 22 trials of S-phase fraction (4763 women) in the Daidone review,⁹⁴ of which 15 found that this was associated with statistically significant decrease in either disease-free or overall survival. The review found nine studies (1865 women) of mitotic count. Five of these found this was a significant prognostic factor in multivariate analyses. Finally, Daidone and Silvestrini⁹⁴ found four studies (1158 women) that investigated the relationship between Ki-67-MB-1 and disease-free or overall survival using multivariate analysis. Two of these found that high Ki-67-MB-1 was associated with decreased disease free survival. In summary, Daidone and Silvestrini⁹⁴ did not combine the studies they identified in a meta-analysis, but conclude that the four proliferation indices they investigated are each associated with poor prognosis.

Three of the five studies in the Mirza review⁸³ of S-phase fraction conducted a univariate analysis of the relationship with disease-free or overall survival. It was a significant positive prognostic factor in two studies, but negative in the other. Four studies used a multivariate analysis. S-phase fraction was a prognostic factor for overall survival in three of these studies and was not a prognostic

factor for disease-free survival in the fourth. On the basis of these findings, Mirza and colleagues⁸³ conclude that S-phase fraction is a “useful” prognostic factor.

Three of the four studies in the Mirza review⁸³ of mitotic index conducted a univariate analysis of the relationship with disease-free or overall survival, and all three found it to be a significant prognostic factor. The fourth study was the only one to use a multivariate analysis and it also found that mitotic index was a significant positive prognostic factor. On the basis of these findings, Mirza and colleagues⁸³ conclude that mitotic index is a “useful” prognostic factor.

Four of the five Ki-67 studies in the Mirza review⁸³ conducted a univariate analysis of the relationship with disease-free or overall survival. It was a significant prognostic factor in three of these studies. All five studies carried out a multivariate analysis, and Ki-67 was significant for overall and/or disease-free survival in four. On the basis of these findings, Mirza and colleagues⁸³ conclude that Ki-67 is a “useful” prognostic factor.

The conclusions of these reviews of S-phase fraction (that high values are associated with poor prognosis) appear robust. The conclusions of the Mirza⁸³ and of the Daidone⁹⁴ reviews on mitotic index and Ki-67, and of the latter on thymidine-labelling index, are also supported by the evidence but seem less robust, based as they are on vote counting, a lack of clarity about the methods of the reviews and the possibility of publication bias.

Class of prognostic factor: genetic instability and checkpoints

Some of the genes that transform normal breast cells into cancer cells (oncogenes) can induce genetic instability by BRCA1, ataxia telangiectasia and BRCA2. In addition, as cells proliferate, genetic damage occurs and uncontrolled proliferation without proper ‘check-points’ on duplicating DNA results in losses and gains of chromosomes. A downstream final end-point of many of these changes is whether there is the normal number of chromosomes and amount of DNA per cell, or abnormal amounts. This is assessable by measuring ‘ploidy’ of tumours (being aneuploid is having an abnormal amount of DNA). It may be expected that those tumours with the most abnormality can generate variants with more aggressive behaviour.

The cell cycle ‘check-points’ include a nuclear protein p53 that stops cells dividing if the DNA is damaged or in response to other stresses. One of the proteins that does this is p21, itself up-regulated by p53. Thus high expression may indicate the ability of the cancer cells to respond to stress.

Prognostic factor: aneuploidy or DNA ploidy

Two reviews were identified of studies of the prognostic nature of DNA ploidy or aneuploidy: Sunderland⁹⁵ and Mirza and colleagues.⁸³ Both were in the context of broader reports on prognostic factors. The older review⁹⁵ is within a general discussion of a variety of prognostic factors but remains relevant because it covers a period before that assessed by the second review. The more recent review⁸³ is a review of several prognostic factors in node-negative breast cancer.

Scope of the reviews

The Sunderland review⁹⁵ contains information on nine studies (2865 patients; range: 71 to 690), including women with node-negative and node-positive breast cancer, from studies done before the 1990s. The Mirza review⁸³ was restricted to women with node-negative disease and, in the main, to studies published between 1996 and 2000. Thus the content of the two reviews appears mutually exclusive, and none of the studies reported in the former seem to have been included (based on more recent follow-up) in the latter. The Mirza review⁸³ includes four studies (1230 patients), with the largest containing 421 women and the smallest 212.

No details of the searching are provided for the Sunderland review.⁹⁵ The searching for the Mirza review⁸³ involved a search of MEDLINE for articles published during 1996–2000, and the search terms used are reported in the review. Some papers published before this time were also included but no information is given on how these were selected, and two of the four studies in the review were published in 1992 and 1994, respectively. The Mirza review was restricted to studies with at least 200 women and more than 5 years of follow-up. Both reviews considered the possible effect of aneuploidy on disease-free and overall survival.

Quality of the reviews

The eligibility criteria for the review were clearly stated in the Mirza review,⁸³ but details were not given of how these criteria were applied. Sunderland⁹⁵ does not report the eligibility

criteria that were used. Both reviews appear to have relied exclusively on the published results of the studies they identified. Neither discussed the quality of the studies that were brought together or the possible impact of publication bias on the findings of the review.

Conclusions of the reviews and the strength of the conclusions

The Sunderland review⁹⁵ takes a study-by-study approach to describing the association, if any, found between aneuploidy and patient outcome. They conclude that “patients with aneuploid tumours are more likely to have short survival time than patients with diploid tumours”. On the other hand, Mirza and colleagues⁸³ report that DNA ploidy was a significant prognostic factor for overall survival and disease-free survival in two studies, but not in the other two, using univariate analyses, and that it was non-significant in all four when multivariate analyses were done.

The apparent inconsistency in the findings of these two, mutually exclusive, reviews and the methodological shortcomings of the Sunderland review⁹⁵ mean that there is insufficient evidence to draw any reliable conclusions about the relationship between DNA ploidy or aneuploidy and patient outcome.

Prognostic factor: p53

Five reviews of studies of the prognostic nature of p53 were identified. Three of these focused entirely, or almost entirely, on p53⁹⁶⁻⁹⁸ and one investigated p53 within a review of prognostic factors in node-negative breast cancer.⁸³ The Porter-Jordan review⁸⁴ of biological markers in general also contained a section on p53 but it is superseded by the more recent reviews and is not discussed further here.

Scope of the reviews

The four reviews all examined the association between p53 and disease- or recurrence-free survival and overall survival, but appear to have approached the problem from slightly different standpoints, which might be reflected in the poor overlap of the studies included in each of them. Each review appears to contain studies that were relevant to the other reviews, but which were not included in the others. Perhaps the most striking of these is a study by Silvestrini and colleagues,⁹⁹ which was published in 1993 and includes 1400 node-negative patients. This is in the Mirza review⁸³ but not in the other three. The Mirza review⁸³ was restricted to women with node-negative breast cancer, but the other three reviews

included both node-negative and node-positive patients.

The stated aim of the Barbareschi review⁹⁶ was to “analyze most of the published studies on the prognostic value of p53 immunohistochemical over-expression in breast carcinomas, trying to compare and weigh their results”. However, it is unclear how he selected which studies to include or exclude. The review includes 37 studies of 9860 patients, of whom about 60% were node negative.

Elledge and Allred⁹⁷ reviewed 57 studies, with a total of approximately 13,000 patients, which assessed the association of prognosis with inactivation of the p53 gene.

The stated aim of the Pharoah review⁹⁸ was to “identify all the published studies which have investigated the association between somatic mutations in the p53 gene and breast cancer prognosis”. Unlike the other reviews of p53, it includes meta-analyses combining the results of the included studies. The reviewers sought studies that reported survival analysis for women with breast cancer who had been tested for the presence of somatic mutations in the p53 gene. They identified 16 eligible studies (2993 patients). Twelve of these included unselected breast cancer patients, three were restricted to node-negative patients only and one small study (24 patients) was of women with inflammatory breast cancer. Overall, 18% of the patients had alterations to the p53 gene.

The Mirza review⁸³ was restricted to studies of at least 200 women who were node negative, with more than 5 years of follow-up. It included 16 articles (7586 patients), with a median follow-up of 98 months. One of these articles was Pharoah and colleagues,⁹⁸ meta-analysis of 736 node-negative patients. This is included in the Mirza review⁸³ even though the two studies in that meta-analysis which involved more than 200 patients^{100,101} are also included in their own right, contributing a total of 622 patients.

Quality of the reviews

Neither the search strategy, eligibility criteria nor the use of any quality assessment for the included studies is reported for the Barbareschi⁹⁶ or the Elledge review.⁹⁷

The Pharoah review⁹⁸ was based on searches of the MEDLINE and BIDS databases for articles published between 1983 and 1998 using ‘breast neoplasms’, ‘p53’ and ‘mutation’ as the search terms. The references in identified studies were

checked for possible studies. The reviewers used the most recent report, with the most complete dataset, for any studies that had been reported more than once. The eligibility criteria for the review were not clearly reported and no information is given on how the eligibility and quality of the identified studies were assessed.

The study identification for the Mirza review⁸³ involved a search of MEDLINE for articles published during 1996–2000, and the search terms used are reported in the review. Some papers published before this period were also included but no information is given on how these six studies were selected.

All four reviews relied on the published results of the studies identified and do not appear to have contacted the authors for additional information. The Pharoah review⁹⁸ used the reported relative hazards from multivariate analyses where available. They used reported 95% CIs for these or, if not reported, estimated them from the reported *p*-value and the HR estimate. In the two studies that reported no significant association between survival and p53 mutations, but included a survival curve but no relative hazard, the reviewers used a relative hazard of 1.0 in their meta-analyses.

Conclusions of the reviews and the strength of the conclusions

The Barbareschi review⁹⁶ divides the reports for the 37 included studies into three categories: “positive” studies that show an independent prognostic value for p53 expression in multivariate analyses; “borderline positive” studies that show a significant association in univariate analyses; and studies that “do not show any prognostic value for p53 expression”. The first category included 12 reports (4510 patients) with the individual study’s relative risks of dying of breast cancer ranging from 1.3 to 3.2; the second category included 11 reports (2331 patients) and the last, 14 reports (3021 patients). Barbareschi comments on important differences in the length of follow-up in the studies in these three categories and in the mix of women who were node negative and node positive. The studies that showed an association were more likely to have shorter follow-up and to contain a higher proportion of node-positive patients. Barbareschi concludes that “the prognostic ... value of p53 over-expression is probably weaker than was hoped ... [but] ... deserves further investigation”. The type of investigation he supports would be very large studies, with follow-up of more than 10 years.

The Elledge review⁹⁷ also divides the included studies on the basis of their results. They separately tabulate 40 studies (approximately 10,000 women) as showing that inactivation of p53 was associated with a worse outcome; and 17 studies (3300 patients) “which failed to demonstrate a significantly worse overall or disease-free survival”. They note that some of the latter group “showed trends towards worse outcome” and speculate that the lack of significance may be due to a lack of statistical power because of the smaller number of patients in some of these studies. Elledge and Allred⁹⁷ write, “overall, one can conclude from the aggregate of these studies that inactivation of p53 is associated with a worse prognosis and increases the relative risk of relapse by at least 50%”.

As noted above, the Pharoah review⁹⁸ conducted a meta-analysis, rather than rely on the approach of categorising studies on the basis of the statistical significance of their results. Eleven of the studies in the review (2319 patients) investigated overall survival. The relative hazards in these studies range from 1 to 23.4, with a combined estimate of 2.0 (95% CI: 1.7 to 2.5), but with significant heterogeneity among the results of the studies (*p* = 0.01). Pharoah and colleagues⁹⁸ were able to perform separate meta-analyses for node-negative and node-positive patients, where the relevant data had been reported in the included studies. They obtained a relative hazard of 1.7 (95% CI: 1.2 to 2.3) from a meta-analysis of 736 women who were node negative and 2.6 (95% CI: 1.7 to 3.9) for 550 women who were node positive. However, they note that there is no significant difference between these two subgroup-derived estimates. They conclude that, “in general, mutations in p53 confer a worse overall survival and disease-free survival in breast cancer cases, and this effect is independent of other risk factors”.

The Mirza review⁸³ reports that eight of the studies reporting univariate analysis showed p53 to be a significant prognostic factor and that six found it to be not significant. Twelve of the 16 studies did a multivariate analysis: six found that p53 is a significant prognostic factor and six did not. On the basis of these findings, p53 is one of the prognostic factors for which Mirza and colleagues⁸³ conclude that there are “mixed results”.

All four reviews discuss, in varying levels of detail, the heterogeneity among the studies they include. This relates to case mix, the immunohistochemical

techniques used and the duration of follow-up. Within the context of their meta-analyses, Pharoah and colleagues⁹⁸ were also able to demonstrate significant statistical heterogeneity among the results of the studies in their overall meta-analysis of survival.

All four reviews are possibly subject to publication bias through a reliance on published studies and published data only. The Pharoah review⁹⁸ is the only one to discuss the possible effect of publication bias on their findings. The reviewers explore this using a funnel plot (published in their report), and speculate that although some small studies may be missing from their review, the effect of such a bias would not be sufficient to overwhelm their findings. They show that if they removed the small, positive studies from their meta-analyses the estimate for the relative hazard would remain statistically significant, changing from 2.0 (95% CI: 1.7 to 2.5) to 1.8 (95% CI: 1.4 to 2.3).

These four reviews approach the association between the p53 gene and prognosis in different ways and reach slightly different conclusions. However, the strength of the evidence does support an association between alterations to the p53 gene and poor prognosis.

Prognostic factor: p21

Two reviews were identified of studies of the prognostic nature of p21. One of these¹⁰² was part of the presentation of the results for a series of patients. The other⁹⁷ considered p21 along with a more detailed review of p53.

Scope of the reviews

The two reviews provide little information on how they were conducted. The Domagala review¹⁰² identified nine other studies of p21 expression, in addition to the one conducted by the authors. The 10 studies included a total of 1317 patients with breast cancer. However, not all of these are mentioned in discussion of the association between p21 and survival. Elledge and Allred⁹⁷ reviewed six studies (1457 patients), including one with more than 800 patients that does not appear in the Domagala review.¹⁰² They provide information on disease-free survival and overall survival for each of these studies.

Quality of the reviews

Neither the search strategy, eligibility criteria nor the use of any quality assessment for the included studies is reported for either of these reviews. Both rely on the published results of the studies identified but the Domagala review¹⁰² did, of

course, have access to the raw data from their own study.

Conclusions of the reviews and the strength of the conclusions

Both reviews report that the associations found between p21 and prognosis are inconsistent among the studies. Some studies have shown that p21 is significantly associated with poorer prognosis, some that it is significantly associated with better prognosis and others that there was no significant association. Given the lack of information on the methods of the reviews, it is not possible to assess their quality, and the findings of both must be treated with caution. However, given the lack of any clear pattern, it is clear that more research would be needed to confirm or refute any association between p21 and prognosis.

Class of prognostic factor: apoptosis

As a breast cancer tumour expands, it can start to outgrow its blood supply, become hypoxic and have insufficient metabolites for growth – leading to areas of cell death. Although there is an increase in cell division, it is often so abnormal that the cell dies. Both of these processes contribute to a high death rate of cancer cells. The mechanism of cell death requires specific biochemical pathways leading to apoptosis – programmed cell death. Certain genes can protect cells against apoptosis, such as bcl-2, and therefore cancers which express these genes may have a better chance of survival and growth. Bcl-2 is a protein that stabilises mitochondria against release of toxic proteins and metabolites that activate apoptosis.

As discussed above, p53 is also important in producing apoptosis in response to stress, thus ensuring that damaged cells do not carry on to produce abnormal daughter cells. Mutations blocking this effect are common in most cancers, allowing cancer cells to have a survival advantage. Restoring normal p53 effect could therefore selectively kill cancer cells and is currently being researched by several companies.

Prognostic factor: bcl-2

We identified one review of studies of the prognostic nature of bcl-2.¹⁰³

Scope of the review

The review contains information on 11 studies, including a total of 3615 patients. Women with

node-negative and node-positive breast cancer were included. All but one of these studies assessed women with non-metastatic breast cancer. The largest study included 979 women and the smallest included 81 women. The review includes details of the number of patients with high and low bcl-2 in each study and, where available, the association with disease-free survival and/or overall survival. These data appear to have been extracted solely from the published reports of the included studies with, in some cases, the relevant statistics being estimated by the reviewers, from a published survival curve.

Quality of the review

The review does not contain information on eligibility criteria or quality assessment of the included studies. No details of the searching are provided and there is no discussion of the possible impact of publication bias on the findings.

Conclusions of the review and the strength of the conclusions

The reviewers report that “patients with high bcl-2 immunostaining have a better clinical outcome than those with low/negative bcl-2 expression” and conclude that this is likely to be explained by the relationship between bcl-2 protein and differentiation. It is difficult to judge the quality of this review in the absence of clearer information on how the reviewers conducted their search, appraised the reports they found and determined what should be included in their review. Hence, it is not possible to determine if their conclusion of a definite relationship between prognosis and bcl-2 is justified by the studies that have been done, as opposed to the studies that they included.

Class of prognostic factor: metastases

For tumour cells to spread from the primary tumour to distant sites (metastasis), destruction of the extracellular matrix around the cells is necessary. This is also necessary to allow new blood vessels to supply the tumour and for the circulating tumour cells to invade distant organs. Initially they need to invade the local blood vessels (vascular invasion) or lymphatics (lymphatic invasion), often scored as lymphovascular invasion. There are many pathways involved in normal degradation and turnover of the extracellular matrix, including multiple proteases and heparinases. Cathepsin D was one of the earliest to be studied because it was found to be oestrogen regulated, but many others are important,

including urokinase and its receptor, which binds the enzyme to the surface of invading cells, several metalloproteases and stromelysins. These are being studied as therapy targets although results so far with metalloproteases, the most advanced in trials, have been disappointing. Another pathway regulating metastasis is the enzyme nm23. This suppresses metastasis and loss of its expression is associated with disease spread, so it is a tumour suppressor gene.

As for cell division, assessment of the final overall effect is a way to include all the different pathways and, for metastasis, the detection of small clusters of cancer cells in the bone marrow or the circulation by highly sensitive immunochemistry and molecular techniques can detect minimal residual disease. These methods may be important in the future to measure the effects of adjuvant therapy within a couple of years of diagnosis and treatment, rather than needing to wait for the completion of 5 years of follow-up.

Although there are some associations between factors in general, they represent different pathways and the variable expression of these contributes to the extensive heterogeneity of cancer, hence the difficulty of assessing individual prognosis and individual optimum therapy.

Prognostic factor: cathepsin D

Three reviews were identified of studies of the prognostic nature of cathepsin D. One of these¹⁰⁴ was a standalone review of cathepsin D in women who had node-negative breast cancer. The others investigated cathepsin D within broader reviews of prognostic factors in node-negative breast cancer⁸³ and biological markers in general.⁸⁴

Scope of the reviews

Two of the reviews were restricted to studies in women with node-negative breast cancer^{83,104} and, although the third⁸⁴ (and oldest) was not restricted in this way, it included very little information on node-positive patients. Details of the searching and eligibility criteria are given for two reviews^{83,104} but not for the third.⁸⁴ There is fairly good overlap of studies in the three reviews but, even taking account of the fact that they were done at different times, there still appear to be some studies missing from each review.

The Ferrandina review¹⁰⁴ used the most comprehensive searching, with searches of several bibliographic databases and the explicit checking of the references in relevant papers. The electronic search covered articles published

between 1985 and 1996. It included 11 studies, on a total of 2690 women. One of these studies¹⁰⁵ is listed in two parts in the review, separating pre- and postmenopausal women.

The Mirza review⁸³ was restricted to studies with at least 200 women. The MEDLINE search for the review covered 1996–2000 but studies that were published before this period were also included and no information is given on how these were selected. Because of the restriction to studies that included at least 200 women, several of the studies found for the Ferrandina review¹⁰⁴ were not relevant, including one with 199 women.¹⁰⁶ The Mirza review⁸³ included four reports on the prognostic relevance of cathepsin D. However, one of these was the Ferrandina review¹⁰⁴ and two were studies that had been included in that review.^{107,108} This is noted in the Mirza paper⁸³ but its potential impact on the findings of the review does not seem to have been taken account of and nor is any reason given for why two other studies in the Ferrandina review¹⁰⁴ which included more than 200 women were not included. The fourth report in the Mirza review⁸³ was by Foekens and colleagues¹⁰⁹ and included 1412 women with node-negative breast cancer. It was published after the Ferrandina review¹⁰⁴ and is much larger than any of the studies in the Ferrandina review,¹⁰⁴ which contained a total of 2690 patients.

The Porter-Jordan review⁸⁴ is the oldest cathepsin D review identified for this project and many of the studies identified by the other two reviews had not been published at the time it was conducted. The review contains five studies, which include the above-mentioned study by Thorpe and colleagues¹⁰⁵ – counted once only in this review. All but one¹¹⁰ of the five studies was in the Ferrandina review.¹⁰⁴ This study included fewer than 200 women, so it was not eligible for the Mirza review,⁸³ but it is unclear why it was not included by Ferrandina and colleagues.¹⁰⁴ On the other hand, the Ferrandina review¹⁰⁴ included several studies that were published before the Porter-Jordan review.⁸⁴ However, because the search strategy (in particular the period covered by the searching) is not reported in the Porter-Jordan review,⁸⁴ it is not possible to determine if these studies were missed by their searching, were identified and judged to be ineligible or were not included for other reasons.

Quality of the reviews

The eligibility criteria for the included studies were clearly stated in both the Ferrandina¹⁰⁴ and the Mirza⁸³ reviews, but details were not given of

how these criteria were applied. All three reviews relied mainly on the published results of the studies they identified, but the Ferrandina review¹⁰⁴ involved a recalculation of the survival analyses for each study through the extraction of information from published survival curves. Where this was not possible, data were sought from the original authors and this was supplied and used for three of the 11 included studies.

None of the three reviews discussed the quality of the studies that were brought together and only one¹⁰⁴ discussed the possible impact of publication bias on the review's findings.

Conclusions of the reviews and the strength of the conclusions

All three reviews concluded that cathepsin D is a prognostic factor for women with node-negative breast cancer, such that a high expression of cathepsin D is related to poorer prognosis. The consideration of women with node-positive disease was restricted to the Porter-Jordan review.⁸⁴ A brief mention was made in this review of a non-statistically significant relationship between cathepsin D and prognosis in such women, but this is based on a citation to a single study.

Although there are potential minor flaws in the reviews identified for this project, the consistency of the findings across the reviews and the quantity of evidence identified justify the conclusion that high expression of cathepsin D is related to poorer prognosis in women with node-negative breast cancer. However, there is insufficient evidence to draw any reliable conclusions about the relationship in women with node-positive disease.

Prognostic factor: urokinase-type plasminogen activator

Two reviews^{111,112} and one pooled analysis¹¹³ were identified of studies of the prognostic nature of urokinase or its receptors. The urokinase-type plasminogen system comprises at least four proteins, all of which have been studied in reviews. These are the uPA, its membrane-bound receptor (uPAR) and two inhibitors (PAI-1 and PAI-2). The studies identified for this project examined these to varying extents.

The two reviews were standalone reviews of uPA, uPAR, PAI-1 and PAI-2¹¹¹ and PAI-1.¹¹² The pooled analyses, done by the European Organisation for Research and Treatment of Cancer – Receptor and Biomarker Group (EORTC-RBG), used a specially compiled dataset to examine uPA and PAI-1.¹¹³

Scope of the reviews

The two reviews were focused on women with non-metastatic breast cancer but do not appear to have been restricted in any other way. The pooled analysis only used data from studies by members of the EORTC-RBG in Europe. Details of the searching and eligibility criteria are not given for the two reviews, but both seem to rely exclusively on published studies. Look and Foekens¹¹¹ noted that they used the most detailed report for each included study. This was often the most recent report with the longest follow-up and most events. The pooled analysis included data from both published and unpublished studies. Comparison of the references in the three reports reveals some overlap of studies between the three, but this overlap is not complete and each report contains some studies that are not in the other two (and this is not due solely to studies that were published after the target reports themselves).

The Look review¹¹¹ examined the relationship between RFS and uPA, uPAR, PAI-1 and PAI-2. It included seven studies (2699 women), two studies (639), five studies (1859) and three studies (1474), respectively. High levels of uPA and PAI-1 were consistently found to be related to poor RFS but the smaller number of published studies available for uPAR and PAI-2 revealed a less clear relationship. The review includes some examination of the importance of uPA and PAI-1 in association with other prognostic factors, including ER status, menopausal status and nodal status.

The Harbeck review¹¹² was restricted to studies of PAI-1. It includes 12 such studies, with a total of 6107 women, the largest being a study of 2780 women.¹¹⁴ It concludes that a high PAI-1 value is associated with poor prognosis in all the included studies.

The Look pooled analysis¹¹³ combines the datasets from 18 studies from nine European countries, conducted by members of the EORTC-RBG. Eleven of these studies had been published (with references given for a total of 12 citations), and seven studies were unpublished. There were a total of 8377 in the whole dataset, comprising 8175 women for the analysis of uPA and 6682 for PAI-1. The large study by Foekens and colleagues,¹¹⁴ involving 2780 women, is included. The dataset was analysed for RFS and overall survival, with consideration of subgroups based on age, menopausal status, nodal status, tumour grade and hormone receptors. The researchers conclude that, aside from nodal status, “high levels of uPA

and PAI-1 were the strongest predictors of both RFS and overall survival” in all patients combined and that this was especially true for node-negative patients.

Quality of the reviews

Neither the eligibility criteria nor the search strategies used were reported in the two reviews. For the pooled analysis, the reliance on European data was clearly stated but no details were given of whether any additional studies were potentially eligible. The two reviews relied on the published results of the studies they identified but the pooled analysis made full use of individual patient data, including those from unpublished research. The Look review¹¹¹ did include some discussion of the importance of heterogeneity within a review of this nature but, perhaps because of the consistency of results in the included studies, this does not appear to have been examined in much detail. None of the reviewers discussed the quality of the studies that they brought together in their review.

Conclusions of the reviews and the strength of the conclusions

All three reports concluded that PAI-1 is a prognostic factor for women with breast cancer, such that a high value is related to poorer prognosis. The consistency of this finding in the reviews and in the pooled analysis despite the different data used in each, justifies this conclusion. The reviews and the pooled analyses are also strongly supportive of the same relationship for uPA. There is insufficient data for uPAR and PAI-2 and this is recognised by the one review that assessed this.¹¹¹

Prognostic factor: nm23

One review was identified of studies of the prognostic nature of nm23. This was within a review of the prognostic relevance of biological markers in general.⁸⁴

Scope of the review

The review contains three studies, all published in the early 1990s, and one of which was of infiltrating ductal carcinoma. The total number of breast cancer patients in these studies is not clearly reported. It is unclear if any eligibility criteria were set or applied.

Quality of the review

The review contains no information on the search strategy, eligibility criteria or the use of any quality assessment for the included studies. It relies on the published results of the three studies identified.

Conclusions of the review and the strength of the conclusions

The review reported that the two studies of women with primary breast cancer showed that expression of nm23 was associated with longer disease-free survival. However, the reviewers state, “large cohorts to confirm the findings and permit subset analysis of patients have not been published”. This is a fair statement of the absence of reliable research on the possible prognostic nature of nm23. Given the lack of information on their methods, it is also not possible to assess the quality of the review itself or whether the few studies identified are a complete sample of the worldwide evidence as it existed in the early 1990s.

Prognostic factor: bone marrow micrometastases and minimal residual disease

One systematic review¹¹⁵ was identified of studies of the prognostic nature of bone marrow micrometastases in a range of cancers, with most studies being in breast cancer, and one review¹¹⁶ of minimal residual disease in breast cancer.

Scope of the reviews

The Funke review¹¹⁵ contains a total of 20 studies (2494 patients), of which 11 studies (1926 patients) are of breast cancer. The results and conclusions of the review are mostly presented for all cancers combined. Of the 11 breast cancer studies, one included women with metastatic breast cancer, but this study was relatively small¹¹⁷ (71 patients). The 11 breast cancer studies ranged in size from 25 to 727 patients, with a prevalence of bone marrow micrometastases from 2 to 48%.

The Diel¹¹⁶ review contains information on six studies, including a total of 1934 patients, with a restriction to studies of more than 100 patients. The review also mentions one smaller study by citing it as the first study of this question. Diel and Cote¹¹⁶ refer to the Funke review¹¹⁵ but criticise it because of the variety of cancers and detection techniques included. The largest study in the Diel review¹¹⁶ included 1026 women, and is a later report of the largest study in the Funke review¹¹⁵ (see below). Of the six reports in the Diel¹¹⁶ review, four are also in the Funke review.¹¹⁵ In addition, there is the aforementioned study by Diel and Cote,¹¹⁶ which is included in the Funke review¹¹⁵ but based on an earlier publication in English,¹¹⁸ with a smaller number of patients. There is also another study that was not reported in English and contained 228 patients.¹¹⁹ The restriction of the Funke review¹¹⁵ to studies published in English is discussed below.

The outcomes considered in both reviews are disease-free survival and overall survival, and the data used appear to have been extracted solely from the published reports of the included studies.

Quality of the reviews

No details of the searching or quality assessment are provided for the Diel review.¹¹⁶ The Funke review¹¹⁵ contains good details on its literature search. MEDLINE and Current Contents were searched from 1980 to 1997 for original articles. Letters to the Editor, abstracts, reviews, book chapters and articles without an English abstract were excluded, as were studies with fewer than 20 patients or with insufficient data to calculate a relative risk for relapse-free or overall survival. The references in retrieved articles were checked for possible studies, but did not reveal any that had not been found by the electronic search. Efforts were made to ensure that the same patient series were not used more than once. Each retrieved report was summarised independently by the two reviewers, with discrepancies resolved by consensus. The review included tests for heterogeneity as part of the pooling of results in meta-analyses.

Conclusions of the reviews and the strength of the conclusions

The Funke review¹¹⁵ calculated a combined relative risk of 1.34 (95% CI: 1.27 to 1.42) for RFS for breast cancer patients with bone marrow micrometastases compared with those without, but there was significant heterogeneity among the studies ($p = 0.001$). For overall survival, the review reports a significant impact of bone marrow micrometastases in four of eight studies, but does not present a combined analysis. The reviewers conclude that “the presence of epithelial cells in bone marrow has to be validated as an independent factor for poor prognosis in cancer patients by further studies with standardised procedures before its official acceptance in the TNM classification”. The Diel¹¹⁶ review concludes that “the question of whether or not tumour cell detection is an independent prognostic factor has not been resolved”.

The Funke review¹¹⁵ was well conducted and is well reported. The searching was extensive and well described. However, its reliance on English language original articles leaves it open to publication bias through the exclusion of research that has been published in other languages or the grey literature, or has not been published. Although the Diel review¹¹⁶ contains insufficient information to assess its quality, its inclusion of

two articles published in German does reveal that a reliance on English language articles may miss important literature. The conclusion of both reviews that more evidence is needed to be certain about the possible prognostic nature of bone marrow micrometastases or minimal residual disease is justified. However, based on the evidence in these reviews, it does appear that the presence of micrometastases is an indicator for poor prognosis.

Class of prognostic factor: standard pathology

Other prognostic factors studied involve standard pathology, which always needs to be considered, because of the robust, extensive data and their routine use in management. These factors include tumour size and grade, lymphovascular invasion, involvement of surgical margins and spread to regional lymph glands. These are the core factors onto which information from new factors can be grafted to see if further refinement is possible.

Because larger or fatter people may have differences in the amount of hormones they make or growth factors, body habitus has also been studied for prognosis and might be especially important because it can be modified by dietary interventions.

Prognostic factor: tumour size

One article¹²⁰ was identified that discussed the importance of various prognostic factors (including nodal status and tumour grade) in women with tumours smaller than 1 cm. However, this is not examined further here because the review's methods (including eligibility criteria, study identification and assessment) are inadequately reported. Rather, one review was identified of studies of the prognostic nature of tumour size. This was within a broader review of prognostic factors in node-negative breast cancer.⁸³

Scope of the review

The Mirza review⁸³ only included studies in women with node-negative breast cancer. It was restricted to studies with at least 200 women and 5 years of follow-up. The review includes nine studies (17,883 women) of the association of tumour size with prognosis. The majority of the patients were in one very large study of 13,464 women.¹²¹ Studies were identified for the review through a MEDLINE search for articles published

during 1996–2000. However, studies published before then were also included although no information is given on how these were selected. There were four such studies in this review, including the very large one just noted. The review focused on the relationship between tumour size and disease-free survival and overall survival.

Quality of the review

Although the eligibility criteria for the review were stated briefly, no details were given of how these criteria were applied. The review relied on the published results of the studies identified, and no additional information appears to have been sought from the original researchers.

There is no discussion of the quality of the studies that were brought together and the possible impact of publication bias is not discussed. Similarly, there is no direct discussion of heterogeneity but mention is made of the use of systemic therapy in each study (although it was not reported in five of the nine studies).

Conclusions of the review and the strength of the conclusions

All nine included studies found a statistically significant association between tumour size and either disease-free survival, overall survival or both. Six of the nine studies reported a univariate analysis of the relationship between tumour size and disease-free survival or overall survival. All six studies showed a significant relationship. Four studies used a multivariate analysis. Tumour size was a significant prognostic factor for overall survival in all these.

On the basis of these findings, Mirza and colleagues⁸³ conclude that there is good evidence that there is an association between tumour size and prognosis. This conclusion, that increasing tumour size is associated with poorer prognosis in women with node-negative breast cancer, is justified.

Prognostic factor: tumour grade

One review was identified of studies of the prognostic nature of tumour grade, within a general review of prognostic factors in node-negative breast cancer.⁸³

Scope of the review

The review only included studies in women with node-negative breast cancer. It was restricted to studies with at least 200 women and 5 years of follow-up. The review includes six studies, with a

total of 3442 women. The largest of these included 1157 women. Studies were identified for the review through a MEDLINE search for articles published during 1996–2000. However, studies published before then were also included although no information is given on how these were selected. There were three such studies in this review, including the largest.

Quality of the review

Although the eligibility criteria for the review were stated, no details were given of how these criteria were applied. The review relied on the published results of the studies identified, and no additional information appears to have been sought from the original researchers.

There is no discussion of the quality of the studies that were brought together and the possible impact of publication bias is not discussed. Similarly, there is no direct discussion of heterogeneity but the potential problem of each study using a different grading system is mentioned.

Conclusions of the review and the strength of the conclusions

Five of the six studies conducted a univariate analysis of the relationship between grade and disease-free or overall survival. All five showed a significant relationship. Three studies used a multivariate analysis and Mirza and colleagues⁸³ highlight the importance of systemic therapy as a variable in such analyses. In the study in which none of the patients received systematic therapy, there was a significant association between grade and overall survival. However, in the two studies in which some patients received systemic therapy, the study that included treatment in the multivariate analysis concluded that grade was a significant prognostic factor but the study that did not include treatment in its model showed no significant association.

Mirza and colleagues⁸³ conclude that there is an association between grade and prognosis for women with node-negative breast cancer. In the light of the fact that the association was seen in almost all the analyses in the studies in the review, this conclusion appears justified. However, the possibility of publication bias and the lack of detail on how studies were assessed as eligible for the review mean that it should still be treated with some caution. In addition, the review provides no information on the association between tumour grade and prognosis in women with node-positive breast cancer.

Prognostic factor: vascular invasion

One review was identified of studies of the prognostic nature of tumour size. This was within a broader review of prognostic factors in node-negative breast cancer.⁸³

Scope of the review

The Mirza review⁸³ only included studies in women with node-negative breast cancer. It was restricted to studies with at least 200 women and 5 years of follow-up. The review includes five studies (3150 women) of the association of vascular invasion with disease-free survival and overall survival, with the largest including 1203 women. Studies were sought for the review through a MEDLINE search for articles published during 1996–2000. However, studies published before then were also included although no information is given on how these were selected and, in this review of vascular invasion, all five of the included studies were published before this period.

Quality of the review

Although the eligibility criteria for the review were stated briefly, no details were given of how these criteria were applied. The review relied on the published results of the studies identified, and no additional information appears to have been sought from the original researchers.

There is no discussion of the quality of the studies that were brought together and the possible impact of publication bias is not discussed. Similarly, there is no direct discussion of heterogeneity but mention is made of the use of systemic therapy in each study.

Conclusions of the review and the strength of the conclusions

Two of the five studies reported a univariate analysis of the relationship between vascular invasion and disease-free survival and overall survival. Both studies showed a significant relationship. One of these two studies and the other three in the review reported a multivariate analysis. Vascular invasion was a statistically significant prognostic factor for overall survival in all these. However, in one of the four, the relative risk for disease-free survival was below the threshold of 1.5 used by Mirza and colleagues⁸³ to judge something as 'positive' for an association (although the relevant *p*-value was 0.002).

Mirza and colleagues⁸³ conclude that there is good evidence that there is an association between vascular invasion and prognosis. This conclusion,

that the presence of vascular invasion is associated with poorer prognosis in women with node-negative breast cancer, appears justified, although the lack of information on how any of the five included studies were identified or judged to be eligible means that some degree of caution is warranted.

Prognostic factor: body size

One review was identified of studies of the prognostic nature of body size.¹²²

Scope of the review

The review identified 14 studies (5525 women), 13 of which were cohort studies and one, with only 25 patients, used a case-control design. The cohort studies ranged in size from 68 to 962 patients. The outcome measures considered by the reviewers were recurrence and survival. None of the studies set out to study body size *per se*, but the information had been collected as part of routine patient histories or within the context of a controlled trial.

Quality of the review

The review contains some information on its literature search but does not provide details of the keywords or index terms used. MEDLINE was searched from 1975 to 1989 for reports published in English, and the reviewers mention that additional articles were also sought from the references in retrieved articles and “additional MEDLINE searches when necessary”. Studies reported in abstracts or letters were excluded because their methodological quality could not be assessed. The reviewers used an explicit methodological quality assessment for each included study and this is reproduced in the review. Both reviewers independently assessed each of the retrieved reports, with discrepancies resolved by consensus. The maximum possible score for a study was 14. The score range for the 13 cohort studies was 4–12. The score for the case-control study was 3. The reviewers do not appear to have contacted the original researchers for any additional information or clarification.

Conclusions of the review and the strength of the conclusions

In general, studies that were judged to be of higher quality were more likely to conclude that there was an association between a body size and prognosis. Six studies found body size to be predictive of outcome and seven found it to be associated with disease recurrence. The effect was consistently indicative of a large body size being associated with a poor prognosis. Five studies reported on the size of the association and, among these, the estimates of relative risk for recurrence

and death in the individual studies were typically 1.5–2.0. The conclusion of the reviewers is that “it appears that body size exerts a modest effect on prognosis in breast cancer that persists after adjustment for the effects of other prognostic factors”. However, they do note that their conclusions “require confirmation in well designed properly conduct prospective studies”.

Although certain aspects of this review are well done, most notably the detailed assessment of the quality of the reports of the studies, the conclusion is based on vote counting and the reliance on English language original articles leaves the review open to publication bias. Hence, the reviewers’ estimate of a moderate association between body size and prognosis may be an overestimate, and their caution in calling for better quality research to resolve this issue is justified.

Other prognostic factors

There are other possible prognostic factors in breast cancer that were not investigated in this project, for example, angiogenesis, which is the process of development of a new blood supply from pre-existing vessels, which is essential for the growth and spread of cancer. Many oncogenes also switch on growth factors for blood vessels (angiogenic factors), for example EGFR, HER2, ER. Amongst the most potent and specific angiogenic factors is a vascular endothelial growth factor (VEGF). This is also a target for therapy with oral inhibitors of its receptor on endothelial cells and antibodies to the growth factor or its receptor. There are many other angiogenic factors, so assessing a final downstream pathway and the number of blood vessels in the tumour (microvessel density) has been extensively evaluated. This is done by staining paraffin sections with antibodies to blood vessels (for example anti-CD31). No search was made for reviews of studies of angiogenesis, other than for those factors discussed above.

Discussion

The advances in cell biology and molecular biology applied to cancer have led to a major increase in understanding of the process of malignancy. Much of this work is initially applied to experimental models including human cell lines grown in tissue culture, animal tumour models and genetic studies in lower organisms and in mice. These studies provide proof of principle that a particular biochemical pathway can control tumour growth or

spread, but do not show that this applies to a specific type of human cancer (there are over 100 different ones), or that it is relevant to a particular clinical problem. That can only be proven eventually in clinical trials designed to modify that pathway and in correlative randomised studies. In the latter, the pathway can be studied in tumour biopsies and, depending on its relationship to response to treatment or survival (or both), be shown to be predictive or prognostic. This provides evidence for a role and for a use of the marker but to show that it is the mechanism requires specific modification of the pathway, for example of ER by tamoxifen.

This project sought systematic reviews of studies of the relationship between many suggested prognostic factors and relapse or death for women diagnosed with breast cancer. The overall quality of the reviews identified was poor, in comparison with what would be regarded as high-quality systematic reviews of, for example, healthcare interventions. The deficiencies relate to both conduct and reporting. Inadequate information was provided for many of the reviews on key factors such as eligibility, study identification and quality assessment. Where these were reported, the reviews were often revealed to have been restricted to published studies, published data and, almost always, these were limited to the English language.

Only a small number of the reviews sought data or further information from the original authors and only Look and colleagues¹¹³ used individual patient data, an approach which is recognised as the benchmark for systematic reviews.¹²³ Similarly, very few of the reviews included a meta-analysis. Most of the reviews relied on vote counting, which was almost always based on statistical significance of the association between a factor and an outcome in each study in the review. This is problematic because of the large number of small studies that might be non-significant because of the lack of power to show that a true relationship is statistically significant. In addition, some of the small studies may have produced false-positive results, which, because of publication bias, entered the public domain whereas similar sized, but non-significant, studies remained unknown. This may vary unpredictably among the prognostic factors and could lead to false positives and false negatives in the reviews.

Vote counting is also unreliable because it gives equal weight to a study of 20 and one of 2000 patients. Some reviews were restricted to studies above a certain size and with a minimum period of follow-up (e.g. Mirza and colleagues⁸³), which

might help to overcome problems associated with biased reporting of small studies. Few of the reviews considered heterogeneity or publication bias. Sensitivity analyses were rare and only two reviews^{91,98} included a funnel plot to investigate publication bias.

However, even with these failings in mind, some conclusions can be drawn about which factors do and do not have good evidence for their prognostic nature in women with breast cancer. In summary, the factors for which the identified systematic reviews provide robust evidence that the factor is prognostic are the following:

- S-phase fraction
- p53
- cathepsin-D (for node-negative women)
- PAI-1
- tumour size (for node-negative women).

The evidence on the following appears supportive of the prognostic nature of the factor:

- HER2 (CerbB2) (for node-positive women)
- C-MYC amplification
- thymidine-labelling index (TLI-BrdULI)
- mitotic count
- Ki-67
- uPA
- bone marrow micrometastases
- minimal residual disease
- tumour grade (for node-negative women)
- vascular invasion (for node-negative women).

The evidence from systematic reviews that was identified for the following factors is insufficient to conclude whether or not they are prognostic:

- ER status
- ps2
- EGFR
- HER2 (CerbB2) (for node-negative women)
- aneuploidy or DNA ploidy
- p21
- bcl-2
- cathepsin-D (for node-positive women)
- uPAR
- PAI-2
- nm23
- tumour size (for node-positive women)
- tumour grade (for node-positive women)
- vascular invasion (for node-positive women)
- body size.

Table 6 summarises the strength of evidence for each of the prognostic factors examined in this chapter.

TABLE 6 Summary of findings from a systematic review of reviews of prognostic factors in breast cancer

Prognostic factor	Finding
Cathepsin-D	++ (for node-negative), +/- (for node-positive)
Epidermal growth factor	+/-
HER2/CerbB2	+/- (for node-negative), + (for node-positive)
Urokinase and its receptors	
PAI-1	++
uPA	+
uPAR	+/-
PAI-2	+/-
p53	++
p21	+/-
bcl-2	+/-
C-MYC amplification	+
Proliferation indices	
S-phase fraction	++
Thymidine-labelling index	+
Mitotic count	+
Ki-67-MB-1	+
ps2	+/-
nm23	+/-
Aneuploidy or DNA ploidy	+/-
Tumour grade	+ (for node-negative), +/- (for node-positive)
Tumour size	++ (for node-negative), +/- (for node-positive)
ER status	+/-
Vascular invasion	+ (for node-negative), +/- (for node-positive)
Minimal residual disease	+
Bone marrow micrometastases	+
Body size	+/-

–, No relationship between factor and survival; +/-, insufficient evidence to identify a relationship between factor and survival; +, evidence of a relationship between factor and survival; ++, clear evidence of a relationship between factor and survival.

Chapter 5

Prognostic models in breast cancer

Introduction

Statistical models for predicting patient outcome are termed prognostic models. They are widely used in cancer for investigating patient outcome in relation to multiple patient and disease characteristics. The focus of this chapter is the review of published models that have been developed to predict the outcome of future breast cancer patients. There are two broad ways in which such a model may be useful. First, it may allow the (reasonably) reliable classification of patients into two or more groups with different prognoses. Such classification schemes can be used to influence therapy and perhaps save patients from unnecessary referrals or tests. Second, a prognostic model can be used to estimate the prognosis of individual patients. Specific reasons for wishing to predict a patient's outcome include:

1. to inform treatment or other clinical decisions for individual patients
2. to inform patients and their families
3. to create clinical risk groups for informing treatment or for stratifying patients by disease severity in clinical trials.¹²⁴

With regard to treatment, it may be of particular interest to identify patients with such a good prognosis that adjuvant therapy would not be (cost-)beneficial, or a group with such poor prognosis that more aggressive adjuvant therapy would be justified.¹²⁵

A prognostic model was defined as a combination of at least two separate variables to predict patient outcome. A focus was made on those papers where the specific aim was either to develop a new prognostic model or to attempt to validate an existing model. Studies were excluded where the primary focus was to investigate the prognostic or predictive importance of one or more specific markers that included their evaluation in the context of a multivariate model. Such studies are considered in Chapters 4 and 6. In practice, it is not easy to distinguish the two types of study, and we have included a few studies with aims that were somewhat ambiguous. For example, the study of Cooke and colleagues¹²⁶ was included, which

examined the impact of adding HER2 to the Nottingham Prognostic Index (NPI), and the study of Chan and colleagues,¹²⁷ in which the monoclonal antibody BRE-3 was explored in the context of an overall prognostic model. For inclusion, a study had to consider at least one of the end-points of death, cancer death or recurrence of disease in newly diagnosed patients, and these were the only outcomes that we considered.

Although one can derive a multiple regression model from any data set, it is clear that a prognostic model will have no clinical value unless it has been shown to predict outcome with some success. As Burstein¹²⁸ noted, "Any classification system, be it nominal, ordinal, or scalar, should be proved to be a workable tool before it is used in a discriminatory or predictive manner". Another way of expressing this idea is that unless the model is shown to be useful it will be quickly forgotten.¹²⁹ Prognostic models were therefore sought that reliably distinguish clinically important variations in prognosis among groups of women with newly diagnosed breast cancer. The aim of identifying reliable information about prognosis meant that the main interest was in identifying prognostic models for which there has been an evaluation of how successfully the models have been when used in a different setting, that is, models which have been validated externally.^{124,125,130}

Developing reliable prognostic models

"The prediction of whether or not a woman will relapse within a given time of the primary tumour being removed is a very hard problem"¹³¹

The difficulties of developing reliable multiple regression models have been much discussed.^{40,59,132,133} Both clinical and statistical aspects are critical. Previous work has perhaps focused on methods of analysis rather too much and neglected the importance of study design and data quality. *Table 7* shows the main challenges in conducting prognostic analysis, suggested by Concato.¹³⁴

TABLE 7 Challenges in prognostic analysis¹³⁴

Clinical issues	Statistical issues
Criteria to identify 'candidate' subjects	Suitable coding of variables
Choice of zero time	Appropriate analytic techniques
Exclusion of patients	Desired format of results
Data describing patients at zero time	Measures of accomplishment
Evaluating the impact of interventions	
Defining the outcome state	

The American Joint Committee on Cancer proposed detailed criteria for evaluating putative prognostic factors:¹³⁵

“They must be (1) significant, (2) independent, and (3) clinically important. Furthermore, we suggest the criteria for selecting a prognostic system that includes TNM and new prognostic factors. These criteria are: (1) easy for physicians to use; (2) provides predictions for all types of cancer; (3) provides the most accurate relapse and survival predictions at diagnosis and for every year lived for each patient; (4) provides group survival curves, where the grouping can be by any variable including outcome and therapy; (5) accommodates missing data and censored patients and is tolerant of noisy and biased data; (6) makes no *a priori* assumptions regarding the type of data, the distribution of the variables, or the relationships among the variables; (7) can test putative prognostic factors for significance, independence, and clinical importance; (8) accommodates treatment information in the evaluation of prognostic factors; (9) accommodates new putative prognostic factors without changing the model; (10) accommodates emerging diagnostic techniques; (11) provides information regarding the importance of each predictive variable; and (12) is automatic.”

Although this list presents an unachievable goal, it does indicate well the scope of the issues that need to be considered when developing and using prognostic models. Some of the clinical and methodological issues are considered in the following sections.

Clinical issues

As noted, prognostic models can be of great clinical assistance. However, few prognostic models are in common use (not just in cancer).¹³⁶ As Wyatt and Altman¹²⁹ observed: “However accurate a model is in statistical terms, doctors will be reluctant to use it to inform their patient management decisions unless they believe in the model and its predictions.”

They suggested the following prerequisites for clinical credibility:

1. All clinically relevant patient data should have been tested for inclusion in the model.
2. It should be simple for doctors to obtain all the patient data required, reliably and without expending undue resources, in time to generate the prediction and guide decisions. Data should be obtainable with high reliability, particularly in those patients for which the model's prediction are most likely to be needed.
3. Model builders should try to avoid arbitrary thresholds for continuous variables.
4. The model's structure should be apparent and its predictions should make sense to the doctors who will rely on them, as only then will the law treat users of the model as 'learned intermediaries' in a case of alleged negligence.
5. It should be simple for doctors to calculate the model's prediction for a patient.

The first suggestion is rarely discussed, but it seems that in practice models are developed in retrospective studies using those variables that happen to have been collected already for other reasons. Hence many studies omit potentially valuable variables as they do not have data for them.

A further aspect, only implicit in the above quotation, is that the patients whose data are used to develop a model are precisely those for whom such a model would be used in clinical practice. It follows that it is essential for patients' characteristics and the sampling method to be described in reports.

Note that point 4 implies that the statistical modelling method must be correctly applied, and also suggests that 'black box' models such as artificial neural networks (ANNs) are less suitable for clinical applications.

It may well be that the failure of models to be taken into clinical practice reflects the (justified) view that they have not been adequately demonstrated to be useful, that is, a “healthy scepticism of undocumented new technology”.¹²⁹

Study design

Aspects of the design of studies of specific prognostic markers were considered in Chapter 3 (Table 2). Most issues apply to studies designed to develop a prognostic model, such as the advantages of a prospective study over a retrospective study, the need for patients to be followed up from a common event (such as surgery or diagnosis) and the quality of measurement.

Several authors have addressed the issue of sample size for prognostic studies.^{52,137} It is important to recognise that the power of a study depends on the number of observed events, not the number of patients. Thus a small sample with long follow-up may well yield better information than a large study with short follow-up. It follows that a study will have less power to investigate rarer end-points; in the present context, that means that there is more power to investigate recurrence than death.

For studies which aim to develop a prognostic model, the sample size needs to be large enough to override the problems of multiple comparisons in the selection of variables and the comparison of models. Harrell and colleagues¹³⁸ suggested that the number of events should be at least 10 times the number of potential prognostic variables investigated, a value supported by a simulation study.¹³⁹ Feinstein¹⁴⁰ suggested that a minimum event prognostic value (EPV) of 20 is safer and Schumacher and colleagues¹³³ suggested 10–25. Most studies fail to achieve an EPV of even 10, which is likely to be a major source of unreliability in their findings, especially when they have used some stepwise algorithm for selecting the model (as most researchers do).

Other factors may suggest a larger sample size. For example, if continuous markers are dichotomised, as is common, the effective sample size is reduced by 30% or more, so that considerably more patients would be needed to achieve the same statistical power. Any investigation of interactions between prognostic factors and consideration of multiple cut-points will further increase the sample size required. Altman and Lyman⁵⁹ suggested that such studies should be based on at least 250–500 events. Although a large sample size can improve

precision, it cannot compensate for other weaknesses of a study.

One reason for preferring prospective studies is that data sets are likely to be much more complete than in retrospective studies using clinical databases collected for other purposes. Incomplete data are a common and often serious problem for studies developing prognostic models. Thus, although the sample size may be large, patients missing one or more variables will generally need to be excluded from a modelling exercise. (Recent developments in imputation of missing data have as yet found little uptake in this field – see, for example, Clark and Altman.¹⁴¹) The obvious effect will be to reduce power, but a much more serious possibility is the risk of introducing bias, notably the use of banked tumour material, which is likely to be unavailable after some time in subjects with small initial tumour size⁴¹ and may be associated with other prognostic factors.¹⁴² Such selection bias cannot be discerned by readers unless published articles report on the selection of individuals for inclusion in a study and a comparison of the characteristics of those with and without available tumour material.¹⁴²

The alternative to excluding patients with missing data is to impute the missing values. Various strategies are available,^{143,144} but they make some fairly strong and unverifiable assumptions about the reasons for the incompleteness in the data. However, such imputation may be preferable to using only complete cases, which can be a small proportion of the whole sample.

The completeness of the data should clearly be reported, both by variable and overall.¹⁴⁵ Some authors include data completeness as an inclusion criterion, making it impossible for the reader to know how representative the sample was. This practice is not recommended.

Statistical analysis

Continuous variables

A particularly important aspect of modelling is the handling of continuous variables. The choice is primarily between keeping such variables continuous, usually leading to the specification of a linear relation between the variable and lnHR, or creating categories and thus largely avoiding the problem of model specification. There are considerable advantages in keeping variables continuous.^{146,147} However, categorisation is extremely common in oncology – indeed, splitting into two groups (dichotomisation) may be considered the norm. The reasons for this

widespread practice are unknown. It seems that there is inadequate understanding of the implications of reducing the data in this way.

Categorising patients into high- and low-risk groups based on a marker threshold or cut-point effectively assumes a constant risk up to the threshold and then a different constant risk for all values beyond the threshold. Such dichotomisation is artificial and often unnecessary. Further, it discards potentially important quantitative information, thus reducing the power to detect a real association with survival.⁶⁵ A continuous, if not linear, relation between the value of marker and prognosis is in most cases far more plausible *a priori* than a jump in risk at some (unknown) value of the marker.

Further, the best method for selecting an appropriate marker cut-point is unclear. Patients are often divided into two equal groups by splitting at the median value. Although this approach will be unbiased, there is no *a priori* reason to suppose that half of the patients are at higher risk (indeed, this is implausible). Some investigators compute the statistical significance level for all possible cut-points and then select the cut-point giving the smallest *p*-value. There are several serious problems associated with this so-called 'optimal cut-point' approach. The *p*-values, survival curves and regression coefficients resulting from these analyses are biased by preselection of the cut-point using the same data.^{66,68} Also, the actual value of the cut-point is not well estimated and has no clinical meaning. The actual Type I error rate for this procedure is close to 40% rather than the nominal 5%.⁶⁶ Although an 'optimum' cut-point may give the best discrimination within a sample, it is unlikely to do so in the entire population of similar patients, which should be the aim of the study. Finally, the bias associated with this method is carried across into subsequent multiple regression analyses.^{66,146}

Rather than just fit all the possible ('candidate') variables in a prognostic model, many studies seek parsimonious prediction models by retaining only the most important prognostic factors. The most common approach is some form of stepwise variable selection. The model may be constructed by identifying the most important prognostic variable and then including also the one which adds most to the first variable, and so on until no more variables add significantly to the predictive performance. Alternatively, a full model can be constructed including all variables and the least significant variable is removed, a process repeated

until all remaining variables are statistically significant. These methods are known as forward stepwise and backward elimination, respectively. There are some variants of these approaches. It is also possible, and often sensible, to force certain known important variables into the model and use selection for the remainder. Recognised prognostic factors should generally not be subjected to the selection process. If they are excluded because by chance they do not reach a specified level of significance in that particular study, the resulting model can be misleading. (Note that inclusion may be less likely if such variables such as tumour size and grade are reduced to just two categories.)

Models

The most common analytic technique in prognostic studies in cancer is usually known as Cox regression. The Cox proportional hazards model is a survival analysis regression model, which describes the relation between the event incidence, as expressed by the hazard function and a set of covariates.^{133,148} The Cox model is essentially a multiple linear regression of the logarithm of the hazard on a set of predictor variables (or covariates), with the baseline hazard being an 'intercept' term that varies with time. Under the model, the covariates act multiplicatively on the hazard at any point in time, and this provides us with the key assumption that the hazard of the event in any group is a constant multiple of the hazard in any other. It follows that the fitted hazard curves (and hence also survival curves) for different groups cannot cross. The exponentials of the regression coefficients are called HRs. A covariate with an HR greater than one (equivalent to a regression coefficient greater than zero) indicates that as the covariate increases, the event hazard increases and hence the length of survival decreases. This proportionality assumption is often appropriate for survival time data but ought to be verified for each data set.

Unfortunately, the results of stepwise regression analyses are likely to be misleading.^{40,59} The regression coefficients in the final selected model may be biased, being on average too large. This effect is the result of the inclusion and exclusion of variables based on their association with outcome. Significance tests associated with these inflated coefficients are not strictly valid and the *p*-values are too small. It is common practice to include all variables significant at an arbitrary level of significance of 0.05 in the final model. The selection of variables on this basis has no direct relationship to clinical importance. Also, the classification of certain variables as important (and

others as not important) misrepresents the fact that models based on very different sets of variables may predict equally well. All of these difficulties are exacerbated when there are few events per variable, as discussed below. The preceding remarks apply in particular to weak prognostic factors.

Nonetheless, stepwise methods are widely used; it is important that researchers should be aware of the statistical properties of stepwise selection procedures and not over-interpret their findings. In the current context, the wide use of such methods is another argument supporting the need to carry out a validation study before claiming that a model is useful. Failing that, the stability of a model can be explored using the bootstrap approach.¹⁴⁹

A similar form of model selection is the 'all subsets' approach, in which each combination of variables is examined. In essence, the best model with each number of variables is ascertained and there is a penalty for each additional variable included in the model. This approach is not often used.

Cox proportional hazards regression is the most widely used method for examining several prognostic variables simultaneously and can reasonably be considered to be the conventional approach for such studies. Some other approaches that are occasionally used in studies exploring multiple factors simultaneously^{133,137} are described briefly below.

Parametric proportional hazards models work in broadly the same way as Cox models but the underlying hazard function is estimated by means of assuming a particular distribution for the survival times (such as the exponential or Weibull).^{148,150,151} Rather different are accelerated failure time models, in which effects of covariates are modelled in terms of increased survival time rather than HRs. Parametric models are not in wide use in statistical practice in general and are very rarely used when developing prognostic models in oncology. As discussed below, however, they have particular attractions in this setting.

Classification trees (also called regression trees and recursive partitioning), using 'CART', 'RECPAM', or similar methods, work by selecting the variable which best splits the patients into high- and low-risk groups and simultaneously selects the best cutpoint.^{152,153} The same procedure is then carried out within each group thus formed, and so

on until a stopping rule comes into play. The resulting classification rule of variables is called a tree. Multiple testing is a highly relevant concern here also. Successive choices are made on more highly specified subgroups with diminishing sample size. The method in effect examines many interactions between variables; it is highly data dependent and may give an over-optimistic result. Classification trees are appealing in that they provide ready-made groups which may be mapped on to different treatment options. Although their success has been noted in specific cases, there is no good evidence that they tend to improve on Cox regression (or logistic regression). Classification trees have not gained wide acceptability, partly because they require specialist software, but they do appear occasionally in cancer journals. Erlichman and colleagues¹⁵⁴ suggested that the tree method is better able to handle missing data than conventional regression modelling.

A more recent approach, although no longer really a new technique, is the construction of an ANN.¹⁵⁵⁻¹⁵⁷ Neural networks automatically allow arbitrary non-linear relations between the independent and dependent variables, and all possible interactions between the dependent variables. Cox regression requires additional modelling to allow this flexibility, so that ANNs ought to have an advantage.

The development of an ANN is a complex, computer-intensive process, and so its main disadvantage is that it is a 'black box'. Not only is it difficult to understand what the model is doing, but also the model cannot be written down simply and hence is not at all straightforward to transfer to another centre. Also, the development of ANNs requires considerable computational power and specialist software.

Sargent¹⁵⁸ reviewed the literature comparing conventional modelling with neural networks. Although there are published articles that show improved prediction with an ANN, Sargent's review of 28 comparative studies¹⁵⁸ (not only in cancer) found no clear evidence of an overall benefit of ANNs. He concluded that more research is needed and that ANNs should not replace standard statistical approaches as the method of choice for the classification of medical data. Schwarzer and colleagues¹⁵⁹ reviewed applications of ANNs in oncology (1991-5) and found that serious methodological errors were common; they also concluded that there is as yet no evidence that "application of ANNs represents real progress in the field of diagnosis and prognosis in oncology".

They drew the same conclusion from a more recent literature survey of applications of ANNs in prostate cancer (1999–2001).¹⁶⁰

Creation of risk groups

Regardless of whether specific prognostic variables are kept as continuous or categorised when developing a model, the predicted outcomes will often need to be grouped in some way to apply that model (the exception is when the model contains just two or perhaps three binary variables). With a model containing just binary or categorical predictors, risk groups may be created by collapsing a multi-way categorisation. With one or more continuous variables it is usually necessary to calculate a prognostic index and divide the range of values into bands representing different levels of risk. (The same approach could also be applied when the model contains many categorical predictor variables.) Kaplan–Meier survival curves are a good way to illustrate the degree of prognostic separation achieved by the classification.

The prognostic index (PI) is the weighted combination of the variables in the model (X_i), with the regression coefficients (b_i) as weights. Thus $PI = \sum b_i X_i$. There is no consensus on how many groups should be created, or on how to choose the cut-points.¹⁴⁶ Equally spaced intervals will tend to give extreme groups with rather few patients, so some researchers choose groups with similar numbers of patients, at the expense of some reduction in the separation of the survival curves. Although the separation between groups will generally increase with more groups, the clinical use of such a classification needs to be kept in mind. Elston and Ellis¹⁶¹ observed that because prognosis needs to be considered in relation to the available treatment options, there was little value in having more than three risk groups for categorising breast cancer patients. Risk groups may be created for various reasons, however, and more may be advisable in some situations.

Validation of a model

The idea of validating a prognostic or diagnostic model is generally taken to mean establishing that it works satisfactorily for patients other than those from whose data the model was derived.¹²⁴ Although it is customary to refer to the process of applying an existing model to new data as ‘validation’, there is an important distinction between the process of evaluating a model and the outcome of that process. Strictly, a model would be described as validation only when that evaluation gave results deemed satisfactory in some sense.

It is to be expected that the performance of a model will be poorer in new data than in the data set on which the model was derived. One main reason is the data-dependent choices made when developing a prognostic model, most obviously the choice of a model that is in some sense ‘best’ among many alternative models. Another important reason would be differences in the patients’ settings in the two data sets. A validation exercise may be carried out on additional patients in the same centre(s) as the original data or elsewhere. The latter is a harder, and more valuable, test of a model, as it mimics the reality of taking a model and applying it in clinical practice in different settings. Regrettably, such studies are rare.

However, it does not matter if the performance of a model is less good in a different context if that performance remains clinically useful. Various measures have been proposed to quantify the performance of a model. Arguably the best indicator of usefulness is the separation between risk groups in a plot showing Kaplan–Meier survival curves, which can be quantified using the *D* statistic of Royston and Sauerbrei.¹⁶² Their method can be used regardless of whether groups are created or how many there are.

Quality assessment

The assessment of the quality of prognostic studies was considered in Chapter 3. It was noted there that there are no established criteria for assessing the quality of prognostic studies. Here additional issues are considered specific to the development of a prognostic model. Publications across all medical areas are drawn on, especially systematic reviews of prognostic models in other medical areas. No systematic search was carried out to identify such studies.

Quality is sometimes addressed by levels of evidence or quality scores. The American Society of Clinical Oncology (ASCO) Expert Panel¹⁶³ used five levels of evidence for assessing the information about prognostic markers:

- I. meta-analysis or large, high-powered concurrently controlled studies in which the primary objective of the trial design was to test the utility of the marker
- II. prospective clinical trials designed to test a therapeutic hypothesis in which tumour marker evaluation was a secondary, but prospectively described, objective

- III. retrospective studies of large size (>200 patients per subgroup) and/or inclusion of multivariate analysis
- IV. retrospective studies of small size and/or no multivariate analysis
- V. small, retrospective studies not designed to examine relation between marker results and clinical outcome.

A major problem with levels of evidence, however, is that they indicate the nature of the evidence, not its quality.¹⁶⁴ In addition, the studies we have reviewed are almost all of level III, with some at level II, so such a scheme does not discriminate.

Quality scores also are problematic, as they generally combine information that is important (in the sense of being associated with the risk of bias) with other information.^{72,165} In this review therefore no attempt has been made to produce a simple statement of the overall 'quality' of each study.

The specific quality aspects that relate to a study developing a prognostic model are

- study design (including sample size)
- sample selection criteria
- choice of method of modelling (e.g. Cox regression)
- selection of candidate variables
- number of events per variable
- missing data
- method of reducing the number of variables in the 'final' model
- assessing modelling assumptions
- whether the model is presented in such a way that it could be used by others
- whether any validation was undertaken.

All of these issues were considered when reviewing published studies. In addition, the completeness of reporting of key information in the publications was considered.

Searches

Papers were sought that presented new prognostic models for patients with operable breast cancer or which evaluated a previously published model (validation study), or both of these. A total of 4791 abstracts were initially identified for question A from across all databases: BIOSIS, 311; CancerLit, 3424; MEDLINE, 580; EMBASE, 127; Cochrane Library, 342; CCTR, 7.

Appendix 1 gives details of the question A search strategy. De-duplication was undertaken in Procite using the terms 'Author/Title/Date', 'Title/Date', 'Author/Date', 'Author' and 'Title'. Title screening involved the removal of papers that were non-breast cancer, non-adjuvant therapy and exclusively focused on advanced/metastatic disease, whereby non-breast cancer was the principal reason for the exclusion of these papers. Screening by SB reduced the number of abstracts to 387, of which 58 were deemed possibly relevant. The 222 records with no abstracts were not considered further, owing to the time constraints on the project and because it seemed unlikely that the papers sought would be published without an abstract.

In addition, 21 papers that had been identified as possibly relevant to prognostic models during the screening of the databases for the other questions within the project were added to the 58 question A references (all from the QB database). Four additional abstracts were identified through hand-searching of *Breast Cancer Research and Treatment*; three of these had been identified through the initial search strategy, but had been rejected at the abstract screening stage. During the hand-searching (undertaken in the summer of 2002), these abstracts were deemed to be relevant to the project question and were added to the set of abstracts to be screened by the other two reviewers for this research question. The same person who performed the initial screening of the abstracts (SJB) undertook the majority of the hand-searching.

Two other reviewers (DA, CW) screened the 85 abstracts deemed relevant to this topic. A total of 52 papers were retrieved for data extraction. Some additional articles were identified by DA, after the main searches as the review progressed, including some identified by inspecting the references of included studies. The final number of papers reviewed was 78.

It is likely that some studies have been missed because there is no reliable search strategy for the types of study being sought. Hand-searching of one journal yielded some studies that had been rejected on the basis of abstracts. Also, some studies were not found by searching but by serendipity, from reference lists of included studies and by citation searching of key articles.

Data extraction

Assessment of each identified paper was made using a pre-piloted extraction sheet (Appendix 2).

For studies reporting the development of a new prognostic model, the following information was extracted:

- study characteristics
- design
- sample characteristics
- treatment
- end-points for outcome analysis
- follow-up
- data quality
- prognostic factors used in developing the model
- univariate analysis
- multivariate analysis
- presentation of multivariate model
- treatment of continuous prognostic variables.

If relevant, additional information was extracted relating to the validation of the prognostic model or of a previously published model. Any additional comments were recorded at the end of the form.

In the next two sections, the methods used in these studies are considered first, including observations on the quality of reporting, and then their findings.

Published prognostic models – review of methodology

After assessment, 17 of the 78 articles were excluded as ineligible (*Table 8*). The findings that follow are thus based on analysis of data from 61 studies: 42 presented one or more new prognostic models and 19 included model validation, of

which seven presented solely a validation of an existing prognostic model. Even this distinction was not clear as many studies did not have clearly stated objectives.

Some studies were the subject of two or more publications, possibly with different lengths of follow-up, in which the authors did not always take a consistent approach. For example, Fisher and colleagues¹⁶⁶ presented prognostic models based on a 10-year follow-up of 1090 node-negative and 651 node-positive patients enrolled in NSABP Protocol B-06, but they omitted to mention that the study was a randomised trial or that they were analysing only one arm of that trial. The paper reporting a model based on a 15-year follow-up¹⁶⁷ clarified the design but included only 1039 node-negative patients. As a second example, in two papers based on the same series of patients, Chapman and colleagues¹⁶⁸ presented results of five different multivariate models based on the same series of patients (but different lengths of follow-up), but the sample sizes were 378 and 293, respectively.

Table 9 gives characteristics of the studies and the patients who were included and *Table 10* shows details of the statistical analysis and development of the prognostic model.

Study characteristics

Any study of patient survival can be defined by three dates – the dates of the start and end of recruitment of patients and the date of the end of follow-up. Occasionally follow-up may be terminated at the same fixed time point after recruitment for all individuals.

TABLE 8 Ineligible studies, with reasons for exclusion

Blamey, 1996 ¹⁶⁹	Not developing or validating a model
Burke <i>et al.</i> , 1995 ¹⁷⁰	Not developing or validating a model
Clark <i>et al.</i> , 1995 ¹⁷¹	Abstract
Cooke <i>et al.</i> , 2001 ¹²⁶	Review of importance of HER2
De Laurentiis <i>et al.</i> , 1993 ¹⁷²	Abstract
Denic, 1996 ¹⁷³	No clinical data
Erllichman <i>et al.</i> , 1990 ¹⁵⁴	Not all patients had operable cancer
Frkovic-Grazio and Bracko, 2002 ¹⁷⁴	Not developing or validating a model
Grumett and Snow, 2000 ¹⁷⁵	Editorial
Hasebe <i>et al.</i> , 2000 ¹⁷⁶	Focus on a single factor; no prognostic model
Knox <i>et al.</i> , 1993 ¹⁷⁷	Abstract
Marchevsky <i>et al.</i> , 1999 ¹⁷⁸	Not developing or validating a model
Pinder <i>et al.</i> , 1998 ¹⁷⁹	Not developing or validating a model
Quantin <i>et al.</i> , 1996 ¹⁸⁰	Not developing or validating a model
Rostgaard <i>et al.</i> , 2001 ¹⁸¹	Not developing or validating a model
Rudolph <i>et al.</i> , 1999 ¹⁸²	Focus on a single factor
Schwarzer <i>et al.</i> , 2000 ¹⁵⁹	Not developing or validating a model
Yamamoto <i>et al.</i> , 1998 ¹⁸³	Metastatic disease

TABLE 9 Characteristics of studies presenting and/or validating prognostic models for breast cancer patients (n = 61)

Study characteristics	n	%	Study characteristics	n	%
Focus of the paper			<i>Patients' end-points analysed</i>		
Development of a prognostic model	42	69	Death		
Evaluation of an existing model	7	11	Explicitly any death	13	21
Both	12	20	Not explicitly any death (includes 3 reporting 'overall survival')	22	35
Study dates			Specifically cancer death	16	26
Date of start of recruitment given ^a	46	75	Relapse (including recurrence, progression or disease-free survival) (includes 15 studies where it was unclear if death was taken as an event or censored, and some studies which looked separately at local and distant recurrence)	29	48
Date of end of recruitment given ^a	44	72	Any event (death or relapse)	2	3
Date of end of follow-up given ^a	14	23	<i>Time origin specified</i>	17	29
All 3 study dates specified (includes one study with fixed-term follow-up)	13	21	Diagnosis	4	
Design			Biopsy	1	
<i>Study design</i>			Surgery	8	
RCT	4	7	Randomisation	1	
Cohort study	54	89	First treatment	1	
Unclear	3	5	'Date of first observation'	1	
Prospective	10	16	Not stated	44	71
Retrospective	48	79	Follow-up		
Unclear	3	5	<i>Summary of length of follow-up</i>		
<i>Justification of sample size</i>	0	0	Median (or mean) only	19	31
Patients' characteristics			Median (or mean) and range	15	25
<i>Inclusion criteria</i>			Other	11	18
Not stated	21	34	Not stated	16	26
Stated explicitly	23	38	<i>Median (or mean) follow-up</i>		
Partly stated	17	28	<5 years	9	15
<i>Menopausal status</i>			5–9.9 years	22	36
Pre- and post-menopausal	39	64	10+ years	7	11
Post-menopausal only	1	2	Not stated	23	38
Not stated	21	34	<i>Clear statement on number of patients lost to follow-up^b</i>	14	23
<i>Adjuvant therapy (at least some patients)</i>			<i>Clear statement on how losses to follow-up treated was in analysis^b</i>	11	18
Chemotherapy	30	49	Data quality		
Hormone therapy	23	38	<i>Discussion in text of missing data?</i>	19	32
Radiotherapy	26	43	(includes 6 studies with no missing data)		
Immunotherapy	1	2	<i>Number of excluded patients due to missing data was reported (includes 6 studies with no missing data)</i>	29	48
End-points for outcome analysis					
<i>Number of end-points used in univariate survival analysis</i>					
0	15	25			
1	28	46			
2	13	21			
3+	5	8			
<i>Number of end-points used in multivariate survival analyses</i>					
0	2	3			
1	41	67			
2	15	25			
3+	3	5			

^a At least the year was specified.

^b Including two studies that stated that no patients were lost.

It is desirable for investigators to present all three dates,⁷⁴ thus fixing the study in time, indicating the recruitment rate and giving an idea of the range of follow-up. Curiously, it is rare for all three dates to be provided in published survival studies; the date defining the end of follow-up was given in less than one-quarter of these papers.

Design

Descriptions of key aspects of study design were very poor. Not one paper gave a justification for the sample size for the study. Although it is likely that in some of the randomised trials a formal power calculation was used, that would have related to the detection of a treatment effect

TABLE 10 Characteristics of studies developing prognostic models for breast cancer patients (n = 54 unless stated otherwise)

Study characteristics	n	%	Study characteristics	n	%
Number of candidate prognostic variables used in developing the model			<i>Any interaction(s) examined</i> (includes all studies using only ANN)	14/53	26
2–5	6	11	<i>Model assumptions discussed</i>	9/53	17
6–10	30	56	<i>Model assumptions assessed</i>	9/53	17
11–20	10	19	<i>Goodness of fit assessed</i>	8/53	15
21–30	3	6	<i>How continuous prognostic variables were treated in multivariate analysis</i>		
>30	3	6	No continuous variables	3/53	6
Not stated	2	4	All kept continuous	13/53	24
Univariate analysis			All categorised (of which 13 dichotomised all continuous variables and 5 used a data-dependent method)	24/53	44
<i>Univariate analyses reported</i>	45	83	Some categorised	10/53	19
<i>Results presented in a tabulated form</i>	28/45	62	Unclear	4/53	7
<i>Survival proportions presented per variable</i>			Presentation of multivariate model		
All variables	4/45	9	<i>Regression coefficients presented</i>	30/49	61
Some variables (e.g. only if $p < 0.05$)	7/45	16	<i>HRs (relative risks) presented</i> (including 2 only partially)	19/49	39
<i>HRs presented</i>	7/45	16	<i>SE or CI of regression coefficients or risk/hazard ratios presented</i> (1 partially)	24/49	49
<i>CI presented</i>	4/45	9	<i>p-Values from final model</i> (8 partially)	39/49	80
Multivariate analysis			<i>Calculation of prognostic index</i>	28/49	57
<i>Multivariate method used</i> (6 studies used 2 different methods)			<i>Risk groups created</i>	32	59
Cox proportional hazards (including one using a time-dependent model)	45	83	2 groups	6	11
Accelerated failure time	3	6	3 groups	11	21
Classification tree/recursive partitioning	1	2	4 groups	11	21
Logistic regression	2	4	5+ groups	3	6
Discriminant analysis	1	2	Not done	29	54
ANN	6	11	<i>Method used to create risk groups</i>		
None	1	2	Count factors present	5/32	16
<i>Choice of variables to include</i>			Data dependent	6/32	19
All available variables/all used in univariate analyses	34/53	64	Equal size (e.g. at quartiles)	3/32	9
All variables with $p < 0.05$ in univariate	10/53	19	Other non-data-dependent method	8/32	25
Other	5/53	9	Unclear	10/32	31
Unclear	4/53	8	<i>Graph presented showing expected survival for risk groups</i>	29	54
<i>Strategy for building the multivariate model</i>			<i>Prognostic model compared with other published models</i>	7	13
Forward stepwise selection	8/53	15	<i>Full model specified</i> (so it could be applied to new patients)		
Backward stepwise selection	10/53	19	Yes	13	24
Stepwise selection (unspecified)	7/53	13	No	32	59
All significant in univariate (no further selection)	5/53	9	Unclear	9	17
Other	6/53	11			
Unclear	12/53	23			
Not relevant	5/53	9			
<i>Forced inclusion in full model of variables known a priori to affect survival</i>	2/53	4			

^a At least the year was specified.

^b Including two studies that stated that no patients were lost.

rather than the development of a model. It was very hard to identify which studies were prospective, in the sense that there was a clear *a priori* intent to collect data with the aim of generating a prognostic model. Also, studies stated to be prospective may well not be, for example one with follow-up of 108 women for between 1 and 18 years.¹⁸⁴ A few data sets were

subsets of patients who had been enrolled in one or more randomised trials, which by definition are prospective studies.

Sample size, missing data and number of events

Ideally, we would like to know for each study the number of eligible patients, the number included

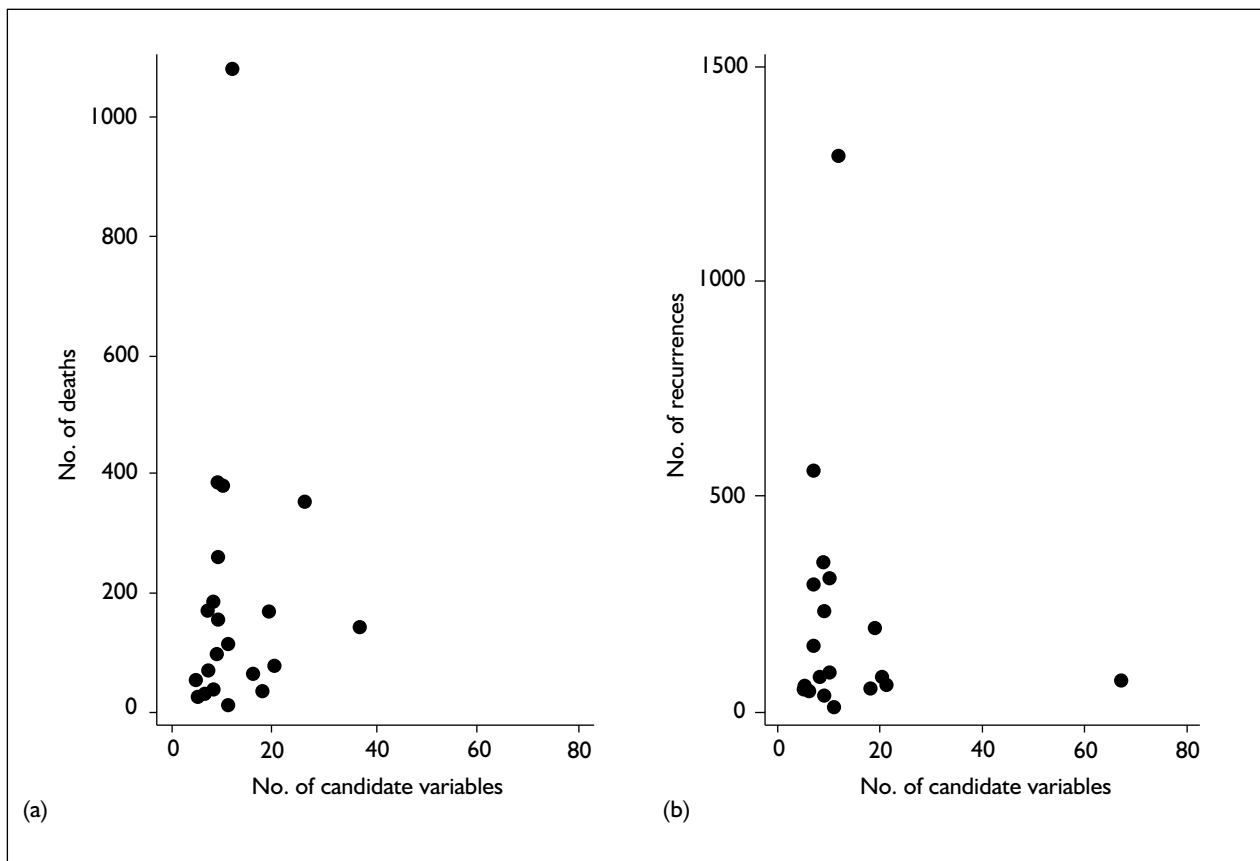


FIGURE 3 Relation between numbers of events and number of candidate variables for the two most common end-points: (a) death; (b) relapse

in univariate analyses and the number contributing data to the generation of the prognostic model. In practice, it was rarely possible to derive all of these numbers. The number of patients included in univariate analyses could often be deduced from tables of patients' characteristics, but the number included in the multivariate analysis to develop a model was often not stated explicitly. Thus the number of patients with complete data was often not stated, yet missing data can lead to a substantial reduction in overall sample size. As an example of good reporting, Chapman and colleagues¹⁶⁸ reported an initial sample size of 448 of whom 70 were ineligible (with reasons given), and they state explicitly that 323 patients had complete data and were included in the multivariate analysis. As an example of the extent of missing data in such retrospective studies, Erlichman and colleagues¹⁵⁴ showed the amount of missing data for 21 candidate variables considered in their modelling. Eleven variables were missing for more than 10% of patients and five variables for more than one-third of patients. The number of patients with complete data was not reported. By contrast,

Lundin and colleagues¹⁸⁵ reported that only 36 of 1050 patients had missing data in a series going back to 1945.

Similar comments apply to the numbers of events, deaths and/or recurrences. Hence the number of events per variable could only be calculated for about half of the studies. As shown in *Figure 3*, many studies had fewer than 10 events per candidate variable. An extreme case is the study of Kaufmann and colleagues,¹⁸⁶ who investigated 11 variables in a data set of 57 patients with only four events (EPV <0.5).

Sample characteristics

Patients' characteristics were somewhat determined by inclusion criteria but, as noted, these were poorly reported. Stage of breast cancer was rarely reported explicitly, but all studies included only operable cancer (or, in a few cases, did not specify). In 33 studies, all patients were reported as having surgery and in the study of Collan and colleagues¹⁸⁷ 1/120 did not have surgery. This information was missing for 23 studies. Menopausal status (and hence age

distribution) was often not indicated. Some studies gave no information about patients' characteristics (e.g. Burke and colleagues;¹⁸⁸ Lockwood and colleagues¹⁸⁹).

Sample size

None of the studies gave a justification for the sample size. Although in most studies there were some patients with missing data, many papers did not indicate clearly the sample size used to develop the final model. Further, only the minority of papers reported the number of events (e.g. deaths), without which it is not possible to judge the reliability of a study.

Treatment

Information about adjuvant therapy was not well reported. *Table 9* shows the numbers of studies for which some patients were reported to have received each type of adjuvant therapy. In only a very few cases was there explicit information that no patients had received a certain type of adjuvant therapy. In almost all cases the treatment received was not included among the candidate prognostic variables (see below).

End-points for outcome analysis

The majority of papers (56/61) analysed two outcomes in multivariate analyses, usually death and recurrence, each defined in various ways.

Most of the papers (51/61) presented models for survival but for only about one-quarter (13/51) was it unambiguous that the authors were including deaths from any cause. Several authors referred to 'overall survival', which usually means that the event of interest is death from any cause, but some authors use this term when talking of only deaths from breast cancer,¹⁸⁹ a confusion also noted by Altman and colleagues.⁷⁴ In 16 studies (26%), the outcome was explicitly only death from cancer. In such studies it was not generally noted how such judgements were made.

Similar ambiguity afflicted the reporting of analyses of time to recurrence. We therefore have not distinguished between terms such as time to recurrence, disease-free survival and relapse-free interval. A key aspect of such outcomes is whether deaths without recurrence are treated as events or censored (presumably on the grounds that the deaths were not due to cancer). A few papers presented results for two or more different recurrence outcomes, such as local and distant recurrence. Multiple recurrence outcomes are indicated by a digit in column 2 in *Table 12*.

Follow-up

Most studies (45/61) reported the length of follow-up, usually as the median with or without the range, but very few explained how the median was calculated.

Data quality

Data completeness was generally poorly described. In several studies, absence of missing data was an inclusion criterion, but the number excluded for that reason was usually not stated. In other papers, the initial number of patients was reported but the number included in the final model was not stated. Even when such numbers were given, they were usually hard to find and were not reported in the abstract. It is therefore impossible to tell how representative the sample was. That such omissions could be of major significance is illustrated by the study of Kaufmann and colleagues.¹⁸⁶ They reported the relevant numbers so we know that they began with 108 patients of whom 11 had died, but their prognostic model (derived from 11 variables) was based on only 57 patients and just four deaths. A few studies excluded patients lost to follow-up early in order to have a fixed-length follow-up (e.g. Burke and colleagues¹⁸⁸).

Prognostic factors used in developing the model

For all but two studies, the number of candidate variables was reported. *Figure 4* shows the distribution. Baak and colleagues¹⁹⁰ identified only the eight variables that were significant in univariate analyses and Collan and colleagues¹⁸⁷ may have analysed up to eight variables.

Handling of continuous variables

The handling of continuous variables was generally reported poorly, and was often different for univariate and multivariate analyses. *Table 10* shows the methods for just the multivariate analyses. A few papers that carried out different analyses both keeping variables continuous and then categorising have been classified in the table as keeping the data continuous.

In just 13 studies (24%), all continuous variables were kept as continuous in the modelling – in virtually all cases this was using linear terms, occasionally after log transformation. Only three studies did not have any continuous variables. The large majority of studies categorised some or all of the continuous variables, in 13 cases by dichotomising all of them. The method for choosing cut-points was usually not specified, but in at least five studies the cut-points were chosen

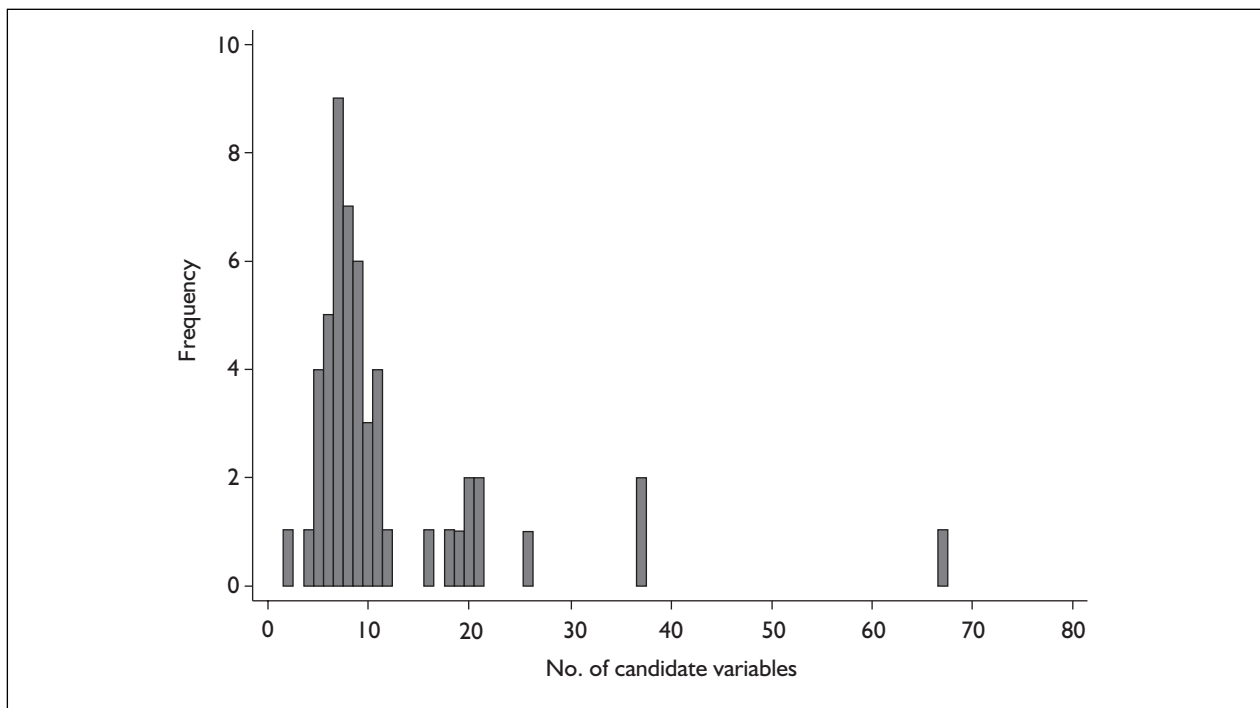


FIGURE 4 Distribution of number of candidate variables for developing prognostic model (n = 48)

in a data-dependent way (by choosing which cut-point from several gave the 'best' results). By contrast, a few studies carefully examined alternative models for continuous data and concluded that some variables had non-linear relations with survival and that it was therefore preferable to keep variables continuous and model them in a non-linear fashion.^{69,191}

Univariate analysis

It is generally helpful to present the results of univariate analyses to demonstrate the extent of association of each candidate variable to patients' outcome. Most studies (83%) did describe the results of such analyses, but only 28/45 (62%) reported the results in a table. Presentation often focused on *p*-values. Reporting of quantitative information, such as survival probability or HR, was sparse, and CIs were rarely presented.

Multivariate analysis

A total of 49 out of 50 papers presented the results of a multivariate analysis. Henson and colleagues¹⁹² did not fit a multivariate model but examined survival within a cross-classification of prognostic variables. The large majority of studies (83%) reported the results of Cox regression analysis. Of these, five also investigated one or more additional models (mainly accelerated failure time models or ANNs).

One-third of studies used all the available variables as candidates in the multivariate analysis, but 10 (19%) included only those significant with $p < 0.05$ in univariate analyses. Five studies simply took as their prognostic model all the variables that were statistically significant in univariate analysis, a highly questionable approach. About half the studies (25/53) used stepwise selection to derive a final prognostic model. Of these, two studies also used the all subsets approach. In only two studies were specific variables 'forced' into the model regardless of statistical significance.

The distribution of the number of variables in the final model is shown in *Figure 5*. For studies with more than one such model (e.g. for different outcomes), the model with the largest number of variables was taken.

Presentation of multivariate model

Rather than stop after producing a model, many authors (32/54) used their prognostic model to create risk groups, typically three or four groups (22/32). These groups were created in a variety of ways, with six studies choosing cut-points in some data-dependent manner. One-third of studies (10/32) did not indicate how the risk groups were created. Many papers did not note explicitly how many patients fell into each of their risk groups. Plots of survival in the different risk groups were presented in most papers (29/32).

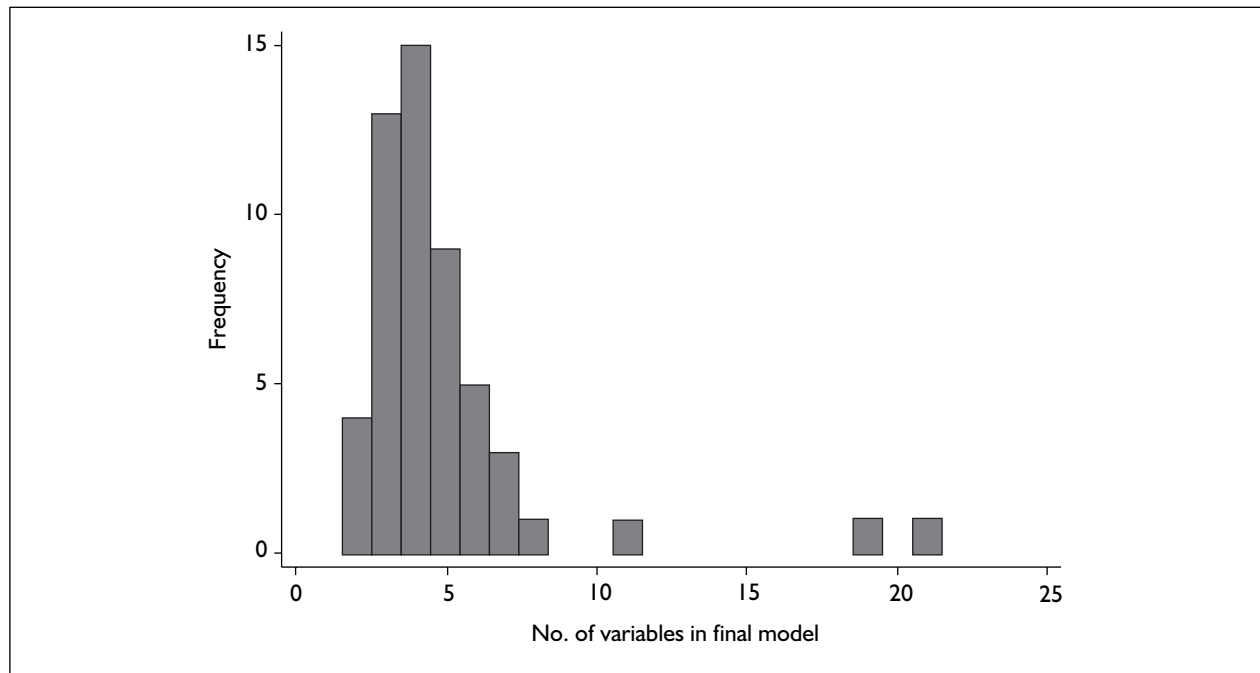


FIGURE 5 Distribution of number of variables in final prognostic model (n = 53)

To be of any potential value to other investigators or clinicians, the model needs to be presented adequately, with regression coefficients (or HRs) for all variables in the model – this was the case for only 13/54 studies (24%). We note, though, that the regression coefficients allow only statements about the relative survival of different patients. To permit an estimate of survival of individual patients ('absolute' rather than relative survival), the baseline hazard function is needed. No study gave this information (and indeed it is very rare in the medical literature at large). The baseline hazard function used in a Cox model cannot be specified simply; there is a clear advantage here (in principle) for parametric models.

Published prognostic models – review of findings

We consider first articles related to developing or validating the NPI, and then articles describing the development (and, occasionally, validation) of other models. *Table 12* shows brief details of the studies that developed models, indicating which variables were included in the models.

The NPI and derivatives

The NPI is one of the oldest indices proposed for breast cancer patients. It is one of the relatively few such indices that is actually used in clinical

practice. Its use is particularly common in the UK (Chapter 7). Likely explanations for its wide uptake include its simplicity, clinical credibility and especially the demonstration that it performs well in different populations (validation).¹²⁹

The NPI was first described by Haybittle and colleagues,¹⁹³ although a similar model had been suggested in an earlier, preliminary study.¹⁹⁴ The model was fitted to data from 387 of a cohort of 500 consecutive breast cancer patients – 79 had missing data (for variables not included in the final model), 11 had non-invasive cancer and 23 were excluded for other reasons. Nine potential prognostic variables were explored as predictors of death from any cause: age, menopausal status, tumour size, lymph node stage, tumour grade (Bloom and Richardson¹⁹⁵ criteria), cellular reaction, sinus histiocytosis, ER status and adjuvant chemotherapy (15 patients had received this). Age and tumour size were treated as continuous variables. Four ordinal variables were treated as scores from 1 to 3 (or 4) and also fitted as continuous. Thus the coefficients for these variables related to a change from one category to the next. Among these variables was lymph node involvement, coded as:

- A: tumour absent from all three nodes sampled
- B: tumour in low axillary node only
- C: tumour in apical axillary and/or internal mammary node.

TABLE 11 Cox regression model used to derive the NPI

Variable ^a	β	Z	p
Tumour size (cm)	0.172	2.92	0.004
Lymph node stage ^b	0.763	5.29	<0.0001
Tumour grade ^b	0.822	4.56	<0.0001
Postmenopausal	0.524	1.50	0.13
ER positive	-0.340	1.72	0.09

^a Only the first three variables were included in the model.
^b 3 level variables scored 1, 2, 3.
Based on Haybittle and colleagues.¹⁹³

A 'full' model with all nine variables was fitted. Three variables were statistically significant ($p < 0.01$): tumour size, lymph node stage, and tumour grade (Table 11). Also shown in Table 11 are the results for the two other variables that were closest to reaching $p < 0.05$.

Rather than fit a new model including just those three significant variables, which incidentally would have allowed them to include 79 additional patients, the authors used the coefficients from this nine-variable model to derive their prognostic index (they argued why they wanted to adjust for adjuvant therapy, but not the other variables).

The index was

$$I = 0.17 \times \text{size} + 0.76 \times \text{lymph node stage} + 0.82 \times \text{tumour grade}$$

where large values of I indicate a worse prognosis.

The index is adjusted for six other variables, and it is unknown what the model would have looked like without that adjustment. The authors noted that there was negligible impact on the index of including the other variables that were in the model.

Further investigations were made using only 298 patients – the period where some patients received adjuvant therapy was excluded. The authors created three risk groups by splitting values of the NPI at cut-points of 3.65 and 4.5, giving groups of size 154, 95 and 49 from low to high values. In a second analysis they used cut-points of 2.8 and 4.4, giving groups of size 64, 169 and 65. The choice of cut-points was not explained.

The index I was simplified to

$$I = 0.2 \times \text{size} + \text{lymph node stage} + \text{tumour grade}$$

which has become known as the NPI. The authors noted that for this modified index the cut-points of 3.4 and 5.4 correspond to the groupings using 2.8 and 4.4 for the original index. These cut-points have become standard when using the NPI.

Todd and colleagues¹⁹⁶ revisited the NPI using the extended Nottingham database. First they refitted the model to the original 387 patients but with follow-up extended from a maximum of 6.5 years to 11 years. They presented the two models for the five variables shown in Table 11. The three variables in the NPI remained the only significant ones. The regression coefficients in this updated model showed a rather larger effect of grade (0.72) and rather smaller for tumour size (0.11), but there was no suggestion that the NPI should be modified. These authors used the same cut-points of 3.4 and 5.4 to examine survival of risk groups in these patients and also a new cohort of 320 further patients; they did not present a Cox model for the new cohort. They also showed survival for all 707 patients in five groups using integer cutpoints of 3, 4, 5 and 6.

A third publication from the Nottingham group¹⁹⁷ examined the survival of a cohort of 1629 patients, including those already analysed. The index was again shown (graphically) to produce three risk groups (cut-points 3.4 and 5.4) with well-differentiated survival, similar to that seen in the original study. In this larger series, the proportions in the three risk groups were 29, 54 and 17%. They did not present a Cox model for the extended cohort. These authors made the important suggestion that lymph node stage could be replaced in the NPI by the number of nodes involved. They suggested using groups of 0, 1–3 and 4+ involved nodes. They did not present any analyses to show the impact of this change.

Several aspects of the design and analysis of the study in which the NPI was developed could be criticised. However, as noted by Altman and Royston,¹²⁴ 'clinical validity' is more important than 'statistical validity'. Despite some deviations from what is now common statistical practice, the model clearly has very good discrimination both in the original sample and in subsequent evaluations elsewhere, as described below.

Other prognostic models

No attempt has been made to summarise the discriminatory ability of the many models, partly because of a lack of a standard metric used in all papers, and partly because such measures could be influenced by major variations in case mix. By

definition, in almost all cases the variables in the models were shown to be statistically significant in the study samples, although many of the studies were small enough for concerns about over-optimism.

Rather, here the focus is on the variables which have been found important enough to feature in the various models, as shown in *Table 12*. Despite considerable heterogeneity of both clinical characteristics, the variables studied and the statistical approach to deriving a model, some clear features can be seen. A relatively small number of variables feature in more than one or two of the models. The most common variables are nodal status, tumour size and grade. Other variables often included are age, ER status and PR. A number of variables feature in just one published model. In many cases those studies were the only papers to investigate those particular factors, and presumably their inclusion reflected a particular research interest of that group. Although age featured fairly often for the end-points of death and recurrence, it was rarely important in models for predicting cancer death (see also Chapter 9).

Given the wide variety of approaches and findings, we sought to identify models that had been shown to perform well outside the original study data set and setting, as described in the next section.

Validation of prognostic models

Studies that evaluate a pre-existing model in new data are particularly valuable, especially when so many models have been developed in studies with small samples, data deficiencies and data-dependent modelling. Common practice is followed in calling such studies ‘validation studies’, although there is a case for saying that this term prejudices the outcome.¹²⁴

Table 13 gives some details of 19 identified validation studies of prognostic models in breast cancer, for 12 studies which included validation in the original report and seven that just evaluated a pre-existing model. Note that some form of validation is customary when developing a model based on an ANN, hence such validation is not included in *Table 13*. However, one ANN is included that was also evaluated on an independent data set.¹⁹⁸

Only three models have been evaluated in external data sets – that is, on new data from

different locations. The relevant studies are discussed in the following section.

Summary of validation studies

Given the need to demonstrate that a model does indeed have prognostic value, those models which have been evaluated in separate data sets are of particular importance. In this section, other studies are described which have validated previously developed models, with comments on the methodology of both the original study and the validation study.

NPI

The studies of Todd and colleagues¹⁹⁶ and Galea and colleagues,¹⁹⁷ which re-evaluated the NPI in the centre where the NPI was developed, have been discussed above. A few other studies have considered the NPI in other settings. In addition to those discussed below, Guerra and colleagues¹⁸⁴ reported the prognostic ability of the NPI in younger women with breast cancer (age <35 years) and Kollias and colleagues²²⁰ examined the performance of the NPI for women with tumours of 1 cm or smaller.

Balslev and colleagues²³⁶ and Hansen and colleagues²¹⁵

These authors evaluated the NPI in 9149 patients. This was a large, high-quality study in which patients were all enrolled in prospective, protocolled studies performed by the Danish Breast Cancer Group. Using three risk groups they found that the survival was broadly similar in the Danish cohort to that in the original Nottingham cohort, although 10-year survival was better in the poor prognostic group in the Danish series. This similarity was despite some differences in the details of assessing stage and grade from the original Nottingham study.

Hansen and colleagues²¹⁵ also evaluated the NPI in Danish patients; it seems possible that the 836 women in their study were a subset of the larger Danish cohort just discussed. They confirmed the prognostic separation of the NPI (three groups) and also found that vascular grade added significant prognostic information to the NPI.

Collett and colleagues²³⁷

These authors evaluated the NPI in 1223 patients and compared their results with those of Balslev and colleagues.²³⁶ They used the modified NPI using number of involved nodes (0, 1–3, 4+). They split the NPI into three groups at values 3.4 and 5.4 and reported 10-year survival rates similar to those of Balslev and colleagues,²³⁶

TABLE 12 Studies that developed prognostic models: variables included in final models

Study (first author)	End-point (D/C/R) ^a	Number of variables examined	Number of variables in the final model	Age	Age at menarche	Age/menopausal status	Menopausal status	Race	Surgery	Adjuvant therapy	Nodal status/+ nodes	Tumour size	TNM stage	Tumour grade	Tumour necrosis	Tumour/hist. type	Nuclear grade	Nipple invasion	Dermal infiltration	ER	PR	EGFR	Thymidine labelling index	c-erbB2/HER2	DNA ploidy/histotype	S phase fraction	MIB1	Mitotic activity index	Mucin expression	Other			
Albain, 1992 ¹⁹⁹	R	9	3																														
(Model 1 – Cox)																																	
(Model 2 – CART)	R	9	3																														
Albain, 1992 ¹⁹⁹																																	
(Model 1 – Cox)	D	9	2																														
(Model 2 – CART)	D	9	3																														
Anderson, 2000 ²⁰⁰	C	4	4																														
Aubele, 1995 ²⁰¹	R	67	4																														
Baak, 1985 ¹⁹⁰	D		3																														
Bryan, 1986 ²⁰²																																	
(Model 1)	D	10	4																														
(Model 2)	D	9	4																														
Caleffi, 1989 ²⁰³	C, R	?	4																														
Ceriani, 1992 ²⁰⁴	C, R	8	6																														
Chan, 1994 ²⁰⁵																																	
(Model 1)	C	8	4																														
(Model 2)	R	8	3																														
Chapman, 1992 ¹⁶⁸	R	7	4																														
Chapman, 1996 ²⁰⁶	R	10	5																														
Chevallier, 1990 ²⁰⁷																																	
(Model 1)	D	8	2																														
(Model 2)	R	8	3																														
Clark, 1993 ²⁰⁸	D	7	6																														
Clark, 1994 ¹²⁵	D	7	6																														
Collan, 1998 ¹⁸⁷	D	?	3																														
Coradini, 2001 ²⁰⁹	R	5	2																														
D'Eredita, 2001 ²¹⁰	D	8	3																														

continued

TABLE 12 Studies that developed prognostic models: variables included in final models

Study (first author)	End-point (D/C/R) ^a	Number of variables examined	Number of variables in the final model	Age	Age at menarche	Age/menopausal status	Menopausal status	Race	Surgery	Adjuvant therapy	Nodal status/+ nodes	Tumour size	TNM stage	Tumour grade	Tumour necrosis	Tumour/hist. type	Nuclear grade	Nipple invasion	Dermal infiltration	ER	PR	EGRF	Thymidine labelling index	c-erbB2/HER2	DNA ploidy/histotype	S phase fraction	MIB1	Mitotic activity index	Mucin expression	Other						
Eskelinen, 1990 ²¹¹	D?	6	3																																	
Fisher, 1993 ¹⁶⁶	D	11	5	•																											Elastosis					
(Model 1 – N+)	D	11	3																																	
(Model 2 – N-)	D	26	3																																	
Fisher, 1993 ²¹²	D	37	7	•																																
Fisher, 2001 ¹⁶⁷	D	12	8																																	
Foekens, 2000 ¹¹⁴	D, R	37	5																																	
Fuster, 1983 ²¹³	D	9	5																																	
Gago, 1998 ²¹⁴	M	8	4																																	
Guerra, 2003 ¹⁸⁴	D, R	10	5	•																																
Hansen, 2000 ²¹⁵	D	10	4	•																																
(Model 1)	R	10	4																																	
(Model 2)	D, R3	18	4																																	
Hawkins, 1996 ²¹⁶	D	20	4	•																																
Hawkins, 2002 ²¹⁷	R	20	4																																	
(Model 1)	D	20	4																																	
(Model 2)	R	20	4																																	
Haybittle, 1982 ¹⁹³	D	9	3																																	
Haybittle, 1986 ²¹⁸	D	6	3																																	
Henson, 1991 ¹⁹²	D	2	2																																	
Huang, 1988 ²¹⁹	D	11	5	•																																
Kaufmann, 1994 ¹⁸⁶	D, R	11	3																																	
Knorr, 1992 ¹⁹¹	D	7	7	•																																
Kollias, 1999 ²²⁰	C	5	2																																	
Korzeniowski 1994 ²²¹	D, R	6	3																																	
Lockwood, 1999 ¹⁸⁹	C, R	6	3																																	

continued

TABLE 12 Studies that developed prognostic models: variables included in final models

Study (first author)	End-point (D/C/R) ^a	Number of variables examined	Number of variables in the final model	Age	Age at menarche	Age/menopausal status	Menopausal status	Race	Surgery	Adjuvant therapy	Nodal status/+ nodes	Tumour size	TNM stage	Tumour grade	Tumour necrosis	Tumour/hist. type	Nuclear grade	Nipple invasion	Dermal infiltration	ER	PR	EGFR	Thymidine labelling index	c-erbB2/HER2	DNA ploidy/histotype	S phase fraction	MIB1	Mitotic activity index	Mucin expression	Other								
Lundin, 1999 ¹⁸⁵	C	8	6																																			
(Model 1 – Logistic)																																						
(Model 2 – ANN)	C	8	8																																			
McCready, 2000 ²²²	R	20	5																																			
Menard, 1994 ²²³	C	8	4																																			
Mitra, 1995 ²²⁴	D	8	4																																			
Parham, 1992 ²²⁵	D	5	2																																			
Perrone, 1996 ²²⁶	D, R	7	5																																			
Ravdin, 1992 ²²⁷	R	7	6																																			
(Model 1 – Cox)																																						
(Model 2 – ANN)	R	7	7																																			
Sauerbrei, 1997 ²²⁸	R	7	3																																			
(Model 1 – Cox)																																						
(Model 2 – CART)	R	7	4																																			
Sauerbrei, 1999 ⁶⁹	D	7	2																																			
(Model 1)																																						
(Model 2)	R	7	4																																			
Scholl, 1996 ²²⁹	C, R2	16	5																																			
Seshadri, 1997 ²³⁰	C, R	9	3																																			
Shek, 1988 ²³¹	C	9	4																																			
Tubiana Hulien 1993 ²³²	R	21	5																																			

continued

TABLE 12 Studies that developed prognostic models: variables included in final models

Study (first author)	End-point (D/C/R) ^a	Number of variables examined	Number of variables in the final model	Age	Age at menarche	Age/menopausal status	Menopausal status	Race	Surgery	Adjuvant therapy	Nodal status/+ nodes	Tumour size	TNM stage	Tumour grade	Tumour necrosis	Tumour/hist. type	Nuclear grade	Nipple invasion	Dermal infiltration	ER	PR	EGFR	Thymidine labelling index	c-erbB2/HER2	DNA ploidy/histotype	S phase fraction	MIB1	Mitotic activity index	Mucin expression	Other		
van der Linden 1989 ²³³ (Model 1)	D	5	3								•	•																				
(Model 2)	R	5	4								•	•																				
Younes, 1997 ²³⁴	D	6	4								•	•																				
Neural networks																																
Burke, 1997 ¹⁸⁸	C	22	22	•			•	•	•	•	•	•	•	•	•	•	•	•	•	•	•										Family history, previous biopsy, other cancer, other breast carcinoma, nipple discharge, mammogram, location of carcinoma in the breast, distant metastasis, number of lymph nodes examined, payment method	
De Laurentiis, 1999 ¹⁹⁸	R	7	7	•							•	•																			Lymph vessel invasion, hormone receptor	
Harbeck, 2000 ²³⁵	R	17	17	•			•	•	•	•	•	•	•	•	•	•	•	•	•	•	•										status, synthesis of Ki67 and S-phase fraction, uPA, PAI-1, cathepsin D, cathepsin L, cathepsin B	
Ripley, 1998 ¹⁵⁵	R	8	8	•							•	•																				
Tarassenko, 1996 ¹³¹	R	5	5	•							•	•																				

CART, classification and regression trees (or similar); N+, node positive; N-, node negative; PCNA, proliferating cell nuclear antigen.

^a End-points: C, cancer death; D, death; R, relapse/recurrence (number indicates multiple end-points, e.g. local and distant).

^b Time-dependent covariate.

TABLE 13 Characteristics of studies carrying out a validation study of a prognostic model for patients with breast cancer (n = 19)

Validation	n
Type of validation	
Internal (e.g. bootstrap)	1
Random split	5
Temporal	4
Larger series including original sample	2
External	7
By whom validation was done	
Same investigators as derived model	12
Different investigators	7

although slightly better in the lowest two groups. From an analysis of about 700 women they concluded that adding ER and PR to the NPI gave additional prognostic information, but they did not present a model with all those variables included.

These authors showed that the prognostic separation achieved by the NPI was weaker in the second 5 years than the first 5 years. Such a difference is not unexpected, but hardly any papers have considered this issue.

Sundquist and colleagues²³⁸

These authors evaluated the NPI in 608 Swedish patients and confirmed the high degree of prognostic separation between groups defined by the index (cut-points 4, 5, 6 and 7). They also fitted a new model using the same variables which they called the Kalmar Prognostic Index (KPI). The KPI was similar to the NPI, but tumour size and especially grade had higher coefficients in the KPI model:

$$\text{KPI} = 0.31 \times \text{size} + 0.79 \times \text{nodal stage} + 1.57 \times \text{grade}$$

The authors 'normalised' their model by multiplying by 0.78 so that the two indices had the same means. The two models agreed well and gave similar discrimination. The results were very similar for all-cause mortality and cancer mortality.

Other models

Just three papers reported validation studies of two models other than the NPI, and only one was not by the group who developed the model. Thus the study by Collan and colleagues,¹⁸⁷ of just 120 women, represents the only independent evaluation of any prognostic model other than the NPI.

Alexander and colleagues²³⁹

In this paper, two previously published models of Bryan and colleagues²⁰² were evaluated in a new data set. The two papers are from the same group of researchers. The two data sets were apparently from the same hospital, the first from patients recruited between July 1977 and March 1983 and the second between March 1977 and December 1983 – the authors do not mention the relation between the two datasets but do note that the second set of women had not been investigated before.

The original sample comprised those 796 women out of 3005 with adequate data, of whom 115 had died.²⁰² A total of 711 women had a full axillary clearance, of whom 34 had an unrecorded number of involved nodes. The number of nodes was imputed for these women using an unclear method based on the distribution in the remainder. A total of 694 women had their tumour size measured. Another 28 women had multifocal tumours; they were stated to have been treated as a separate group in analysis but there is no evidence of this.

Two models were derived from Cox regression analysis using 11 candidate variables. The authors investigated different ways of modelling several variables, including age and tumour size, with cut-points determined by data-dependent methods. The 'node-based' and size-based prognostic indexes were each the sum of four quantities where the score for each variable was derived from the regression coefficient in the Cox model:

Node-based

Nodes: 0 if none, 13 if 1–3, 31 if >3
 ER: 15 if ER <10 fmol
 PR: 12.5 if PR <10 fmol
 Age: # years >65

Size-based

Size: 25 if ≥4 cm
 ER: 17 if ER <10 fmol
 PR: 23 if PR <10 fmol
 Age: # years >65

Kaplan–Meier curves showed good separation across six risk groups for the node-based index, although there were few women in the extreme groups.

The second study²³⁹ included 383 and 424 women for evaluating the node-based and size-based models, respectively. An unstated number of women were excluded for not having complete data. Their Figure 2²³⁹ shows separation of the

survival curves for three risk groups based on the node-based prognostic index, but noticeably inferior discrimination to that in the original study.

Collan and colleagues¹⁸⁷

In addition to developing their own model (with variables tumour size, lymph node status and mitotic activity index), these authors evaluated the model of Baak and colleagues¹⁹⁰ (with the same three variables) on a sample of 120 women. The Baak model¹⁹⁰ did not perform particularly well using the median as a cut-point, with about two-thirds of patients correctly predicted as dead or alive at 5.5 years. They also evaluated and compared the models using three groups (using various cut-points) where the middle group was treated as uncertain prognosis and ignored in the assessment of performance. Such analyses do not give a useful assessment of performance.

van der Linden and colleagues²³³

These authors also evaluated the model of Baak and colleagues;¹⁹⁰ the two papers are from the same group of researchers. The original data were from 271 women treated at six centres from 1969 to 1976. The second data set of 195 women came from one centre from 1980 to 1983.

The prognostic model included three variables: tumour size, lymph node status and (square root of) the mitotic activity index. However, the model was developed using survival as outcome whereas the validation study used distant recurrence, partly in order to have more events to analyse (but 14 local recurrences were excluded). Even so, there were only 37 events in the validation sample. The maximum follow-up of the validation data was only about 48 months compared with about 130 months in the original study.

Despite limited data, the authors drew a very strong conclusion that “the prognostic index is, indeed, an accurate predictor of distant recurrence”.

Comments

Remarkably few published prognostic models have been re-examined by independent groups in independent settings. Most validation studies have been by the investigators themselves. The few validation studies have been carried out on ill-defined samples, sometimes of smaller size and short follow-up, and authors in general are unclear about how to summarise the performance beyond showing Kaplan–Meier plots of survival. Even when the full model is presented, only regression coefficients are provided. Estimation of actual

survival of specific patients requires that in addition the baseline hazard is specified. This part of the Cox model is non-parametric and thus the baseline hazard function is a step function that cannot be described simply. By contrast, a parametric model allows a parsimonious description of the full model to allow both absolute and relative survival to be predicted. Such models are therefore eminently more transportable to other settings. The wide preference for Cox models thus mitigates against useful transfer of models between clinical settings. These issues influenced the modelling described in Chapter 9.

As noted above, some authors used a different patient outcome when validating a model than that used to develop it, as if these outcomes were interchangeable. Few of the investigators suggested modifications to the original model in the light of the validation process. Any such changes would, of course, themselves need validation.

Overall, the only clear message from the published validation studies is support for the prognostic value of the NPI.

Discussion

In this chapter we have reviewed over 60 published studies in which authors have presented one or more prognostic models for women with newly diagnosed operable breast cancer. As noted, we have not attempted any formal assessment of the quality of these studies, although the preceding text and the information in the tables should indicate both the heterogeneity of the methodology and overall poor quality of reporting.

Some context is given by examining other published reviews of multiple prognostic models. It seems unlikely that there are many published reviews of multiple prognostic models in any medical area. Searching for such studies is even more difficult than identifying studies reporting single models, and has not been attempted; some are described below, mainly in cancer.

The identified reviews have generally not considered quality in a formal way. Nonetheless, it is clear that reviewers typically find widespread methodological deficiencies in studies that develop multivariate prognostic models. Not all of these studies have compared the prognostic performance of multiple prognostic indexes on the same set of patients.

First, Stoll²⁴⁰ reviewed many prognostic studies in breast cancer, but the emphasis was firmly on identifying which variables were prognostic, not on studies developing multivariable prognostic models.

An early relevant example is a study by Vollmer,²⁴¹ who summarised the findings of 54 multivariate analyses of survival from melanoma. He noted that there was considerable variation in the methods of analysis, the variables examined, the coding of those variables and the variables found to be significant in a multivariate model. Many of the difficulties of this review, and such reviews in general, are well captured by the opening of Vollmer's final section:²⁴¹

"In spite of 54 studies using multivariate techniques, there remain uncertainties about which prognostic factors to use in melanoma and how well we can predict the course of this disease. To some degree we must blame the methods of these studies. They often used too few patients followed for too short a time, so that the numbers of uncensored patients were limited. In general, they did not optimize the coding of prognosticators, and although the number of factors they studied was large for the number of uncensored patients, they often omitted key factors. They seldom published the coefficients of their models so that others could validate the results, and they almost never validated their models with their own test data."

Indeed, these 54 studies yielded clear evidence of the prognostic importance of just one variable, tumour thickness, and for that variable there remained uncertainty about how best to include it in a model. Such controversies may persist for many years.²⁴²

Ross and colleagues²⁴³ reviewed prognostic models for survival from prostate cancer. (We note that they inappropriately refer to these models as nomograms, a term that should be reserved for a graphical depiction of a model.) These authors observed that independent validation is necessary to see how well a model might perform in practice. Only 18 of the 42 identified models had undergone validation, of which two "partially failed". They noted that "decreased performance accuracy during validation is more often the rule than the exception". Authors reported some measure of model performance (accuracy) for only 23 of the 42 models. Vollmer and colleagues²⁴⁴ had earlier compared 13 of these published models to predict the outcome of radical prostatectomy on a single sample of men. They noted that the models did not have good discriminatory performance for binary outcomes.

Gobbi and colleagues²⁴⁵ compared seven prognostic models for patients with Hodgkin disease, five derived using Cox regression and two using parametric regression models. The models included similar but not identical sets of variables and treated the common variables in different ways. For example, age was variously handled as binary, linear or quadratic. When applied to a single data set from 315 patients, there was wide variation in the predictive ability of these models. In a stepwise Cox regression analysis with all seven indices as candidates, three were simultaneously statistically significant, leading the authors to suggest combining three prognostic indexes to achieve better performance than with any individual index.

Counsell and Dennis²⁴⁶ reviewed 83 prognostic models for stroke patients, of which 55 had examined variables prognostic for survival. They noted that over 150 different predictors had been evaluated but most of these only in one or two studies. Unlike most of the reviews already described, these authors considered carefully the quality of the individual studies. They developed a set of criteria of internal and statistical validity that they considered the minimum for studies of good quality:

- an adequate inception cohort (most patients seen within 7 days of stroke)
- prospective data collection
- less than 10% loss to follow-up
- assessment of a reliable outcome and at a fixed time point
- inclusion of age and a measure of stroke severity among the predictive variables
- at least 10 events per prognostic variable (EPV) considered for inclusion in the model
- the use of stepwise regression.

Some of these criteria are not appropriate for cancer studies, such as the use of a fixed point, although others could be questioned. Nonetheless, just four of their studies met all the criteria; these studies related to survival at three different time points (30 days, 3 months or 3 years), and none had been externally validated. The only three models that predicted survival that had been externally validated had all been evaluated on fewer than 200 patients. The authors concluded that, despite the large number of studies already done, there remained a need for better quality models.

Clinical issues

Apart from aspects of poor methodological quality and poor reporting of studies, interpretation of

the literature is further hindered by variability in the clinical aspects of the studies. Some studies focused on only node-negative or only node-positive cancers, or considered these groups separately, whereas most studies included both with nodal status included in the modelling. As noted above, the variables examined seem largely to have been determined by the data that had already been collected, although it is clear that in a few cases specific examination of stored samples allowed examination of some specific tumour characteristics.

End-points

Studies varied in which end-points they considered, with rarely any indication of the reasons for the choice. The most common end-points were death from all causes and recurrence (also known as disease-free survival), each with variations. Some studies examined both of these end-points, some just one, and others examined different end-points either instead of or in addition to these – the most common of these was cancer death. There was no consensus on definitions of clinical end-points. In some papers it was not completely clear which end-point was used, when the end-point was death it was often unclear if all deaths or only cancer deaths were considered and when the end-point was recurrence of disease it was often unclear whether deaths without recurrence were treated as events or censored.

The failure to be specific perhaps indicates an implicit view that the choice of end-point is not important, because the factors that are predictive of recurrence are the same as those predictive of death. Indeed, some authors have validated previously published models using a different clinical end-point from that used to develop the model. The assumption that the same factors apply to all end-points may be reasonable, but it is generally made without comment and there does not seem to be clear evidence to support or refute the idea.

Treatment

Most studies have used retrospective series in which patients received a variety of adjuvant therapies. Although it is likely that in some studies there was a unit protocol that determined treatments, information given was usually limited to noting which types of adjuvant therapy were received by at least some of the women. Given that the spread of prognosis is typically very much larger than the impact of any particular therapy, it is probably reasonable to disregard treatment

when developing a prognostic model. A few studies did consider type of adjuvant therapy candidate in the modelling, and in a few this variable contributed to the model.

Statistical issues

Few of the studies reviewed can be considered to be of high methodological quality. A lot of the problems stem from the general need to carry out retrospective studies, using data that have been collected for clinical purposes. Such databases are often deficient with regard to data completeness, may have problems of standardisation of measurements and the quality of the follow-up information may not be good. Also, they often by definition do not include data on recently identified markers, although there is sometimes scope to use stored samples for new assays. The potential impact of missing data has rarely been appreciated. The standard approach is to omit women without complete data, but in addition to reducing the sample size this approach will give biased results in some circumstances. The possibility of imputing missing data¹⁴¹ and the advantages and disadvantages of doing so need wider appreciation, and these methods need more empirical investigation in this context.

Nonetheless, it is clear that many studies compound the difficulties of less than ideal data with less than ideal statistical analysis methods. Many researchers have developed models on data sets which are too small. It should be more widely recognised that models developed from small data sets are unreliable. Further, when some form of variable selection is used, as is the case in most of the studies reviewed, there is a considerable risk of over-optimism, that is, the results are biased to show too much prognostic discrimination. Such problems can be alleviated by having a very large sample, by starting with a small number of important predictors, by not reducing the number of variables and by not making data-dependent choices regarding the modelling of continuous variables. Bootstrap investigation can help to investigate the stability of a prognostic model.¹⁴⁹ In addition, models should be evaluated with independent data, preferably in a different location, as discussed later.

One particular statistical issue is the handling of continuous covariates. Tumour size, age and most tumour markers are continuous variables, yet the majority of studies categorise these variables, and many dichotomise all continuous variables. Much the same remarks apply to the number of affected lymph nodes. Categorisation of variables greatly

reduces the power of a study (which is probably not large enough in the first place), and will diminish the apparent prognostic importance of those variables. Further, cut-points should be chosen in a non-data-dependent way, for example by splitting at the median or standard values. As discussed earlier in this chapter, cut-points derived by selecting the value that minimises the p -value are seriously biased and will lead to highly misleading models.

Given the long survival of many women with operable breast cancer, prospective studies are difficult. The main way to avoid the problems associated with database studies is to embed the collection of prognostic information and/or the collection of tissue and blood samples within large randomised trials as a resource for future research.

Quality of reporting

Methodological deficiencies in published studies are often compounded by deficiencies in reporting. Regardless of the specific details of a study, studies should be reported completely and accurately. It is clear that the reporting quality of the studies in this sample share many of the deficiencies seen in previous reviews.^{74,247} All papers should provide basic information about important aspects of the study including the sample selection, patient characteristics, markers examined, clinical end-points, statistical methods of analysis and the results of model fitting. *Tables 9 and 10* show that many studies failed to provide such information. Journals should ensure that reports of prognostic studies adhere to basic requirements for sound scientific reporting.

Which variables are prognostic?

Evidence on which variables are prognostic in newly diagnosed breast cancer patients is to some extent constrained by the variables that have been investigated. For few studies was it clear how the investigated variables were selected. As it is likely that the majority of these studies were retrospective, it is also likely that in most cases the choice was at least partly constrained by the information available on the local database. Also, it may be presumed that in other cases researchers choose what to study in the light of the findings of the studies already published. We cannot be sure, but it seems unlikely that the published studies have examined all known factors likely to be of major importance.

As shown in *Table 12*, some variables featured in the majority of prognostic models – in particular nodal status (number of positive nodes), tumour

size and grade, followed by age, ER and PR. Few other variables were explored in more than a handful of studies, and some factors have been studied in only one publication. As mentioned earlier, only those studies were considered that were aiming to develop a prognostic model. For many of the variables listed in *Table 12* there are many published studies that have investigated their prognostic role, but which did not develop a prognostic model. Also, some of the studies did not investigate all of the most commonly prognostic variables. Indeed, a few studies focused on a specific subset of possible prognostic information – for example, Parham and colleagues²²⁵ set out to investigate only histological information and thus addressed a narrower question.

The case of ER deserves further comment. Although ER was found to be prognostic in many studies (*Table 12*), it is likely that this was due to greater use of, and greater benefit of, of tamoxifen in patients with ER-positive tumours. At least 38% of the studies included some patients who had received hormonal treatments. As this information was very poorly reported, the true proportion may well be much higher. It must be questionable whether ER is truly prognostic, therefore, despite its statistical significance in many of the prognostic models. Its predictive ability (Chapter 6) may thus be obscuring its relative lack of prognostic importance.

Which models are useful?

A (long-term) goal is to be able to make precise forecasts of the prognosis of individual patients. Whether this could ever be achieved remains open to serious doubt. Nonetheless, prognostic models are undoubtedly useful for classifying groups of patients, for example to help choose appropriate adjuvant therapy. The extent to which a prognostic model is clinically useful has so far not been considered objectively, reflecting the lack of an agreed metric for judging the value of a prognostic model. As noted by Graf and colleagues,²⁴⁸ a measure of inaccuracy that aims to assess the value of a given prognostic model should compare the estimated event-free probabilities with the observed individual outcome. They observed that various *ad hoc* measures commonly used are of only limited value, in particular methods associated with receiver operating characteristic curves that have been borrowed from the evaluation of diagnostic tests. A recently developed index of separation, which can be used for grouped or continuous prognostic scores, may offer a valuable step forward in this regard.¹⁶²

Regardless of how well a model is able to identify groups with differing prognosis, it has to be said that no model is of any use if it is not published in enough detail. For regression models, this means that the regression equation should be published – either the HRs or lnHR should be quoted for all variables in the model. In addition, it is essential that the exact definition and numerical coding of each variable is specified.¹²⁹ For prognostic schemes based on classification trees, each subgroup needs to be clearly defined. Neural networks pose a problem in this regard as there is no easily described model. The lack of easy portability could be overcome (e.g. using the Internet) if these models had demonstrated superior performance. As yet there is little evidence to support more than marginal benefit and it is therefore unlikely that such models will become at all widely used, although they may be used in those centres in which they are developed. As noted above, however, even the familiar Cox model is not easily transported in full. The part of the model that is generally provided, the ‘prognostic model’, indicates the relative risk of different patients according to prognostic factors. Assessment of a patient’s actual risk also requires the baseline

hazard function, which is never published. Parametric models are much simpler in this regard, although thought needs to be given to whether the baseline hazard is as transportable to different settings as the PI.

Concluding comments

Despite much research effort over two decades, no new prognostic factors have been shown to add substantially to those identified in the 1980s. As Haybittle²¹⁸ noted, “Any improvement [on the NPI] in prediction must now depend on finding factors which are as important as, but independent of, lymphnode stage and pathological grade.” The NPI remains a useful clinical tool, although additional factors may enhance its use. Such factors have proved surprisingly elusive,¹⁶⁹ as is evidenced by the continued widespread use of the NPI in clinical practice (Chapter 7) after 20 years.

Further, no other prognostic model has emerged that is clearly superior to the NPI. That said, it seems clear that there is a small set of prognostic variables, perhaps especially ER (even though it is more often viewed as of predictive value), that may usefully add to the variables included in the NPI: grade, tumour size and positive lymph nodes.

Chapter 6

Predictive factors in breast cancer

The question addressed in this chapter is the use of patient/tumour factors as possible predictors of response (overall survival, disease-free survival, death) to treatment. Both individual studies reporting such predictive factors in breast cancer and reviews of these were searched for. The treatment focus for this project was systemic adjuvant therapy: hormonal therapy or chemotherapy.

A total of 7151 abstracts were initially identified for question B from across all databases: BIOSIS, 1302; CancerLit, 1542; MEDLINE, 942; EMBASE, 2906; Cochrane Library and CCTR, 459. Through de-duplication, this was reduced to 4769. De-duplication was undertaken in Procite using the terms 'Author/Title/Date', 'Title/Date', 'Author/Date', 'Author' and 'Title'. Through title screening, this was reduced to 3090. Title screening involved the removal of papers that were non-breast cancer, non-adjuvant therapy and exclusively focused on advanced/metastatic disease, whereby non-breast cancer was the principal reason for the exclusion of these papers.

Screening by SB of the 3090 abstracts reduced the number to 598, of which 298 were deemed possibly relevant to question B; 20 to question A; 69 to question C; 15 to question E; and there was no abstract for 185.

Papers potentially pertinent for questions A, C and E were entered onto the relevant Procite database for these questions. The records with no abstracts were not considered further, owing to the time constraints of the project.

Papers that had been identified as possibly relevant to question B during the screening of the databases for the other questions within the project were added to the 298 question B references identified within the screening of the hits from the question B search strategy (see Appendix 1 for details of the question B search strategy). This added a further 79 references: 59 from question A and 20 from question C. (See Appendix 1 for a full breakdown of figures.)

Five additional abstracts were identified through the hand-searching of *Breast Cancer Research and Treatment*. It was noted that two of these five had been identified through the initial search strategy, but had been excluded at the abstract screening stage. During the hand-searching (undertaken in the summer of 2002), these abstracts were deemed as relevant to the project question and were re-added to the set of abstracts to be screened by the other two reviewers for this research question. The same person who performed the initial screening of the abstracts (SB) undertook the majority of the hand-searching.

Three additional references were identified by a member of the project team during 2002.

The next stage involved two other reviewers (MC, CW) screening the 391 abstracts deemed relevant to question B. A total of 124 full papers were retrieved for data extraction; four non-English language papers were excluded at the data extraction stage. In all cases, there was a summary in English that identified that these papers did not meet the eligibility criteria for this question.

It should be noted that a number of papers were identified as relevant to both questions B and C. They were screened separately for each question, and included within the reference list for question B and/or question C where they met the relevant question eligibility criteria.

Twenty-four primary study papers were excluded and 13 review papers were excluded at the data extraction stage. Data were not extracted from 26 primary study papers. These were primary study papers of predictive factors, for which there was a good-quality, up-to-date review, and owing to a time shortage it was decided not to data extract from primary study factors for which we had at least one good-quality, up-to-date review.

Forty-eight papers were included in the report (11 reviews and 37 primary studies). An additional nine papers were deemed not to be suitable (mainly owing to poor quality), but they contained important background information, and hence were not excluded outright.

Quality and type of studies used to identify useful predictive factors

Studies designed to identify useful predictive factors are difficult to carry out. It is essential that evidence is gathered in RCTs testing the utility of the treatment. A good example of this is the systematic reviews of RCTs of adjuvant therapy of early breast cancer.^{12,13} These included RCTs of tamoxifen that included a no-tamoxifen control. In addition, many of these RCTs collected data on ER or PR status regardless of the treatment group. The systematic review of these data clearly shows that the disease-free interval and survival is only improved by tamoxifen in the group of women who are hormone receptor positive (*Figure 1*). This type of evidence is essential if we are to demonstrate that presence or absence of a particular factor is truly predictive of outcome.

All too often researchers have attempted to impute the predictive utility of specific patient or tumour factors on the basis of evidence from uncontrolled and often retrospective data. Many of these studies are relatively small in size. These practices are dangerous and likely to lead to misleading results.

In addition, showing that a factor predicts response to treatment effectively requires a test of factor \times treatment interaction. The power available to detect such an effect is markedly lower than for detecting a prognostic effect – roughly four times the sample size is necessary to have equivalent power. Analyses based on randomised trials are not protected from these dangers.

Oestrogen and progesterone receptors (ER and PR)

ER and PR are routinely used to select patients to receive or not hormone therapy such as tamoxifen. There are extensive data regarding the adjuvant therapy of early disease and also treatment for advanced and metastatic disease.^{12,13} These data come from RCTs and provide extremely strong evidence that these hormone receptors are useful predictors of response in the adjuvant therapy of breast cancer. In addition there is evidence that PR is routinely used in the selection of patients for adjuvant tamoxifen therapy in the UK (see Chapter 7 for details of a survey in the UK). For these reasons, they will not be considered in detail in this chapter.

HER2 and trastuzumab (Herceptin)

There is clear evidence from RCTs that trastuzumab is active in breast cancer in women who are strongly positive for the HER2 receptor.²⁴⁹ HER2 protein over-expression can be established by measuring expressed HER2 protein using IHC methodology. In the clinical trial studies, specimens were tested with the CTA and scored as 0, 1+, 2+ or 3+, with 3+ indicating the strongest positivity. Only patients with 2+ or 3+ positive tumours were eligible (about 33% of those screened).

Data from the randomised trial suggest that the beneficial treatment effects were largely limited to patients with the highest level of HER2 protein over-expression (3+) (see *Table 14*). In an exploratory analysis, the relative risk (RR) for time to progression was lower in the patients whose tumours tested as CTA 3+ (RR = 0.42 with 95% CI 0.33 to 0.54) than in those tested as CTA 2+ (RR = 0.76 with 95% CI 0.50 to 1.15). The RR represents the risk of progression in the trastuzumab plus chemotherapy arm versus the chemotherapy arm. Therefore, a lower ratio represents longer time to progression in the trastuzumab arm.

HER2 gene amplification detection methods

As a surrogate for protein overexpression, measurement of the number of HER2 gene copies using FISH to detect gene amplification may be employed. An exploratory, retrospective assessment of known CTA 2+ or 3+ tumour specimens was performed to detect HER2 gene amplification using PathVysion[®], a FISH assay. Data from this retrospective analysis involving 660 of 691 (96%) patients enrolled in the clinical studies (all scoring 2+ or 3+ by the CTA) suggested that the beneficial treatment effects were greater in patients whose tumours tested as FISH (+) than in those that were FISH (-); however, time to progression was prolonged for patients on the trastuzumab arm, regardless of the FISH result (*Table 14*).

These data are from advanced disease; the results of RCTs of adjuvant trastuzumab are still awaited. Because there is clear evidence that the presence of HER2 is predictive of response to trastuzumab, we will not consider this use of HER2 as a predictor of response to trastuzumab further in this chapter, which concentrates on factors where there is less robust evidence on their utility as a predictor of outcome.

TABLE 14 Treatment effect versus level of HER2 expression. Phase III randomised trial (N = 469): Herceptin plus chemotherapy versus chemotherapy

HER2 assay result	No. of patients (N)	RR ^b for time to disease progression (95% CI)	RR ^b for mortality (95% CI)
CTA 2+ or 3+	469	0.49 (0.40 to 0.61)	0.80 (0.64 to 1.00)
FISH (+) ^a	325	0.44 (0.34 to 0.57)	0.70 (0.53 to 0.91)
FISH (-) ^a	126	0.62 (0.42 to 0.94)	1.06 (0.70 to 1.63)
CTA 2+	120	0.76 (0.50 to 1.15)	1.26 (0.82 to 1.94)
FISH (+)	32	0.54 (0.21 to 1.35)	1.31 (0.53 to 3.27)
FISH (-)	83	0.77 (0.48 to 1.25)	1.11 (0.68 to 1.82)
CTA 3+	349	0.42 (0.33 to 0.54)	0.70 (0.51 to 0.90)
FISH (+)	293	0.42 (0.32 to 0.55)	0.67 (0.51 to 0.89)
FISH (-)	43	0.43 (0.20 to 0.94)	0.88 (0.39 to 1.98)

^a FISH testing results were available for 451 of the 469 patients enrolled in the study.
^b The RR represents the risk of progression or death in the trastuzumab (Herceptin) plus chemotherapy arm versus the chemotherapy arm.

RCTs of adjuvant trastuzumab

Three major trials of adjuvant trastuzumab were presented at the 2005 ASCO meeting as late-breaking news (Piccart-Gebhart and colleagues, Romond and colleagues and Perez and colleagues, all authored in 2005) and were reviewed by Sledge. The results of the joint (NSABP-B31 and NCCTG-N9831) analysis for distant disease-free survival (DDFS) were HR = 0.47 ($2p = 8 \times 10^{-10}$) and for overall survival HR = 0.67 ($2p = 0.015$). The results of the HERA trial were DDFS HR = 0.51 ($p < 0.0001$) and overall survival HR = 0.76 ($p = 0.26$). Although the results are still preliminary and await peer review publication, there is clear evidence that adjuvant trastuzumab will improve recurrence-free survival and probably overall survival. These papers were presented at a special plenary session of ASCO in Orlando, FL, in 2005, and as late-breaking news did not have an abstract for reference purposes.

Cathepsin D

The searches identified two reports of studies evaluating cathepsin D as a predictor of benefit from adjuvant treatment. One²⁵⁰ was a moderate-sized regional study of cathepsin D and the other²⁵¹ a larger regional study of cathepsin D and plasminogen activator inhibitor-1.

What was the scope of these studies?

Tetu and colleagues²⁵⁰ reported on 638 pre- and postmenopausal women with node-positive breast cancer diagnosed between 1980 and 1986. Cathepsin D was evaluated by immunohistochemistry. Follow-up was between 2.5 and 9.5 years with a median of 4.8 years. The

study was retrospective in nature and the main end-point was occurrence of distant metastases. Adjuvant therapy consisted of chemotherapy (CMF) in 180 women, hormone therapy (tamoxifen) in 148 women and a combination of both in 154.

Billgren and colleagues²⁵¹ reported on 1851 pre- and postmenopausal women with node-positive and node-negative breast cancer diagnosed between 1988 and 1992. Cathepsin D levels were measured using an enzyme-linked immunosorbent assay (ELISA). Follow-up was from 39 to 88 months with a median of 59 months. The study was retrospective in nature and the main end-point was distant recurrence-free interval. Adjuvant treatment consisted of chemotherapy (CMF) in 198 women and hormone therapy (tamoxifen) in 1136 women.

What was the quality of these studies?

The study by Tetu and colleagues²⁵⁰ is retrospective in nature, patients coming from 21 hospitals in the Quebec region of Canada. Those carrying out the assay were blinded to the clinical outcome of the women in the study; data were available for all 636 women. There is detailed description of the immunohistochemical and other laboratory procedures, but the report is lacking detail in the clinical sections. Analysis of the data was appropriate, but the absolute numbers of recurrences and deaths were not reported and there was no clear statement of how loss to follow-up was handled. How cut-points were selected was not reported. Adjuvant treatment was given according to normal practice in the various hospitals.

Longer term follow-up of this study has recently been reported.^{252,253} Node-positive and node-negative breast cancers (1348 women in total) diagnosed between 1980 and 1986 and with a minimum follow-up of 5.2 years were included. Cathepsin D expression by cancer cells did not predict distant metastases-free survival or overall survival but, by univariate analysis, cathepsin D expression by reactive stromal cells was associated with earlier recurrence and shorter survival in women with node-negative breast cancer ($p = 0.0425$) and node-positive breast cancer given adjuvant chemotherapy ($p = 0.0234$). However, cathepsin D expression by reactive stromal cells remained a significant predictor of recurrence by multivariate analyses only in a subgroup of node-positive women given adjuvant chemotherapy.

The report of Billgren and colleagues²⁵¹ is also retrospective, patients coming from hospitals in the Stockholm and Gotland region of Sweden. It is not clear whether those carrying out the assay were blinded to the clinical outcome although this seems likely given that the patients came from many hospitals and the assays were carried out in one centre. Owing to missing data, 1671 women were available for multivariate analysis of cathepsin D. There is a brief description of the assay method used and discussion of how cut-points were selected. Clinical aspects were generally well described, but the absolute numbers of recurrences and deaths were not reported and how loss to follow-up was handled was unclear. Analysis of the data was appropriate and the description of methods used was detailed. Treatment was according to regional practice guidelines.

What did the studies conclude and how strong are these conclusions?

The studies by Tetu and colleagues^{250,252,253} reported that cathepsin D expression by stromal cells was strongly associated with a worse prognosis in the subgroup of women given adjuvant chemotherapy but no such relationship was seen in women who received adjuvant hormone therapy. This could be interpreted as showing that the presence of stromal cells expressing cathepsin D predicts benefit from hormone therapy. However, the authors²⁵⁰ caution against this interpretation since a mechanism of action to explain this finding was unknown. In addition, selection of patients for different modalities of therapy may have affected the outcome and the conclusions rely on subgroup analysis.

Billgren and colleagues²⁵¹ reported on the potential of cathepsin D as a predictor of response to tamoxifen. They concluded that the level of cathepsin D appeared to predict benefit with tamoxifen in ER-positive women although the results did not reach statistical significance ($p = 0.09$).

Both of these studies provide weak evidence that cathepsin D can be used to predict response to hormone therapy (tamoxifen). However, in view of the potential for reporting bias in this literature and the relative weakness of the evidence, we conclude that there is no good evidence to support the use of cathepsin D as a predictor of response to systemic adjuvant therapy.

HER2 (HER2/neu or c-erb-B2) as a predictive factor for adjuvant chemotherapy and hormone therapy

There are a number of reviews that include some of the features of systematic reviews and these will be discussed.

What was the scope of these reviews?

Lohrisch and Piccart²⁵⁴ reviewed the usefulness of HER2/neu (HER2 or c-erb-B2) as a predictor of outcome in women with breast cancer receiving hormonal therapy or chemotherapy. They reported on eight studies in hormonal therapy, of which four were of adjuvant therapy (1784 women). They also reviewed the data from 14 trials of chemotherapy, of which 10 were adjuvant trials (6126 women).

Piccart and colleagues²⁵⁵ also reviewed this topic, but the paper contains fewer studies than that of Lohrisch and Piccart²⁵⁴ and because of this it is not presented here.

Révillion and colleagues⁸⁷ reviewed the use of HER2 extensively. This review contained a short section on the predictive value of HER2, reporting on three studies of adjuvant hormone therapy and eight studies of adjuvant chemotherapy (three high dose).

Hayes and Thor²⁵⁶ also reviewed the usefulness of HER2 as a predictor of response. This paper includes discussion of two trials of adjuvant hormone therapy (795 women). There is brief discussion of data from 14 trials of adjuvant chemotherapy (numbers not given).

Yamauchi and colleagues²⁵⁷ reviewed the use of HER2 as a predictive factor in breast cancer. This paper includes three studies of adjuvant hormone therapy (925 women), five studies of adjuvant alkylating agent therapy (2572 women) and seven studies of adjuvant anthracycline therapy (4290 women).

What was the quality of the reviews?

The review by Lohrisch and Piccart²⁵⁴ did not present an explicit search strategy and eligibility was only partially stated. Sample sizes were given, but with few baseline characteristics. There was no description of assay methods or storage. End-points included survival (no clear definition) and relapse (no definition of how deaths were treated). There was no formal assessment of the quality of reports of included studies. The review did not describe how loss to follow-up was handled. Data were presented as tabular summaries of the trials and RRs were given, but no *p*-values were given. No information was given on what cut-points were used and how they were chosen. There was no meta-analysis; data synthesis was by vote counting.

Lohrisch and Piccart²⁵⁴ concluded that the data suggested a poor outcome with tamoxifen hormone therapy for women over-expressing HER2. However, they suggested that further evidence is needed to confirm these findings. They felt that the data for chemotherapy were inconclusive, there being mixed results.

The features of the review by Yamauchi and colleagues²⁵⁷ related to quality included the following. There was an explicit search strategy which included hand-searching. However, eligibility criteria were not stated. Sample sizes were given, but with few baseline characteristics. They included some description of the assay methods used, but there was no information on storage methods. Survival was generally an end-point although relapse was an end-point in one group. There was no discussion of loss to follow-up. Once again there was no formal assessment of the quality of included studies. The review was a narrative summary of results with no meta-analysis. Data synthesis was by vote counting and HRs were given with *p*-values. Cut-points were not specified. This review concluded that there is evidence suggesting that women whose tumours over-express HER2 are likely to derive greater benefit from anthracycline-containing regimes compared with alkylating agents. They found that the evidence for use of HER2 as a predictor of response to hormone therapy was variable and was not sufficient to recommend the use of HER2 to select women likely to be resistant to hormone therapy.

The review by Révillion and colleagues⁸⁷ included the following features that may have affected quality. There was a search strategy, although this only includes CancerLit. No information on sample sizes and baseline characteristics was given. Similarly, no data on assay methods or storage were given. The main end-points were overall survival and recurrence (unclear how deaths were treated). The review did not include any formal assessment of quality and there was no discussion of loss to follow-up. There was a narrative summary of the results although no HRs or *p*-values were given. Once again cut-points were not specified, there was no meta-analysis and synthesis was by vote counting. The authors concluded that patients whose tumour over-expresses HER2 are unlikely to benefit from hormonal therapy or chemotherapy. They found that the lack of benefit from chemotherapy is prevalent for alkylating agent regimes and not anthracycline regimes, but decided that it was too early to draw definitive conclusions and that new trials were needed.

Hayes and Thor's review²⁵⁶ included the following features. There was no explicit search strategy and eligibility criteria were not specified. Sample sizes were given, but not baseline characteristics. In this review, details on assay methods were given. Overall survival and disease-free survival were the main end-points. There was no formal assessment of quality or discussion of loss to follow-up. The results were presented as a narrative summary; HRs were not given, although *p*-values were. Once again cut-points were not stated, there was no meta-analysis and synthesis was by vote counting. These reviewers concluded that HER2 over-expression predicts for relative but not absolute resistance to hormone therapy in women with ER-positive tumours. They also reported that HER2 over-expression suggests that there is a greater benefit from anthracycline chemotherapy compared with alkylating agent therapy but also felt that new trials designed to test the role of HER2 were needed. Their summary stated that the evidence was not sufficient to provide clear-cut guidelines on the role of HER2 in selecting specific treatments.

Overall conclusions

There is a consistency in the reviews. Despite suggestive evidence of predictive effects of HER2, the overall consensus is that there is insufficient evidence to develop guidelines on the use of HER2 to select specific treatment for individual patients other than for treatment with trastuzumab.

p53 as a predictive factor

Two reviews of a partly systematic nature were found and this summary is based on these reviews.

What was the scope of these studies?

Elledge and Allred²⁵⁸ reviewed 16 studies (including 3695 women) that examined the role of p53 as a predictive factor for response to chemotherapy (11 studies), tamoxifen (three studies) and radiotherapy (two studies). Some of these studies included women with metastatic disease.

Hamilton and Piccart²⁵⁹ reviewed 15 studies of p53 as a predictor of response to systemic therapy in women with breast cancer. Of these, eight studies were in women receiving systemic adjuvant therapy (one hormones, three chemotherapy/hormones, four chemotherapy).

What was the quality of the reviews?

Overall, the reviews included some of the features of a systematic review, but both were lacking essential information. The review by Elledge and Allred²⁵⁸ had no search strategy and the selection criteria of studies to be included were not stated. There was marked clinical heterogeneity within the included studies. The assay technique was described for some studies but not all. Overall survival and RFS were reported, although it was not clear how deaths were handled for recurrence and there was no discussion of loss to follow-up in the included studies. There was no discussion about the quality of included studies and whether this was assessed. Once again cut-points were not described. A *p*-value was given for individual studies, but there were no HRs. The results were presented in tabular form with no meta-analysis. Data synthesis was by vote counting.

The review by Hamilton and Piccart²⁵⁹ included an explicit search strategy, although this was restricted to MEDLINE and gave little detailed information. Selection criteria of studies to be

included were not stated. Sample sizes were given for each included study, but there were few baseline characteristics, and there appeared to be marked clinical heterogeneity among the included studies. Some details on assay methods were given. Outcomes were presented as “predictive value”, with no detail on overall or disease-free survival. Detailed results were presented in tabular form with cut-points and *p*-values being presented in the text for a minority of studies. No meta-analysis was carried out; data synthesis was by vote counting.

What did the reviews conclude and how strong were their conclusions?

Elledge and Allred²⁵⁸ stated that p53 is not predictive of response to chemotherapy and that p53 status is not significantly associated with response to tamoxifen. Hamilton and Piccart²⁵⁹ concluded that there were no data to support the use of p53 as a predictor of response to hormonal therapy or chemotherapy. Both sets of authors stress that more research is needed, but make these clear conclusions based on the evidence they reviewed.

Overall conclusions

Apart from the use of PR and HER2 to predict response to hormones and trastuzumab, respectively, there is a paucity of good-quality data on the use of patient and tumour factors as predictors of outcome. There is a major need for good-quality, well-conducted RCTs designed to assess the usefulness of predictive factors.

Such evidence should ideally only come from adequately powered RCTs testing the utility of the treatment and where data on the factor in question are collected in the great majority of patients.

Systematic review of the evidence for cathepsin D, HER2 (apart from when using trastuzumab) and p53 as predictive factors failed to show that they were useful predictors of outcome.

Chapter 7

Use of prognostic and predictive factors to select women with early breast cancer for adjuvant systemic therapy – survey and review of treatment protocols from UK breast cancer centres and units (2001–2)

Introduction

There is little systematic evidence about the patterns of use of adjuvant therapy for early breast cancer in the UK. Evidence that has been gathered from UK clinical trials suggests that there are major variations in practice and that decisions are not always made on the basis of complete clinical information. A key component of this project was to understand the current clinical use of adjuvant therapy within cancer centres and units in the UK. To facilitate such an understanding, a survey of the current clinical practice in breast cancer centres and units in the UK was undertaken at the start of the project. The aims of the survey were threefold: first, to ascertain what prognostic and predictive factors were available to and used by clinicians when selecting adjuvant systemic therapy; second, to identify the patterns of use of adjuvant therapy arising as a result of using prognostic and predictive information; and third, to ensure that subsequent analyses, undertaken within the overall project, were appropriate, practical and could be implemented within UK breast cancer units, based on knowledge gained about these units from the project survey.

The survey – questionnaires

The lead breast cancer clinician in each breast cancer centre/unit in the UK was invited to participate in a postal survey created specifically by the project team for this study. In order to assess the accuracy of the data supplied by the clinicians, a separate shorter questionnaire was sent to the lead histopathologist serving each breast cancer centre/unit.

Clinicians' questionnaire

The clinicians' questionnaire developed for the survey consisted of 11 questions compiled by the project team. The first two questions required respondents to indicate which of a given set of prognostic and predictive factors were considered to be clinically important when selecting newly diagnosed regionally localised breast cancer patients for adjuvant chemotherapy and adjuvant hormone therapy. Each question was subdivided for women aged 50 years and under and those aged over 50 years. For the purpose of this study, the age of 50 years was taken to be the indicator of menopausal status. Using the same given set of prognostic and predictive factors, the third question asked clinicians to indicate which were available to them when making decisions about adjuvant therapy. The medical members of the project team provided the set of 16 prognostic and predictive factors used in these first three questions. This set consisted of patient characteristics and histopathological, biochemical, molecular and proliferative markers. For each of these questions, space was provided for the clinician to indicate any additional factors that were available and had not been included in the initial list.

Questions four and five asked respondents to indicate if there were any other factors that were used in clinical practice or that they would like available when making decisions about adjuvant therapy for newly diagnosed regionally localised breast cancer patients. The remaining six questions were fact based, with clinicians asked to indicate:

- The minimum number of lymph nodes sampled for staging of the axilla.

- If there was a written protocol and, if there was, whether it was based on a particular prognostic model (respondents indicating the presence of a written protocol were asked to provide a copy).
- The number of newly diagnosed regionally localised breast cancer patients seen on an annual basis.
- The proportion of patients receiving adjuvant treatments and what conventional and unconventional treatment protocols were utilised within the unit.
- Whether they would be willing to take part in developing further utilities for an economic model. Contact details of clinicians agreeing to do so were passed on to the health economists on the project team.

An introductory letter providing information and instructions was devised for use with the questionnaire. This letter outlined the reason for the survey and how the information gathered would be used and stressed the anonymity of all information provided. Full contact details for a member of the project team were provided on the letter and on the front page of the questionnaire. Clinicians were asked to contact the project team should they have any questions about the questionnaire and/or the overall project.

Piloting

The questionnaire was successfully piloted using a sample of clinicians not identified as the lead oncologists at UK breast cancer centres/units. This sample included clinicians based in the UK and Australia (chosen for its similarities with the UK healthcare system) and was identified using contacts of members of the project team. All persons receiving the pilot survey were asked to complete the questionnaire and comment on its content, layout and ease of completion, particularly whether the questions and instructions were clear and understandable.

Five pilot questionnaires were returned by clinicians from the UK and Australia. All questionnaires were completed in full: there were no missing answers, and all answers were deemed appropriate (by the medical members of the project team) to the questions that were asked. No comments were received regarding the content, layout and ease of completion of the questionnaire so it was assumed by the project team that no further modifications were required. The questionnaire and the instruction letter are given in Appendix 4.

Histopathologists' questionnaire

A survey of histopathologists was included in the project in order to assess the accuracy of the data supplied by the lead breast cancer clinicians on the availability of prognostic and predictive information for use in making decisions about adjuvant therapy (question 3 of the clinicians' questionnaire). Histopathology is not concerned with all of the prognostic and predictive factors that comprised the initial response set for the clinicians' questionnaire, therefore only 11 of these 16 factors were included in the response set for the histopathology questionnaire.

The questionnaire consisted of just one question, which asked the histopathologist to indicate which of the 11 given prognostic and predictive factors were available on a pathology report to breast cancer clinicians making treatment decisions about newly diagnosed breast cancer patients.

This questionnaire (given in Appendix 4) was also accompanied by an instruction letter similar to that developed for use with the clinicians' questionnaire, and respondents were again asked to contact the project team should they have any questions about the questionnaire and/or the overall project. It was assumed the ease with which clinicians had been able to complete this question in the pilot study would extend to histopathologists, and therefore this questionnaire was not piloted separately.

The survey – study participants

Clinicians

The sample for the main study consisted of all the lead breast cancer clinicians within the UK. A database of lead breast cancer clinician contact details was created specifically for this project. Their contact details were obtained through the Cancer Director's Office (England), the Cancer Registries and Cancer Project Teams in England, Scotland, Wales and Northern Ireland and the British Breast Group Membership List for 2001. Duplicate clinician entries were removed and the database was cleaned by cross-checking details between these information sources. Where a disagreement occurred over clinician and/or hospital contact details, the Internet was initially searched to try to clarify the correct details. The Medical Directory was an additional resource that was utilised to clarify differing contact details. All disagreements were clarified through one of these

two sources. Where more than one potential lead breast cancer clinician was identified in a particular unit, members of the project team decided on the individual to whom the survey was addressed. A total of 230 clinicians were identified for the main study.

Histopathologists

The sample for the histopathologists' survey consisted of the lead histopathologist serving each identified breast cancer unit. The questionnaire was simply addressed to the 'Lead Histopathologist, Department of Histopathology, Hospital X'. The names of the relevant histopathologists were not obtained for this project. A total of 230 histopathologists received the survey.

The survey – questionnaire administration

The first mailing of both questionnaires occurred in the week commencing 10 September 2001. Each respondent received an introductory/instruction letter, a copy of the relevant questionnaire and a return stamped, addressed envelope. The lead project member personally signed each introductory letter, which, along with the questionnaire, was printed on paper headed with the logos of the HTA and the Cochrane Cancer Network. Each centre was allocated a study number which was entered on both the clinicians' and histopathologists' questionnaires. The stamped, addressed envelope identified the project member who would receive the questionnaire on its return.

Whereas the histopathologists' questionnaire was addressed only to the lead histopathologist managing breast cancer diagnosis, named individuals had been identified for receipt of the clinicians' questionnaire. In the event that this questionnaire had been addressed to the wrong person, the clinician was asked to pass on the survey to the most appropriate person.

A date for return of completed questionnaires was indicated on the first page of both questionnaires. For clinicians this was 2 weeks after the questionnaire had first been mailed out and for histopathologists 3 weeks. A second mailing of questionnaires to all non-responders was initiated 1 week after the first closing date, with recipients being asked to return the questionnaire blank if they did not wish to or felt unable to complete the survey.

Overall response from the histopathologists was high and no further reminders were necessary. Telephone calls were made to the secretaries of clinicians failing to respond to both initial and reminder questionnaires primarily to verify the contact details and addresses held by the project team. In the event that these details were incorrect, they were amended and an introductory letter and clinicians' questionnaire sent to the correct individual/address. In a number of cases, secretaries indicated that there was no lead breast cancer clinician within that unit, rather the unit was served by oncologists from larger local units or centres. For completeness of data, these oncologists were contacted by letter and asked to indicate whether the management of patients and treatment protocols utilised within the smaller units was identical with that used in the larger centres. Further, these oncologists were provided with a copy of the original questionnaire and asked to indicate whether the prognostic factors available for use in the smaller units, when making a decision about adjuvant therapy, were identical with those factors available for use in their larger centre. This information was not included in the final analysis of the questionnaire data.

One month after initial telephone contact was made, it was noted that a number of large regional centres had not responded to the questionnaire. Personal contact was made to the specific clinician or unit secretary by the project lead to encourage response to the survey.

Unit protocols

Following the above sequence of events, returned clinician questionnaires were reviewed to identify how many clinicians had reported using a written protocol to inform decisions on adjuvant therapy and had specified that they would be prepared to provide a copy (question seven). This number was then cross-checked against the number of unit protocols returned and a letter sent to ask those clinicians who were prepared to send their protocol but had not, if they would still be willing to do so.

Data analysis

Data pertaining to the importance and availability of prognostic and predictive factors were summarised using percentages, as were questions requiring a binary response. Where relevant, answers given to free text questions were grouped by a member of the project team and summarised accordingly.

Data were also analysed by responding centre type, with the distinction made between ‘regional centres’ (RC) (defined as centres providing chemotherapy, radiotherapy and possibly surgical oncology) and ‘smaller units’ (SU) (centres providing surgical oncology and hormone therapy, but referring patients to other hospitals for chemotherapy and radiotherapy). The Cancer Networks of England and the Cancer Registries of Scotland, Wales and Northern Ireland provided the information required to categorise responding centres. Responses by centre type were compared using the χ^2 test.

The survey – results

Of the 230 clinicians’ and 230 histopathologists’ questionnaires mailed, 168 clinicians’ and 196 histopathologists’ questionnaires were returned; 12 of the 230 units indicated that they did not have a breast cancer unit, giving overall response rates for clinicians of 77% (168/218) and for histopathologists 90% (196/218).

Of the 168 returned clinicians’ questionnaires, 114 were returned from the initial mail-out (68%), 33 following the first reminder and 21 following enquiry telephone calls (out of a total of 86 units telephoned). Of the 196 returned histopathologists’ questionnaires, 173 were returned from the initial mail-out (88%) and 23 following the first reminder.

Clinicians’ questionnaire – findings

Importance of prognostic and predictive factors

Table 15 shows prognostic and predictive factors identified by clinicians as being clinically important when selecting women newly diagnosed with regionally localised breast cancer for adjuvant systemic therapy. When deciding whether to administer adjuvant chemotherapy for women aged 50 years and younger, clinicians cited the most important clinical factors as nodal status (pathological) (96%), tumour grade (95%), tumour size (86%) and ER status (79%). These same factors were cited when selecting systemic adjuvant chemotherapy for newly diagnosed women aged over 50 years with regionally localised breast cancer: 96, 93, 86 and 82% respectively. Factors deemed to be the least clinically important in both patient populations were margins positive with invasive cancer (50 years and younger, 14%; over 50 years 13%); nodal status (clinical) (13% both age groups); proliferation index (5 and 6%, respectively) and bcl-2 (2% both age groups).

When selecting systemic adjuvant hormone therapy for newly diagnosed women with regionally localised breast cancer, ER was cited by clinicians as the most important clinical factor: 50 years and younger, 97%; over 50 years, 96%. For both age groups the next most important

TABLE 15 Number (%) of 168 responding clinicians indicating clinical importance of prognostic and predictive factors in selecting women newly diagnosed with regionally localised breast cancer for adjuvant systemic therapy

Factor	Chemotherapy		Hormone therapy	
	≤ 50 years	>50 years	≤ 50 years	>50 years
Age	110 (66)	108 (64)	52 (31)	52 (31)
Menopausal status	74 (44)	64 (38)	89 (53)	69 (41)
Physiological age	49 (29)	131 (78)	22 (13)	25 (15)
Grade	160 (95)	156 (93)	59 (35)	56 (33)
Histological subtype	62 (37)	57 (34)	29 (17)	30 (18)
Margins positive with invasive cancer	23 (14)	21 (13)	15 (9)	15 (9)
Nodal status (pathological)	162 (96)	162 (96)	81 (48)	78 (46)
Nodal status (clinical)	22 (13)	21 (13)	15 (9)	12 (7)
PI (e.g. Nottingham)	107 (64)	105 (63)	51 (30)	50 (30)
Size	145 (86)	144 (86)	61 (36)	59 (35)
Vascular invasion	125 (74)	124 (74)	41 (24)	39 (23)
bcl-2	4 (2)	4 (2)	2 (1)	2 (1)
ER status	132 (79)	137 (82)	163 (97)	161 (96)
HER2 ERBB2	49 (29)	40 (24)	20 (12)	18 (11)
PR	66 (39)	69 (41)	103 (61)	100 (60)
Proliferation index	9 (5)	10 (6)	3 (2)	3 (2)

factor was cited as PR (61 and 60%, respectively), followed by menopausal status (53%) for women aged 50 years and younger and nodal status (pathological) (46%) for women aged over 50 years.

Factors deemed to be the least clinically important were identical with those cited when selecting systemic adjuvant chemotherapy: margins positive with invasive cancer (9% both age groups); nodal status (clinical) (9 and 7%, respectively); proliferation index (2% both age groups) and bcl-2 (1% both age groups).

Results by 'regional centre' and 'smaller units'

Availability of prognostic and predictive factors

Table 16 shows factors reported by clinicians as being available to them when formulating patient management pathways. Nine factors were available when selecting treatment, to over 96% of responding clinicians [age, grade, vascular invasion, size, histological subtype, nodal status (pathological) (99%), ER status, menopausal status and margins positive with invasive cancer (96%)]. Proliferation index and bcl-2 were available to the least number of clinicians, 7 and 6%, respectively.

TABLE 16 Number (%) of 168 responding clinicians indicating availability of prognostic and predictive factors when selecting women newly diagnosed with regionally localised breast cancer for adjuvant systemic therapy

Factor	No. (%) indicating availability
Age	167 (99)
Menopausal status	164 (98)
Physiological age	127 (76)
Grade	167 (99)
Histological subtype	166 (99)
Margins positive with invasive cancer	162 (96)
Nodal status (pathological)	166 (99)
Nodal status (clinical)	140 (83)
PI (e.g. Nottingham)	116 (69)
Size	167 (99)
Vascular invasion	167 (99)
bcl-2	10 (6)
ER status	165 (98)
HER2 ERBB2	68 (41)
PR	116 (69)
Proliferation index	11 (7)

Importance versus availability of prognostic and predictive factors

Factors cited as being most clinically important, tumour grade, size, nodal status (pathological), ER status and menopausal status, were widely available to clinicians. Age, vascular invasion and histological subtype were also reported as being widely available yet were not identified as being the most clinically useful factors when selecting therapies for this patient population. The two factors that were least available (proliferation index and bcl-2) were also cited as the least clinically important across both treatment and age subgroups. Margins positive with invasive cancer and nodal status (clinical) were available to 96 and 83% of clinicians, yet were seen to be of minimal clinical importance by clinicians across treatment and age subgroups when considering adjuvant systemic therapy.

Across both treatments and ages, PR was the main factor cited as being clinically important (particularly when selecting hormone therapy) but not always available to the clinician. Other factors that were deemed to be clinically important but not always available were physiological age (particularly in the chemotherapy, aged over 50 years subgroup), HER2 (principally for the chemotherapy treatment groups) and a calculated PI.

Factors unavailable to clinicians

About 42% of clinicians stated that there were one or more factors other than those listed that they used when deciding on systemic adjuvant therapy for newly diagnosed women with regionally localised breast cancer. A wide range of factors were identified (Table 17). These were grouped as 'patient: physical', 'patient: disease', 'patient directed' and 'department directed'. Of these 'patient directed' was the most commonly cited by clinicians.

Some 38% of clinicians stated that there were factors that were not currently available to them that they would like to have available. All factors identified were biological or molecular in nature. Of these, HER2 was the most common factor. Cathepsin D and bcl-2 were identified as being unavailable in some smaller units.

Unit statistics and protocol

Questions 6–11 of the clinicians' questionnaire were used to elicit information on unit throughput and management protocol. A total of 64% of responding clinicians reported working in units treating over 150 new patients each year. The

TABLE 17 Additional factors used by clinicians when selecting adjuvant chemotherapy for women with early breast cancer

Grouping category	Factor description	No. of units citing factor
Patient: physical	Preoperative geriatric assessment	1
	General health	1
	Cardiac function	1
	Fertility of patient	1
	Family history	2
	Co-morbidity	14
	Physiological age and performance of patients >60	1
	Age	1
	Mobility	1
Patient: disease	Extensive in situ components	1
	Extent of DCIS (ductal carcinoma-in-situ)	1
	Grade	1
	HER2	4
	PR	1
	No. of positive nodes	1
	Size	1
Department directed	Money	1
	Availability of therapy	1
	Unit assessment of risk	1
	Oncologist's decision	1
	Rationing	1
Patient directed	Mental state	1
	Patient opinion	4
	Patient choice	22
	Patient preference	4
	Patient request/demands	1
	Patient wishes	7
	Patient acceptance	2
	Patient views	1
	Psychosocial	1
Discussion with patient/spouse	4	

smallest units treated between 20 and 50 new patients per year and had all been classified as smaller units (SUs) for the purposes of the analysis.

Clinicians were asked what proportion of newly diagnosed patients received adjuvant chemotherapy, adjuvant hormonal therapy or a combination of adjuvant chemotherapy and adjuvant hormonal therapy. *Table 18* provides a breakdown of results. This question was not responded to by a number of clinicians: 18% for adjuvant chemotherapy, 15% for adjuvant hormone therapy and 25% for a combination of adjuvant hormone and adjuvant chemotherapy.

Among clinicians who did complete the question, there was a range of responses for each treatment option (from under 20 to 100%). The most commonly cited proportions for adjuvant chemotherapy were 30–59%, accounting for 70% of responding centres. The most commonly cited proportions for adjuvant hormone therapy were

60–100%, accounting for 85% of responding centres. There were 2% of clinicians who, although providing a percentage proportion, further stated that all ER-positive patients received adjuvant hormone therapy. One clinician indicated that 100% of all postmenopausal and 65% of premenopausal patients received adjuvant hormone therapy. About 4% of clinicians reported that all their newly diagnosed patients received adjuvant hormone therapy. The proportions receiving both adjuvant hormone and adjuvant chemotherapy were more diverse. The most commonly cited proportions were 20–39%.

Of the 79% of clinicians who indicated that there was a written protocol in their unit for making decisions on adjuvant systemic therapy, 53% stated that this protocol was based on a prognostic index. The NPI was the most commonly cited index. Other indices mentioned were the Manchester Scale (by three SUs) and the Mount Vernon Index (by one SU). No other index or model was cited.

TABLE 18 Proportion of patients receiving adjuvant chemotherapy, adjuvant hormone therapy or a combination of adjuvant chemotherapy and adjuvant hormone therapy

Patient proportions, <i>p</i> (%)	No. of centres (%)		
	Chemotherapy	Hormone therapy	Chemotherapy and hormone therapy
<20	6 (3.6)	0 (0)	15 (8.9)
20 ≤ <i>p</i> < 30	18 (10.7)	1 (0.6)	32 (19)
30 ≤ <i>p</i> < 40	38 (22.6)	1 (0.6)	26 (15.5)
40 ≤ <i>p</i> < 50	28 (16.6)	6 (3.6)	16 (9.5)
50 ≤ <i>p</i> < 60	31 (18.4)	12 (7.1)	18 (10.7)
60 ≤ <i>p</i> < 70	8 (4.8)	22 (13.1)	6 (3.6)
70 ≤ <i>p</i> < 80	6 (3.6)	36 (21.4)	8 (4.8)
80–100	2 (1.2)	63 (37.5)	5 (3)
All node positive below 70 years	1 (0.6)	–	–
95% ER positive	–	2 (1.2)	–
Non-responders/proportions unknown	30 (17.9)	25 (14.9)	42 (25)

The majority of remaining responses indicated that a number of factors were used to create a 'unit-specific index'. A number of these were derivatives of the NPI, whereby other factors were added to the original NPI.

Forty-nine (37%) clinicians reporting the existence of a written protocol within their unit provided a copy of their protocol with the return of their completed questionnaire.

When sampling nodes for staging of the axilla, 82% of respondents reported that their unit had identified a minimal number of nodes that should be sampled for this purpose. By far the most commonly cited number of nodes was four (64%), the numerical range being 2–10/12. One regional centre stated that all available nodes were sampled, and one SU stated that full clearance was undertaken. One SU indicated that nodal sampling was not undertaken within their unit.

Clinicians were asked if their unit used a conventional and a non-conventional chemotherapy regime for newly diagnosed, regionally localised breast cancer. About 98% of clinicians reported that they used a conventional chemotherapy regime and 49% indicated that they sometimes used a non-conventional chemotherapy regime for this patient population. Clinicians were asked to list the regimes they used: many clinicians listed two or more regimes. The most commonly cited conventional regimes were CMF (99%) and FEC (87%). CMF was more common in the SUs, whereas FEC was more common in the regional units. Other regimes that were listed include AC (38%), Epi-CMF (8%) and CAF (4%). The most commonly cited non-conventional

regimes were 'clinical trials' (17%) and the TACT clinical trial (11%). Other regimes that were listed included FEC plus another agent, 'taxanes' and specific clinical trials. A greater range of non-conventional regimes was listed by clinicians from SUs than by clinicians in regional centres.

Histopathologists' questionnaire – findings

The aim of the questionnaire was to identify factors available on a pathology report to breast cancer clinicians selecting women with early breast cancer for adjuvant systemic therapy. The sampling frame for the questionnaire comprised the histopathologist managing breast cancer pathology within each cancer unit in the UK. All results are presented in *Table 19*.

TABLE 19 Proportion of 196 histopathologists indicating the availability of prognostic and predictive factors for use by clinicians when selecting women with early breast cancer for adjuvant systemic therapy

Factor	No. (%) indicating availability
Grade	196 (100)
Histological subtype	196 (100)
Margins positive with invasive cancer	196 (100)
Nodal status (pathological)	196 (100)
Size	196 (100)
Vascular invasion	196 (100)
bcl-2	39 (20)
ER status	196 (100)
HER2 ERBB2	121 (62)
PR	150 (77)
Proliferation index	20 (10)

The 196 responding histopathologists indicated that seven factors were available on all breast cancer pathology reports: histological subtype; grade; size; ER status; margins positive with invasive cancer; vascular invasion; and nodal status (pathological). PR status was available on 77% of breast cancer pathology reports, HER2 status on 62% of reports, bcl-2 on 20% and proliferation index available on 10% of reports.

Clinician and pathologist responses were received from 146 centres and were compared to determine the level of agreement between both sets of healthcare professionals with regard to the availability of prognostic and predictive factors. There was close agreement on the seven factors identified by histopathologists as being available on all breast cancer pathology reports. All 146 clinicians also reported grade, size and vascular invasion as being available, 145 reported that histological subtype and nodal status (pathological) were available and 144 said they had knowledge of a woman's ER status when making therapy decisions. Margins positive with invasive cancer had been reported by all 146 histopathologists as being available to clinicians; however, only 140 indicated that they would have these data to hand when selecting women for adjuvant systemic therapy.

Of the remaining four factors, 112/146 histopathologists had reported PR as being available to clinicians. Across these 112 centres, however, only 92 clinicians said they had access to such information. HER2 was reported as being available by 90 out of 146 pathologists, but only 49 out of the 90 clinicians at these units indicated that HER2 was available to them. Similarly, 32/146 pathologists indicated bcl-2 as being available yet at only four of these 32 centres did clinicians appear to be aware of this availability. Finally, 16/146 pathologists reported providing proliferation index to clinicians at their units, but of the 16 clinicians serving these units, only four appeared to be aware that this information was available.

The survey – review of unit protocols

Analysis of the 49 written unit protocols received during the survey was undertaken to determine how prognostic and predictive factors identified through the project survey as important and available to clinicians are used in routine clinical practice to select women newly diagnosed with

regionally localised breast cancer for adjuvant systemic therapy.

Of the 49 written protocols received by the project team, information contained within four protocols was insufficient to ascertain what guidelines were in place to select women for adjuvant systemic treatment (one had vital pages missing, one referred only to the symptoms and signs of breast cancer and two comprised only published tables summarising the benefits from adjuvant therapy). Forty-five protocols therefore formed the basis for the following analysis.

As expected written protocols were based in the first instance upon pathological nodal status, tumour grade and tumour size, those factors identified by clinicians during the survey as being the most clinically important when selecting adjuvant systemic therapy for women with early breast cancer. Approximately half of the 45 units (22/45, 49%) combined these factors to generate the NPI, whereas the remaining 23 units (23/45, 51%) recognised the importance of such factors but chose to develop treatment protocols without the use of a single index score.

Use of the NPI

Amongst the 22 units using the NPI to formulate treatment protocol, the number of prognostic groupings into which women could be classified on the basis of their NPI score varied between units and ranged from five down to three. Such differences may be explained by the published literature on the NPI, which over the last two decades has examined the discriminatory performance of various different index values.

Ten of the 22 protocols (45%) advocate using the NPI to split patients into five subgroups:¹⁶⁹

<i>NPI score</i>	<i>Prognosis</i>
<2.4	Excellent
2.4–3.3	Good
3.4–4.3	Moderate 1
4.4–5.3	Moderate 2
>5.4	Poor

In four protocols (18%), the recommended split is into four categories:

<i>NPI score</i>	<i>Prognosis</i>
<2.5	Very good
2.5–<3.4	Good
3.4–<5.4	Moderate
>5.4	Poor

Five protocols (23%) divide patients into three groups using the NPI, but use different index cut

off values. Only one uses the original three NPI groupings:¹⁹³

<i>NPI score</i>	<i>Prognosis</i>
<3.4	Good
3.4–5.41	Moderate
>5.41	Poor

Three use the following groupings:

<i>NPI score</i>	<i>Prognosis</i>
<3.4	Good
3.4–4.4	Moderate
>4.4	Poor

Further, one protocol uses integer scores to group women as follows:¹⁹⁶

<i>NPI score</i>	<i>Prognosis</i>
<3	Good
>3	Moderate
>5	Poor

Finally three protocols (14%) recommend using the NPI as a prognostic tool to aid with the selection of women for adjuvant systemic therapy, but do not specify any prognostic groupings or significant index values.

Supplementary variables – age/menopausal status

The capacity to benefit from adjuvant systemic therapy has been shown to vary across age groups and with menopausal status.^{12,13} In all but two of the 22 protocols using NPI-generated prognostic groupings, women being assessed for adjuvant treatment are further categorised according to either age (14 protocols) or menopausal status (six protocols). Use of age as a categorical variable differs across protocols. Three units separate women according to whether they are above or below the age of 35 years and one unit uses the following classification system: <35, 36–49 and 50–75 years. Nine protocols classify women less than 50 years of age (premenopausal) as a single category, but use various age bands to group women over the age of 50 years, for example two use 51–69, one 50–59 and 60–69 years and a further uses 50–70 and >70 years. One protocol indicates that a woman's age should be taken into consideration when assessing the appropriateness of adjuvant therapy, but does not elaborate.

Supplementary variables – ER status

The assessment of a woman's ER status is a requirement in all 22 protocols using the NPI, reflecting findings from the clinician survey that

ER status is also one of the most important clinical factors when determining suitability for adjuvant systemic therapy. In addition to ER status, nine protocols also consider the PR status of women within each prognostic group to be important, and two recommend that HER2 should also be considered if available. One unit suggests lymphovascular invasion and performance status should also be taken into account.

Protocols not using the NPI

Although 23 (51%) of the 45 available protocols choose not to advocate use of the NPI to combine formally tumour size, pathological nodal status and grade information, 21/23 still consider the three variables when making a decision on adjuvant therapy. One protocol considers only two of the three to be important (nodal status and grade), the other appears to consider hormonal status only. Age and/or menopausal status are considered to be important by 22/23 protocols. Women are assessed according to their menopausal status in 14 units and age is considered in the remaining eight.

Once again age classifications vary by protocol. Ages considered significant by two protocols are <35 and ≥70 years, whereas others suggest categorising women above and below certain ages, e.g. 50, 50–55 or 60 years. The importance of hormone receptor status is again confirmed in that all 23 protocols consider ER status as an important factor to aid with the selection of patients for adjuvant therapy. PR status is also considered by six units, HER2 by one and 10 protocols include lymphovascular invasion as a criterion to be considered.

Guidelines for adjuvant hormone therapy (45 protocols)

Table 20 summarises the guidelines from 45 protocols on prescribing adjuvant hormone therapy to women following surgery for breast cancer.

Thirteen (29%) of the 45 protocols recommend that tamoxifen be given to all ER- or PR-positive women regardless of prognosis. Two protocols contain similar recommendations, but their guidelines support the use of tamoxifen for all women with the exception of those who are ER or PR negative. It is assumed, therefore, that ER- or PR-poor women at these two units may be considered for adjuvant hormone therapy.

TABLE 20 Guidelines on prescribing adjuvant hormone therapy (tamoxifen) (45 protocols)

No. of protocols adopting guideline	Guideline
13	Prescribe for all ER- or PR-positive women
2	Prescribe for all women except those who are ER or PR negative
10	Prescribe for all ER- or PR-positive women with the exception of those identified as having a favourable prognosis
1	Prescribe for all ER- or PR-positive premenopausal women and for all postmenopausal women
1	Prescribe for all ER- or PR-positive premenopausal women and for all postmenopausal and elderly (>75 years) women
1	Prescribe for all ER- or PR-positive women plus all elderly women (>70 years)
2	Prescribe for all ER-positive women and node-positive elderly women (one unit also considers postmenopausal ER-negative node-positive or high-risk node-negative women)
1	Prescribe for all ER- or PR-positive women ≤ 69 years old with the exception of those with a favourable prognosis
4	Prescribe for all women with the exception of those with a favourable prognosis
2	1. Prescribe for any ER-positive woman with an NPI score between 2.4 and 3.3 2. Prescribe for any woman with an NPI score between 3.4 and 4.3 3. Prescribe for no woman with an NPI score >4.4
1	1. Prescribe for any woman with an NPI score <3 2. Prescribe for any ER-positive woman with an NPI score >3
1	Prescribe for any woman with Stage I, II or III tumours
1	1. Consider for some very low-risk women (NPI score <2.4) 2. Prescribe for ER-positive premenopausal and all postmenopausal women with low and moderate risk 3. Not considered beneficial for any high- or very high-risk women
1	Prescribe for non-ER-negative premenopausal women and all postmenopausal women ≤ 65 years with the exception of those with a favourable prognosis
1	Prescribe for all ER-positive premenopausal women with the exception of those with a favourable prognosis. Prescribe also for all postmenopausal women
3	Too vague to ascertain what guidelines are in place

A further 10 protocols (22%) indicate use of tamoxifen in all ER- or PR-positive women with the exception of those who have a 'favourable prognosis', where favourable prognosis is defined differently across protocols. One protocol identifies such women as over the age of 35 years, with size <1 cm, Grade I, Stage I tumours (NPI score equivalent <2.2) whereas another uses a size <2 cm, Grade I, Stage I classification (NPI score equivalent <2.4). One protocol uses an NPI score of ≤ 2.4 in combination with age (<70 years) to identify women for whom tamoxifen may not be necessary, and a further four use size <2.5 cm, Grade I, Stage I (NPI score equivalent <2.5). The usefulness of prescribing tamoxifen for women with an NPI score of ≤ 3.4 is discussed at two units, and finally one unit identifies favourable prognosis women as those having an NPI score of ≤ 3.4 or negative nodes.

Seventeen of the remaining 20 protocols include broadly similar guidelines in that the majority advocate use of tamoxifen for ER- or PR-positive pre- and postmenopausal women. These 17 differ, however, in that they also contain treatment recommendations for additional cohorts of women considered likely to benefit from adjuvant hormone therapy and/or for women for whom therapy may be considered unnecessary or ineffective. Five protocols supporting the use of tamoxifen in ER- or PR-positive pre- and postmenopausal women, each specify additional patients for whom therapy should also be prescribed. One of the five advocates tamoxifen for all non-ER-positive postmenopausal women, another for all non-ER-positive postmenopausal women plus all women over the age of 75 years. Women over the age of 70 years are given the drug at one centre whereas at another elderly

women must have nodal involvement confirmed before being prescribed tamoxifen. Node-positive ER-negative postmenopausal women and high-risk node-negative ER-negative postmenopausal women are considered for tamoxifen by one unit, as are node-positive elderly women.

Four units (who appear to have based their protocols on the 1994 *BMJ* papers by Richards and colleagues^{285,286} indicating tamoxifen for all patients irrespective of hormone receptor status) support the use of tamoxifen in all women with the exception of those presenting with favourable tumour characteristics i.e. size <2 cm, Grade I, Stage I (NPI score equivalent <2.4). One protocol recommends the use of tamoxifen for all ER- or PR-positive women aged 69 years or less, with the exception of those with a favourable prognosis (NPI score ≤ 2.4 and age <50 years).

Two protocols suggest tamoxifen for ER-positive women with an NPI score of 2.4–3.3, for all women with an NPI score of 3.4–4.3 and for no women with an NPI score exceeding 4.4. At one centre tamoxifen is administered to all women with an NPI score <3 and only ER-positive women with an NPI score >3, whereas at another tamoxifen is indicated for all Stage I, II and III women. One protocol suggests that tamoxifen may be beneficial for some very low-risk women (NPI score <2.4), should be given to all postmenopausal women and ER-positive premenopausal women with low or moderate risk, and is not beneficial for any woman with a high or very high risk.

One centre recommends use of adjuvant hormone therapy for non-ER-negative premenopausal women and all postmenopausal women below the age of 65 years, with the exception of those having an extremely favourable prognosis (size <1 cm, Grade I, Stage I tumours). Another supports the use of tamoxifen for all ER-positive premenopausal women without a favourable prognosis (defined as size <2 cm, Grade I, Stage I tumours) and all postmenopausal women.

Finally, it was not possible to ascertain any clear guidelines relating to the use of tamoxifen from three of the protocols received.

Guidelines for adjuvant chemotherapy (45 protocols)

For ease of comparison, chemotherapy guidelines are assessed first for the 22 protocols using the

NPI to group women according to prognosis. Four of the 22 protocols are excluded from this analysis. Three lacked sufficient detail to permit comparison on the basis of NPI scores and one calculated the NPI for node-negative patients only.

NPI score >5.4

Seven (39%) of the 18 protocols analysed indicate use of chemotherapy for all patients (irrespective of age and hormone receptor status) with an NPI score of >5.4. A further nine (50%) recommend administering adjuvant chemotherapy to all women with an NPI score >5.4 provided they are below a certain age threshold. Four centres use 70 years, four use 69 years and one advocates chemotherapy for all women up to the age of 49 years and will consider only fit women between the ages of 50 and 60 years. Only four of these nine protocols mention discussing treatment for ER-negative poor prognosis women above such specified age limits. Of the remaining two protocols, one recommends chemotherapy for all women except those who are postmenopausal and ER or PR positive. Here chemotherapy should be discussed. The other suggests chemotherapy for all women except those who are ER positive and between the ages of 60 and 69 years.

4.4 < NPI score ≤ 5.4

Seven (39%) protocols continue to indicate the use of chemotherapy for all patients (irrespective of age and hormone receptor status) with $4.4 < \text{NPI score} \leq 5.4$. A further seven will also administer chemotherapy to all women with NPI scores falling within the same range provided they are below 70 years (four protocols) or 69 years (three protocols). Four of these seven protocols mention discussing treatment for ER-negative poor prognosis women above such specified age limits. Amongst the remaining four protocols, one selects out all women between the ages of 50 and 75 years who are not ER negative, two recommend chemotherapy for all women except those who are postmenopausal and ER or PR positive, for whom chemotherapy should be discussed, and one chose not to support chemotherapy for any woman aged 60–69 years and for ER-positive women aged 50–59 years.

3.4 < NPI score ≤ 4.4

Just five (28%) protocols advocate the use of adjuvant chemotherapy in all women (irrespective of age and hormone receptor status) with $3.4 < \text{NPI score} \leq 4.4$. One other recommends treatment for any woman below the age of 70 years with such NPI scores. Of the four centres offering adjuvant chemotherapy to women below

the age of 50 years, two also discuss the benefits to be gained by treatment of ER-negative women aged 50–70 years. Three protocols provide treatment for ER-negative women below the age of 69 years and will discuss chemotherapy for ER-positive women only below the age of 50 years.

One protocol advocates adjuvant chemotherapy for all women with the exception of postmenopausal ER- or PR-positive women, for whom it recommends treatment should be discussed, whereas another considers ER- or PR positive women over the age of 35 years to be unsuitable for such treatment. Two protocols suggest that chemotherapy should be offered to all women below the age of 70 years with an NPI score of 4.4. However, once below this threshold score, emphasis is placed upon discussing rather than on offering chemotherapy to such women. Finally, implementation of guidelines from one protocol would see no women categorised as having this ‘moderate’ prognosis offered adjuvant chemotherapy.

2.4 < NPI score ≤ 3.4

Within this NPI range, half (nine) of the protocols contain guidelines, which vary further with NPI scores. At four centres chemotherapy is not routinely provided for any women with an NPI score between 2.41 and <2.5. Its use is supported in all premenopausal and ER-negative postmenopausal women with NPI scores between 2.5 and <3.4, and it is available to all women with an NPI score of 3.4. Two centres not supporting the use of chemotherapy in women with an NPI score >2.4 and ≤3.39 have differing guidelines for women with an NPI score equal to 3.4. One offers treatment to all women with such a score whereas the other indicates treatment for all except postmenopausal ER- or PR-positive women, for whom management should be discussed. Two protocols specify that adjuvant chemotherapy should not be given to women less than 70 years of age with NPI scores between >2.4 and ≤3.3 and should be discussed for all women less than 70 years of age with an NPI score equal to 3.4. According to the one protocol using integer score values to create prognostic groups, no woman with an NPI score between >2.4 and 3 is prescribed chemotherapy, whereas all women with a score >3 are offered treatment.

Of the remaining nine protocols containing uniform guidelines for this range of NPI scores, one supports the use of chemotherapy only in ER-negative women below the age of 35 years, another will discuss therapy only for ER-negative

women below the age of 70 years, and a third does not advocate chemotherapy for any women except those who are below the age of 50 years and are ER negative – for these women treatment should be discussed. Finally, six centres (four specifying an upper age limit of 69 years) do not support the use of chemotherapy for any woman falling into this prognostic group.

NPI score ≤ 2.4

Only one protocol recommends the use of adjuvant chemotherapy for women with an NPI score of <2.4. These women are required to have an ER-negative tumour and be below the age of 35 years.

The majority of the 23 protocols not using the NPI to group women according to prognosis still use pathological nodal status, grade and size along with other variables to assist with the selection of women for adjuvant chemotherapy. Findings from the clinicians’ survey revealed that pathological nodal status is considered to be the most important clinical factor when assessing suitability for therapy followed by tumour grade, tumour size and ER status. It is in this order that each factor and its role in informing patient management will be considered.

Node-positive women

A positive nodal status appears to be the single most important clinical factor indicative of adjuvant chemotherapy. Nine of the 23 protocols (39%) require that adjuvant chemotherapy be offered to any woman with positive nodes (two specify an upper age limit of 60, one of 65 and one of 70 years) and one (4%) supports the use of adjuvant chemotherapy for all node-positive ER-negative women below the age of 70 years (treatment for all other women must be discussed). A further 12 (52%) protocols recommend the use of adjuvant chemotherapy for all node-positive premenopausal women. Guidelines from these 12 protocols for node-positive postmenopausal women appear more cautious – women appear to be considered for rather than offered adjuvant chemotherapy, and some protocols require that in addition to having involved nodes, women also satisfy a number of additional criteria before being considered. Five of the 12 protocols propose considering any node-positive postmenopausal women (less than 65 years of age) for adjuvant chemotherapy. Of the remaining seven, each specifies different criteria that node-positive postmenopausal women must meet before being considered for adjuvant chemotherapy:

1. One protocol recommends considering all node-positive postmenopausal patients, with the exception of ER-positive women with more than 30% of nodes involved, to whom adjuvant chemotherapy should be offered.
2. One protocol supports the use of adjuvant chemotherapy in node-positive postmenopausal women provided that they fulfil several of the following criteria; tumour size >1 cm; Grade >I; vascular invasion; negative ER status.
3. One protocol advocates adjuvant chemotherapy for node-positive postmenopausal women between the ages of 50 and 60 years, but only the following 60–70 year-old women will be considered:
 - (a) ER- or PR-negative women who have 1–3 involved nodes and either a Grade III tumour ≥ 2 cm in size or any tumour ≥ 4 cm in size
 - (b) women with four or more nodes involved and either a Grade III tumour ≥ 2 cm in size or any tumour ≥ 4 cm in size.
4. One protocol offers adjuvant chemotherapy to postmenopausal women (<70 years) who have more than four positive nodes or when fewer nodes are involved, with a Grade III tumour or a negative ER status.
5. One protocol will consider node positive postmenopausal women who are HER2 positive.
6. One protocol indicates adjuvant chemotherapy for all node-positive postmenopausal women with the exception of those who are ER positive and have a Grade I tumour for whom treatment should be discussed.
7. One protocol considers adjuvant chemotherapy to be standard treatment for ER-negative postmenopausal women (<65 years) with 1–3 positive nodes.

Finally, one protocol out of the 23 (4%) was too vague to permit identification of any specific guidance pertaining to nodal status.

Node-negative women

Assuming all women are considered for adjuvant chemotherapy on the basis of their nodal status first, then according to the previous section, the majority with involved nodes are highly likely to be offered adjuvant chemotherapy regardless of tumour grade, size, ER status and so on. These other prognostic and predictive factors are relevant, however, when selecting women for adjuvant chemotherapy without involved nodes. Continuing in order of clinical importance, the role of grade along with size, ER status and any

other factors is considered for node-negative pre- and postmenopausal women.

Node-negative premenopausal women

A tumour grade of III appears to be significant when considering node-negative premenopausal women for adjuvant chemotherapy. Of the 23 protocols, 14 (61%) indicate that adjuvant chemotherapy should be offered to any node-negative premenopausal woman with a Grade III tumour and one (4%) suggests consideration of adjuvant chemotherapy for women with the same characteristics. Five (22%) protocols will offer adjuvant chemotherapy to node-negative premenopausal women provided that their Grade III tumours exceed a certain diameter. Four of the five also specify ER status and/or age requirements:

- One protocol requires that a Grade III tumour be >1 cm in diameter.
- One recommends adjuvant chemotherapy for Grade III, ER-negative and HER2-positive tumours >2 cm in diameter.
- Two protocols support adjuvant chemotherapy for grade III tumours >2 cm in diameter, for Grade III ER-negative tumours <2 cm in diameter or for women below the age of 35 years with a Grade III ER-positive tumour <2 cm in diameter.
- One protocol will support adjuvant chemotherapy for women with Grade III tumours >3 cm in diameter or who are below the age of 35 years.

Information contained within three protocols is too vague to ascertain what guidelines are in place for these women.

Table 21 presents the guidelines on adjuvant chemotherapy for node-negative premenopausal women with tumour of Grades <III. Three protocols are again too vague to be able to determine what guidelines are in place for these women and one protocol recommends that women with a tumour of Grade <III should not be considered for adjuvant chemotherapy. Of the remaining 19 protocols, 17 require that grade is considered in conjunction with one or more other factors. The two that do not, advocate the use of adjuvant chemotherapy for all node-negative premenopausal women with low or intermediate grade tumours.

Tumour size appears to be the most important factor when determining whether to offer adjuvant chemotherapy to node-negative premenopausal women with Grade I and II tumours and is

TABLE 21 Guidelines on adjuvant chemotherapy for node negative pre-menopausal women with a tumour of Grade <III (23 protocols not using the NPI)

Protocol no.	Guideline
1, 2, 3	Too vague to ascertain any guidelines
4	Not recommended for any woman with Grade I or II tumours
5	Offer to women with Grade I or II tumours
6	Offer to women with Grade II tumours
7, 8, 9	Offer to women with Grade I or II tumours >5 cm in diameter
10	Offer to women with Grade I or II tumours >3 cm in diameter
11	Offer to women with Grade I or II tumours >3 cm in diameter Consider ER-negative or -poor women with Grade II tumours <3 cm in diameter
12	1. Offer to women with Grade I or II tumours >5 cm in diameter 2. Consider women with Grade I or II tumours between 2 and 5 cm in diameter
13	1. Offer to ER-negative women with Grade I or II tumours >3 cm in diameter 2. Consider ER-positive women with Grade I or II tumours >3 cm in diameter 3. Consider ER-negative women with Grade I or II tumours <3 cm in diameter
14	Offer to women below the age of 35 years with Grade I or II tumours >3 cm in diameter
15	1. Offer to women with Grade II tumours >2 cm in diameter 2. Offer to ER-negative women with Grade II tumours <2 cm in diameter 3. Offer to ER-negative women with Grade I tumours >2 cm in diameter
16, 17	1. Offer to women with Grade I or II tumours ≥ 2 cm in diameter 2. Offer to women below the age of 35 years with Grade I or II tumours <2 cm in diameter
18	1. Offer to women with Grade I or II tumours ≥ 2 cm in diameter 2. Offer to women with Grade I or II tumours who are below the age of 35 years 3. Offer to women with Grade I or II tumours and venous invasion 4. Offer to women with ER-or PR-negative Grade I or II tumours
19	Offer to ER-negative c-erbB2-positive women with Grade II tumours >2 cm in diameter
20	1. Offer to women with Grade I or II tumours ≥ 2cm in diameter 2. Offer to ER-negative women with Grade I or II tumours 3. Offer to women ≤ 35 years of age with Grade I or II tumours 4. Offer to women with Grade I or II tumours and vascular or lymphatic invasion
21	1. Offer to women with Grade I or II tumours ≥ 2 cm in diameter 2. Offer to ER-negative women with Grade I or II tumours > 1 cm in diameter 3. Offer to women with Grade I or II tumours > 1 cm in diameter and vascular invasion
22	1. Offer to women with Grade I tumours >3 cm in diameter 2. Offer to women with Grade II tumours >2 cm in diameter 3. Consider ER-negative women with Grade I or II tumours 4. Consider women below the age of 35 years with Grade I or II tumours
23	1. Consider women below the age of 35 years with Grade I or II tumours 2. Consider ER- or PR-negative pregnant women with Grade I or II tumours

considered alone or in conjunction with other variables by 16 protocols. Much variation exists between protocols as to the diameter a Grade I or II tumour should exceed before adjuvant chemotherapy is considered beneficial. *Table 21* shows that for protocols considering size only, optimal tumour diameters range from 2 cm to >5 cm.

Seven protocols consider size along with one or more variables when selecting node-negative premenopausal women with Grade I or II tumours for adjuvant chemotherapy. For certain size

tumours, some protocols recommend treating women below the age of 35 years whereas others favour ER-negative women. Four protocols consider grade in conjunction with variables other than size, for example vascular invasion, ER or PR status and age less than 35 years.

Guidelines for node-negative postmenopausal women continue to be much more uncertain than those in place for their premenopausal counterparts. Having a Grade III tumour alone appears to carry far less weight in the decision about whether to offer adjuvant chemotherapy to

these women than is the case for premenopausal women. Only one protocol recommends offering and seven recommend considering adjuvant chemotherapy for any woman on the basis of a Grade III tumour only. Two protocols recommend that adjuvant chemotherapy be offered to node-negative postmenopausal women with Grade III tumours between the ages of 50 and 60 years only, and a further two support the use of adjuvant chemotherapy for node-negative postmenopausal women with Grade III tumours >5 cm in diameter, or for ER-negative node-negative postmenopausal women with Grade III tumours <5 cm in diameter. Of the remaining 11 protocols, information contained within three was too vague to be able to identify any clear guidance. Amongst the other eight, no two protocols used the same supplementary criteria for selecting or considering women with Grade III tumours for adjuvant chemotherapy:

- One protocol recommends considering ER-negative women with a Grade III tumour >3 cm in diameter or for a woman below the age of 35 years.
- One protocol makes no reference to tumour grade or size, but may consider ER/PR-negative women.
- One protocol recommends offering adjuvant chemotherapy to women with Grade III tumours provided that they fulfil several of the following criteria; tumour diameter >1 cm; vascular invasion; negative ER status.
- One protocol does not recommend adjuvant chemotherapy for any postmenopausal woman.
- One protocol recommends that adjuvant chemotherapy be offered to women between the ages of 50 and 60 years with a Grade III tumour, and for women between the ages of 60 and 70 years with a Grade III tumour >4 cm in diameter.
- One protocol will consider ER-negative and c-erbB2-positive women with a Grade III tumour >2 cm in diameter.
- One protocol recommends offering adjuvant chemotherapy to women with Grade III tumours >5 cm in diameter, or when the diameter is <5 cm women must also have a negative ER status. Adjuvant chemotherapy will be considered for ER-positive women with Grade III tumours <5 cm in diameter.
- One protocol requires that adjuvant chemotherapy be offered to ER-negative women with Grade III tumours >1 cm in diameter.

Table 22 presents the guidelines on adjuvant chemotherapy for node-negative postmenopausal

women with low and intermediate grade tumours. Three protocols are again too vague to be able to determine what guidelines are in place for these women and one protocol recommends that node-negative postmenopausal women should not be considered for adjuvant chemotherapy. The number of protocols using tumour size alone to select node-negative postmenopausal women with low and intermediate grade tumours for adjuvant chemotherapy is less than for node-negative premenopausal women. Once again there is little agreement between these protocols as to the size a Grade I or II tumour should exceed before adjuvant chemotherapy is considered beneficial. As with premenopausal women, the range is from 2 to 5 cm. Unlike premenopausal women, guidelines for these women are much more cautious with adjuvant chemotherapy being considered rather than offered.

As with premenopausal women, tumour size is considered in conjunction with age or ER status by several protocols. Four require women with tumours of a certain diameter to be ER negative before they will offer adjuvant chemotherapy, whereas guidelines contained within two protocols recommend that adjuvant chemotherapy is only offered to women between the ages of 50 and 60 years with certain tumour diameters. Two protocols require that in addition to having a tumour size which exceeds a certain threshold value, women must satisfy at least two other criteria before being considered for adjuvant chemotherapy. One protocol will consider ER-negative women with a tumour diameter >3 cm and who have angiolymphatic invasion. One protocol requires women to be ER negative, c-erbB2-positive and to have a tumour diameter >2 cm.

Six protocols consider grade in conjunction with variables other than size, for example ER or PR status, vascular invasion or age.

Comparison of protocols using and not using the NPI to aid with the selection of patients for adjuvant chemotherapy

Table 23 compares NPI- and non-NPI based protocols in terms of the pathological and biological characteristics that women should possess before they are considered for or offered adjuvant chemotherapy.

TABLE 22 Guidelines on adjuvant chemotherapy for node-negative postmenopausal women with a tumour of Grade <III (23 protocols not using the NPI)

Protocol no.	Guideline
1, 2, 3	Too vague to ascertain any guidelines
4	Not recommended for any woman with Grade I or II tumours
5	1. Offer to women between the ages of 50 and 60 years with a Grade II tumour 2. Offer to women between the ages of 50 and 60 years with a Grade I tumour >2 cm in diameter 3. Offer to ER negative women between the ages of 50 and 60 years with a Grade I tumour
6	Offer to women with Grade II tumours provided they fulfil several of the following criteria; tumour diameter > 1 cm; vascular invasion; ER negative
7, 8, 9	Consider women with Grade I or II tumours >5 cm in diameter
10	1. Offer to women between the ages of 50 and 60 years with a Grade I or II tumour >3 cm in diameter 2. For women over the age of 60 years, discuss based upon the expected benefits and patient fitness
11	No reference made to grade or stage – consider ER- or PR-negative women
12	Consider for women with Grade I or II tumours ≥ 2 cm in diameter
13	Consider for women with Grade I or II tumours >3 cm in diameter
14	Consider ER-negative women with Grade I or II tumours >3 cm in diameter and angiolymphatic invasion
15	1. Offer to ER-negative women with Grade II tumours >2 cm in diameter 2. Consider ER-positive women with Grade II tumours >2 cm in diameter
16, 17	1. Offer to women with Grade I or II tumours >5 cm in diameter 2. Offer to ER-negative women with Grade I or II tumours between 2 and 5 cm in diameter
18	Offer to ER-negative women between the ages of 50 and 60 years with a Grade I or II tumour
19	Consider for ER-negative c-erbB2-positive women with Grade II tumours >2 cm in diameter
20	1. Offer to women with Grade I or II tumours ≥ 2 cm in diameter 2. Offer to ER-negative women with Grade I or II tumours 3. Offer to women with Grade I or II tumours and vascular or lymphatic invasion
21	1. Offer to ER-negative women with Grade I or II tumours ≥ 2 cm in diameter 2. Offer to ER-negative women with Grade I or II tumours > 1 cm in diameter and vascular invasion
22	1. Consider for women with Grade I or II tumours >3 cm in diameter 2. Consider for ER-negative women with Grade I or II tumours
23	1. Consider for women with Grade I or II tumours >2 cm in diameter 2. Consider ER- or PR-negative women with Grade I or II tumours

For women with a prognosis described as ‘excellent’ (i.e. NPI score of 2.4 or less), published literature suggests that survival may not be significantly different from that of the age-matched population, and therefore the probable benefit from adjuvant chemotherapy is negligible.^{169,196} Table 23 shows that these findings are reflected in the guidelines put forward for ‘excellent’ prognosis women by the protocols using the NPI. Between 10 and 55% of the protocols not using the NPI, recommend considering or offering adjuvant chemotherapy to premenopausal women in the same prognostic group. Treatment is guided, however, towards women below the age of 35 years, for whom the prognosis appears particularly unfavourable regardless of tumour characteristics.

Among the protocols using the NPI, chemotherapy guidelines for women with NPI

scores between 2.41 and 3.4 on the whole reflect the fact that these women are generally considered to have a ‘good’ prognosis. Uncertainty does surround the decision about whether to consider or offer adjuvant chemotherapy to ER- or PR-negative premenopausal women with a ‘good’ prognosis. Table 23 shows that approximately half of the protocols suggest treatment may be beneficial whereas the other half do not.

For the NPI-based protocols, the total number recommending women be considered for or offered adjuvant chemotherapy remains constant across different grade/node combinations within each prognostic group. No difference exists between the number of protocols offering chemotherapy to women with a positive nodal status and an NPI score between 2.41 and 3.4 and those women having the same NPI score but achieved without nodal involvement (e.g. 50 and

TABLE 23 Comparison of NPI- and non-NPI-based protocols in the selection of women for whom chemotherapy is discussed or offered

NPI score	Protocols using the NPI				Protocols not using the NPI				
	Premenopausal women		Postmenopausal women		Premenopausal women		Postmenopausal women		
	ER/PR+	ER/PR-	ER/PR+	ER/PR-	ER/PR+	ER/PR-	ER/PR+	ER/PR-	
≤ 2.4									
GI No T ≤ 1 cm	-	-	-	-	40% (8/20) ^{b,c,d,e}	45% (9/20) ^{e,f}	10% (2/20) ^{g,h}	35% (7/20) ^{g,i}	
GI No T < 2 cm	-	-	-	-	40% (8/20) ^{b,c,d,e}	45% (9/20) ^f	10% (2/20) ^{g,h}	40% (8/20) ^{e,g,i}	
GI No T ≤ 2 cm	-	6% (1/18) ^d	-	-	45% (9/20) ^f	50% (10/20)	15% (3/20) ^g	55% (11/20) ^{g,i}	
2.41-3.4									
GI No 2 cm < T < 2.5 cm	0% (0/18)	17% (3/18) ^d	0% (0/18)	6% (1/18)	45% (9/20) ^f	60% (12/20) ^{e,j}	30% (6/20) ^{g,k}	55% (11/20) ^{g,i}	
GI No 2 cm < T < 3 cm	22% (4/18)	39% (7/18) ^d	0% (0/18)	28% (5/18)	45% (9/20) ^f	60% (12/20) ^{e,j}	30% (6/20) ^k	55% (11/20) ^{g,i}	
GI No 2 cm < T < 4 cm	22% (4/18)	39% (7/18) ^d	0% (0/18)	28% (5/18)	65% (13/20) ^f	70% (14/20) ^{e,j}	40% (8/20) ^k	65% (13/20) ^l	
GI No 2 cm < T < 5 cm	22% (4/18)	39% (7/18) ^d	0% (0/18)	28% (5/18)	65% (13/20) ^f	70% (14/20) ^{e,j}	40% (8/20) ^k	65% (13/20) ^l	
GI No 2 cm < T < 6 cm	28% (5/18)	44% (8/18) ^d	6% (1/18)	33% (6/18)	80% (16/20) ^f	85% (17/20) ^{e,j}	60% (12/20) ^k	80% (16/20) ^l	
GI No 2 cm < T < 7 cm	28% (5/18)	44% (8/18) ^d	6% (1/18)	33% (6/18)	80% (16/20) ^f	85% (17/20) ^{e,j}	60% (12/20) ^k	80% (16/20) ^l	
GI No 2 cm < T ≤ 7 cm	50% (9/18)	67% (12/18) ^d	50% (9/18)	56% (10/18)	80% (16/20) ^f	85% (17/20) ^{e,j}	60% (12/20) ^k	80% (16/20) ^l	
GI No T ≤ 1 cm	28% (5/18)	44% (8/18) ^d	6% (1/18)	33% (6/18)	45% (9/20) ^{b,c,d,e}	60% (12/20) ^{e,f,i}	20% (4/20) ^{g,h,k,m}	40% (8/20) ^{g,i,m}	
GI No T < 2 cm	28% (5/18)	44% (8/18) ^d	6% (1/18)	33% (6/18)	45% (9/20) ^{b,c,d,e}	60% (12/20) ^{f,j}	20% (4/20) ^{g,h,k,m}	45% (9/20) ^{e,g,i,m}	
GI No T ≤ 2 cm	50% (9/18)	67% (12/18) ^d	50% (9/18)	56% (10/18)	50% (10/20) ^f	65% (13/20) ^f	25% (5/20) ^{h,k,m}	60% (12/20) ^{g,i,m}	
GI N ₁₋₃ T ≤ 1 cm	28% (5/18)	44% (8/18) ^d	6% (1/18)	33% (6/18)	100% (22/22)	100% (22/22)	91% (20/22) ^{k,m,n}	100% (22/22) ^{k,m,n}	
GI N ₁₋₃ T < 2 cm	28% (5/18)	44% (8/18) ^d	6% (1/18)	33% (6/18)	100% (22/22)	100% (22/22)	91% (20/22) ^{k,m,n}	100% (22/22) ^{k,m,n}	
GI N ₁₋₃ T ≤ 2 cm	50% (9/18)	67% (12/18) ^d	50% (9/18)	56% (10/18)	100% (22/22)	100% (22/22)	91% (20/22) ^{k,m,n}	100% (22/22) ^{k,m,n}	
3.41-4.4									
GI N ₄ T ≤ 1 cm	94% (17/18) ^d	94% (17/18)	50% (9/18)	83% (15/18)	100% (22/22)	100% (22/22)	91% (20/22) ^{k,n}	95% (21/22) ^{k,m,n}	
GI N ₄ T ≤ 2 cm	94% (17/18) ^d	94% (17/18)	50% (9/18)	83% (15/18)	100% (22/22)	100% (22/22)	95% (21/22) ^{k,m,n}	95% (21/22) ^{k,m,n}	
GI N ₁₋₃ T ≤ 1 cm	94% (17/18) ^d	94% (17/18)	50% (9/18)	83% (15/18)	90% (18/20) ^p	90% (18/20) ^a	65% (13/20) ^{a,q}	85% (17/20) ^{a,m,q}	
GI N ₁₋₃ T ≤ 2 cm	94% (17/18) ^d	94% (17/18)	50% (9/18)	83% (15/18)	95% (19/20) ^p	95% (19/20) ^a	70% (14/20) ^{a,m,q}	90% (18/20) ^{a,m,q}	
GI No T ≤ 1 cm	94% (17/18) ^d	94% (17/18)	50% (9/18)	83% (15/18)	100% (22/22)	100% (22/22)	91% (20/22) ^{k,m,n}	100% (22/22) ^{k,m,n}	
GI No T ≤ 2 cm	94% (17/18) ^d	94% (17/18)	50% (9/18)	83% (15/18)	100% (22/22)	100% (22/22)	91% (20/22) ^{k,m,n}	100% (22/22) ^{k,m,n}	
GI N ₁₋₃ 2 cm < T < 3 cm	94% (17/18) ^d	94% (17/18)	50% (9/18)	83% (15/18)	55% (11/20) ^a	70% (14/20) ^{j,n}	35% (7/20) ^{g,k,m}	70% (14/20) ^{g,i,m,n}	
GI N ₁₋₃ 2 cm < T < 4 cm	94% (17/18) ^d	94% (17/18)	50% (9/18)	83% (15/18)	75% (15/20) ^f	80% (16/20) ^{a,j,n}	45% (9/20) ^{k,m}	80% (16/20) ^{l,i,m,n}	
GI N ₁₋₃ 2 cm < T < 5 cm	94% (17/18) ^d	94% (17/18)	50% (9/18)	83% (15/18)	75% (15/20) ^f	80% (16/20) ^{a,j,n}	45% (9/20) ^{k,m}	80% (16/20) ^{l,i,m,n}	
GI N ₁₋₃ 2 cm < T < 6 cm	94% (17/18) ^d	94% (17/18)	50% (9/18)	83% (15/18)	90% (18/20) ^f	95% (19/20) ^{a,j,n}	70% (14/20) ^{k,m}	95% (19/20) ^{l,i,m,n}	
GI N ₁₋₃ 2 cm < T ≤ 7 cm	94% (17/18) ^d	94% (17/18)	50% (9/18)	83% (15/18)	90% (18/20) ^f	95% (19/20) ^{a,j,n}	70% (14/20) ^{k,m}	95% (19/20) ^{l,i,m,n}	
GI No 2 cm < T < 3 cm	94% (17/18) ^d	94% (17/18)	50% (9/18)	83% (15/18)	100% (22/22)	100% (22/22)	91% (20/22) ^{k,m,n}	100% (22/22) ^{k,m,n}	
GI No 2 cm < T < 4 cm	94% (17/18) ^d	94% (17/18)	50% (9/18)	83% (15/18)	100% (22/22)	100% (22/22)	91% (20/22) ^{k,m,n}	100% (22/22) ^{k,m,n}	
GI No 2 cm < T < 5 cm	94% (17/18) ^d	94% (17/18)	50% (9/18)	83% (15/18)	100% (22/22)	100% (22/22)	91% (20/22) ^{k,m,n}	100% (22/22) ^{k,m,n}	
GI No 2 cm < T < 6 cm	94% (17/18) ^d	94% (17/18)	50% (9/18)	83% (15/18)	100% (22/22)	100% (22/22)	91% (20/22) ^{k,m,n}	100% (22/22) ^{k,m,n}	
GI No 2 cm < T < 7 cm	94% (17/18) ^d	94% (17/18)	50% (9/18)	83% (15/18)	100% (22/22)	100% (22/22)	91% (20/22) ^{k,m,n}	100% (22/22) ^{k,m,n}	
GI No 2 cm < T < 8 cm	94% (17/18) ^d	94% (17/18)	50% (9/18)	83% (15/18)	100% (22/22)	100% (22/22)	91% (20/22) ^{k,m,n}	100% (22/22) ^{k,m,n}	
GI No 2 cm < T < 9 cm	94% (17/18) ^d	94% (17/18)	50% (9/18)	83% (15/18)	100% (22/22)	100% (22/22)	91% (20/22) ^{k,m,n}	100% (22/22) ^{k,m,n}	
GI No 2 cm < T < 10 cm	94% (17/18) ^d	94% (17/18)	50% (9/18)	83% (15/18)	100% (22/22)	100% (22/22)	91% (20/22) ^{k,m,n}	100% (22/22) ^{k,m,n}	

continued

TABLE 23 Comparison of NPI- and non-NPI-based protocols in the selection of women for whom chemotherapy is discussed or offered (cont'd)

NPI score	Protocols using the NPI				Protocols not using the NPI			
	Premenopausal women		Postmenopausal women		Premenopausal women		Postmenopausal women	
	ER/PR+	ER/PR-	ER/PR+	ER/PR-	ER/PR+	ER/PR-	ER/PR+	ER/PR-
4.41-5.4								
GII N ₄₊ T ≤ 1 cm	100% (18/18)	100% (18/18)	89% (16/18)	100% (18/18) ^o	100% (22/22)	100% (22/22)	95% (21/22) ^{k,n}	95% (21/22) ^{k,n}
GII N ₄₊ T ≤ 2 cm								
GIII N ₁₋₃ T ≤ 1 cm	100% (18/18)	100% (18/18)	89% (16/18)	100% (18/18) ^o	100% (22/22)	100% (22/22)	95% (21/22) ^{k,n}	100% (22/22) ^{k,n}
GIII N ₁₋₃ T ≤ 2 cm								
GI N ₄₊ 2 cm < T < 3 cm								
GI N ₄₊ 2 cm < T < 4 cm								
GI N ₄₊ 2 cm < T < 5 cm								
GI N ₄₊ 2 cm < T < 6 cm								
GI N ₄₊ 2 cm < T ≤ 7 cm								
GII N ₁₋₃ 2 cm < T < 3 cm								
GII N ₁₋₃ 2 cm < T < 4 cm								
GII N ₁₋₃ 2 cm < T < 5 cm								
GII N ₁₋₃ 2 cm < T < 6 cm								
GII N ₁₋₃ 2 cm < T ≤ 7 cm								
GIII No 2 cm < T < 3 cm								
GIII No 2 cm < T < 4 cm								
GIII No 2 cm < T < 5 cm								
GIII No 2 cm < T < 6 cm								
GIII No 2 cm < T ≤ 7 cm								
>5.4								
GII N ₄₊ 2 cm < T < 3 cm					95% (19/20) ^a	100% (20/20) ^{a,n}	100% (20/22) ^{m,n}	100% (22/22) ^{m,n}
GII N ₄₊ 2 cm < T < 4 cm					95% (19/20)	100% (20/20) ⁿ	100% (20/22) ^{m,n}	100% (22/22) ^{m,n}
GII N ₄₊ 2 cm < T < 5 cm					95% (19/20)	100% (20/20) ⁿ	100% (20/22) ^{m,n}	100% (22/22) ^{m,n}
GII N ₄₊ 2 cm < T < 6 cm					95% (19/20)	100% (20/20) ⁿ	100% (20/22) ^{m,n}	100% (22/22) ^{m,n}
GII N ₄₊ 2 cm < T ≤ 7 cm					95% (19/20)	100% (20/20) ⁿ	100% (20/22) ^{m,n}	100% (22/22) ^{m,n}
GIII No 2 cm < T < 3 cm					95% (19/20) ^a	100% (20/20) ^{a,n}	65% (13/20) ^{m,q}	95% (19/20) ^{a,m,n,q}
GIII No 2 cm < T < 4 cm					95% (19/20)	100% (20/20) ⁿ	65% (13/20) ^{m,q}	95% (19/20) ^{m,n,q}
GIII No 2 cm < T < 5 cm					95% (19/20)	100% (20/20) ⁿ	65% (13/20) ^m	95% (19/20) ^{i,m,n}
GIII No 2 cm < T < 6 cm					95% (19/20)	100% (20/20) ⁿ	75% (15/20) ^{i,m}	95% (19/20) ^{i,m,n}
GIII No 2 cm < T ≤ 7 cm					95% (19/20)	100% (20/20) ⁿ	75% (15/20) ^{i,m}	95% (19/20) ^{i,m,n}
GII N ₄₊ 2 cm < T < 3 cm							95% (21/22) ^{k,m,n}	95% (21/22) ^{k,m,n}
GII N ₄₊ 2 cm < T < 4 cm							95% (21/22) ^{k,m,n}	95% (21/22) ^{k,m,n}
GII N ₄₊ 2 cm < T < 5 cm							95% (21/22) ^{m,n}	95% (21/22) ^{m,n}
GII N ₄₊ 2 cm < T < 6 cm							95% (21/22) ^{m,n}	95% (21/22) ^{m,n}
GII N ₄₊ 2 cm < T ≤ 7 cm							95% (21/22) ^{m,n}	95% (21/22) ^{m,n}
GIII N ₁₋₃ 2 cm < T < 3 cm								
GIII N ₁₋₃ 2 cm < T < 4 cm								
GIII N ₁₋₃ 2 cm < T < 5 cm								
GIII N ₁₋₃ 2 cm < T < 6 cm								
GIII N ₁₋₃ 2 cm < T ≤ 7 cm								
GIII N ₄₊ T any								

^a I protocol selects women <35 years only.

^b 4 protocols select women <35 years only.

^c I protocol selects women <35 years or who have venous invasion only.

continued

TABLE 23 Comparison of NPI- and non-NPI-based protocols in the selection of women for whom chemotherapy is discussed or offered (cont'd)

- ^d | protocol selects women ≤ 35 years or who have vascular/lymphatic invasion only.
- ^e | protocol selects women having a tumour size ≥ 1 cm and vascular invasion only.
- ^f | 2 protocols select women < 35 years only.
- ^g | protocol selects only fit women over the age of 60 years.
- ^h | protocol selects women with vascular/lymphatic invasion only.
- ⁱ | 2 protocols select women between the ages of 50 and 60 years only.
- ^j | protocol selects pregnant women or those < 35 years only.
- ^k | protocol selects women between the ages of 50 and 60 years only.
- ^l | protocol selects women with angiolymphatic invasion only.
- ^m | protocol requires women to fulfil several of following criteria; size > 1 cm; vascular invasion; ER negative; Grade > 1 .
- ⁿ | protocol requires women to be c-erbB2-positive.
- ^o | protocol requires women aged 50–59 years only.
- ^p | 3 protocols select women < 35 years only.
- ^q | 3 protocols select women between the ages of 50 and 60 years only.
- ^r | protocol selects women with a tumour size ≥ 2 cm only.

67% for premenopausal ER-positive and ER-negative women, respectively). In contrast, the requirement for adjuvant chemotherapy is considered by the protocols not signed up to the NPI to vary within each prognosis group according to pathological nodal status, grade or size. *Table 23* shows more clearly the importance that these protocols place upon nodal status when considering adjuvant chemotherapy.

As identified during the review of protocols, almost all advocate adjuvant chemotherapy for node-positive women regardless of tumour grade or size. The difference between the proportion of NPI- and non-NPI-based protocols recommending adjuvant chemotherapy for node-positive women is greatest within the 'good' prognostic group (NPI score 2.41–3.4); for example, only 6% of protocols using the NPI would consider adjuvant chemotherapy for an ER/PR-positive postmenopausal woman with a Grade I, Stage I, size ≤ 1 cm tumour. The corresponding figure amongst units not using the NPI is 86%. For NPI Scores exceeding 3.41, the proportions are not dissimilar. The uncertainty as to how much postmenopausal ER- or PR-positive women can benefit from adjuvant chemotherapy also seems to be shared by both types of protocol across each prognostic group.

Table 23 shows the importance placed on tumour size by the non-NPI-based protocols when determining whether adjuvant chemotherapy is offered to node-negative women with Grade I or II tumours. The greater the tumour size, the higher the risk is considered to be and the greater is the number of protocols that recommend considering or offering treatment. In the 'good' prognosis group, guidelines from these protocols result in more women being considered for or offered chemotherapy than is the case with protocols based on the NPI. This situation is reversed, however, for premenopausal node-negative Grade II women in the moderate I group (NPI score 3.41–4.4). For women with negative nodes to be assigned an NPI score >4.41 , they must have a Grade III tumour, which the review of protocols has shown to be indicative for adjuvant chemotherapy.

In general, it would appear that guidelines from protocols using the NPI result in fewer excellent or good prognosis women being considered for or offered adjuvant chemotherapy than do guidelines from non-NPI-based protocols. For high-risk or poor-prognosis women (NPI score >5.4), the opposite is true. More women are likely to be treated as a consequence of guidelines formulated

for NPI-based protocols than would be the case for women treated at centres not using the NPI to select candidates for adjuvant chemotherapy.

Conclusions and discussion

The aims of this exercise as specified in this chapter's introductory paragraph were to determine what prognostic and predictive factors were available to and used by UK clinicians when selecting women for adjuvant systemic therapy, to identify the patterns of use of adjuvant therapy arising as a result of using prognostic and predictive information and to ensure that subsequent analyses undertaken within the overall project were appropriate, practical and could be implemented within UK breast cancer units, based on knowledge gained about these units from the project survey.

The survey confirmed pathological nodal status, tumour grade, tumour size and ER status as the most clinically important factors for consideration when selecting women with early breast cancer for adjuvant systemic therapy in the UK. Clinicians reported that such factors were readily available to them, and this was verified by histopathologists managing breast cancer pathology within the same centres. Factors identified as useful by clinicians included PR and HER2, but uncertainty surrounding the extent to which these factors are available suggests the need for closer communication between pathology laboratories and clinicians. At centres where histopathologists confirmed the availability of PR, only 82% of clinicians said they had access to such information. The corresponding figure for HER2 was even lower at 54%. Such discrepancies between clinician and histopathologist as to the availability of prognostic and predictive information also extended to factors identified as being least clinically important.

Unit protocols supplied by UK breast cancer units provided the opportunity to explore patterns of use of adjuvant therapy resulting from the use of prognostic and predictive information. Reviewing these protocols revealed that although centres appear to be using the same prognostic and predictive factors when selecting women to receive adjuvant therapy, much variation in clinical practice exists between UK breast cancer centres. Such variation has been shown to occur at two levels. First, centres with protocols based upon the NPI appear to differ from centres not using the single index score. Second, within each of these two categories (NPI and non-NPI users) between-

centre variability exists in guidelines pertaining to the use of adjuvant therapy for women for whom the benefits are uncertain, such as postmenopausal women and women with a negative nodal status and low to intermediate grade tumours.

Consensus amongst units appears to be greatest when selecting women for adjuvant hormone therapy – the decision being based primarily upon ER or PR status rather than combinations of a number of factors. Guidelines as to who should receive adjuvant chemotherapy, however, have been shown to be much less uniform.

Information gleaned from this exercise helped inform the analyses presented in the remainder of the report. Identification of the principle prognostic and predictive factors used in routine practice ensured that such factors were incorporated when modelling the cost-effectiveness of various treatment choices for different groups of women. Given the differential use by centres of the same prognostic and predictive factors, the cost-effectiveness of using prognostic information to inform these treatment choices has also been examined.

Chapter 8

Cost-effectiveness of prognostic models

Introduction

The aim of the literature search for question F in this study was twofold:

1. To identify literature reporting resource use, cost and utility information relating to breast cancer care, in order to inform the model of cost-effectiveness of using prognostic information in making adjuvant therapy choices for patients with breast cancer.
2. To identify any literature in which the cost-effectiveness of using PIs or factors had been evaluated, in breast cancer care or other disease areas.

Having assessed the results of these searches and using the findings from the systematic reviews in Chapters 4, 5 and 6, the intention was then to construct a model to assess the cost-effectiveness of using prognostic information in making adjuvant therapy decisions in breast cancer; to extract parameter values from the literature, the survey and protocol review in Chapter 7 and other sources on transition probabilities, costs, cost-effectiveness and utilities, to refine and run the model and then to assess and report the results.

The search made use of the six electronic databases used in searches (a)–(e):

- BIOSIS (Biological Abstracts)
- CancerLit
- Cochrane Cancer Network's Controlled Trials Register
- CCTR
- EMBASE
- MEDLINE.

In addition, three electronic databases on economics and health economics were searched:

- EconLit
- HEED
- CRD Economic Evaluation database.

Key journals were also identified and hand-searched for the period from 1995 to 2001: *Breast Cancer Research and Treatment*, *European Journal of*

Cancer, *Journal of Clinical Oncology* and *Medical Decision Making*.

The search looked for original articles and review articles, covered the years from 1975 and included only those references where the language of summary was English.

The search strategy was initially developed with the purpose of capturing studies reporting:

1. cost estimates of breast cancer treatments (adjuvant and non-adjuvant treatment)
2. cost-effectiveness estimates [cost per life-years gained (LYG) and cost per quality-adjusted life-year (QALY) gained] of breast cancer treatments
3. decision analytic models (decision analysis, Markov models or simulation models) of cost-effectiveness of breast cancer treatments
4. utility estimates (i.e. values between 0 and 1) (also known as preference weights or quality of life values) of breast cancer health states from EuroQol or visual analogue scale or standard gamble or time trade-off or Q-TWIST
5. costs or cost-effectiveness models incorporating prognostic models (not specific to breast cancer)
6. survival estimates by different breast cancer prognoses.

The search strategy was also designed to exclude literature reporting:

1. estimates of costs of non-breast cancer treatments
2. cost-effectiveness estimates of non-breast cancer treatments
3. results of non-preference-based quality of life measures (i.e. those not measured on a 0–1 scale).

In the first instance, the following words or phrases occurring in the abstract were proposed as search terms:

1. cost(s) or cost effectiveness or cost utility analysis or economic evaluation
2. utility/utilities or preference weights or quality of life values

3. probabilities or recurrence or survival or LYG or QALY
4. model or modelling or decision analysis or Markov model or simulation model
5. prognostic
6. breast cancer
7. adjuvant treatments for breast cancer.

Following initial searches using these terms and review of results, the search strategy was refined. The final detailed strategy is reported in Appendix 1.

As the systematic reviews reported in Chapters 4, 5 and 6 were completed, development of a definitive model to estimate the effectiveness and cost-effectiveness of adjuvant therapy for individual women or groups of women with differing prognostic factors, appeared less viable. As Chapters 4 and 6 revealed, studies systematically reviewing the prognostic and predictive ability of individual factors were generally of poor quality, making specification of a definitive set of prognostic markers difficult. In addition, the absence of validated prognostic models capable of patient-specific prognosis prediction precluded estimation of treatment effectiveness and cost-effectiveness at the individual patient level. In general, it was felt that the published evidence on prognostic markers was insufficient to inform a definitive model.

Despite this finding, and in acknowledgement of the increasing role being played by prognostic information in clinical decision-making, it was still considered important to develop and present an illustrative approach or framework for incorporating prognostic information within a decision analytic model. This framework, presented in Chapter 9 (along with a detailed justification for the approach adopted), made use of a patient-level dataset containing detailed information over the period 1986–2001 on baseline characteristics, treatments and outcomes, for 1058 women treated for breast cancer in Oxford. These data were used to estimate a risk equation incorporating the main prognostic variables identified in Chapter 7 and the adoption of a modelling strategy based directly on this risk equation rather than on a Markov model with specified transition probabilities between health states. This meant that the economics literature search was no longer required for certain aspects of the cost-effectiveness modelling such as transition probabilities, although it was still valuable in providing information on previous attempts to examine the cost-effectiveness of using

prognostic information and data on costs and utilities.

Search results

Table 24 shows the number of references initially identified under search 1 (literature reporting resource use, cost and utility information relating to breast cancer care) and search 2 (cost-effectiveness of using prognostic information in any disease area). Search 1 identified 1690 references and search 2 produced 2441.

Informing of model parameters (search 1)

Abstracts of all 1690 papers were reviewed by one of the authors (KJ) and a decision made on whether to retrieve or reject each paper. Where the potential usefulness of a paper for informing model parameters was unclear, a second author (AG) reviewed the abstract and a consensus decision was made on whether to accept or reject the paper. Using this approach, the 1690 papers initially identified by the search were reduced to 236. As mentioned previously, findings from the systematic reviews presented in Chapters 4, 5 and 6 and the resultant decision to present an illustrative framework, rather than a definitive model, meant that the importance placed upon such papers for informing model parameters was greatly reduced. They were no longer required as a source of model transition probabilities; however, as shown in Chapter 9, some of the unit costs and utility decrements used within the cost-effectiveness model were obtained from the published literature. The focus of the remainder of this chapter is therefore the assessment of published studies evaluating the cost-effectiveness of using prognostic information to guide clinical decision-making.

Cost-effectiveness of using prognostic information (search 2)

To examine the extent to which the role of prognostic information in guiding the cost-effective use of healthcare resources has been evaluated in the past, abstracts of the 2441 articles identified by search two were filtered using the term ‘prognostic’ in conjunction with various terms relating to economic evaluation and cost-effectiveness. *Table 25* presents the terms used in each search and the number of hits returned.

TABLE 24 References initially identified, by source

Source of reference	No. of hits: search 1 (breast cancer)	No. of hits: search 2 (prognostic cost-effectiveness)
Biosis	341	381
CancerLit	126	25
CCTR	111	29
EconLit	3	0
EMBASE	621	1239
HEED	75	559
MEDLINE	321	146
NHSEED	80	38
Unknown/not recorded	12	24
Total	1690	2441

TABLE 25 Terms applied to search two articles and number of hits

Search terms	No. of hits
Prognostic and economic evaluation	3
Prognostic and cost-effectiveness	28
Prognostic and cost effectiveness	8
Prognostic and cost-utility	0
Prognostic and cost utility	1
Prognostic and cost-benefit	1
Prognostic and cost-benefit	0
Prognostic and cost-consequence	0
Prognostic and cost consequence	0
Prognostic and cost-minimisation	0
Prognostic and cost minimisation	0
Prognostic and cost-effective	17
Prognostic and cost effective	63
Total	121

Following removal of duplicate articles, 79 abstracts remained. Thirty-five (44%) were cancer related and 17 (22%) cardiac related. Of the remaining 27, 26 abstracts detailed research covering 23 different disease areas and treatments.

Each of the abstracts was reviewed for data indicative of a formal assessment of the cost-effectiveness of using prognostic information to inform patient treatment decisions. Following an initial screen, 49 abstracts were excluded, 26 (53%) of which were detailed primary studies examining prognostic factors but not within a cost-effectiveness framework. Sixteen (33%) abstracts described reviews or commentaries reporting on current use of prognostic markers in a clinical setting and contained a suggestion that the implications in terms of cost-effectiveness require consideration. Of the remaining seven (14%) abstracts, three described the use of databases and computing systems in clinical practice, one

provided a description of an ongoing trial, one detailed the aims of a conference, one detailed a review of modelling in economic evaluation and one was not related to healthcare.

Two reviewers undertook a second screening of the remaining 30 abstracts. A further 21 were excluded for reasons analogous to those detailed above. To the remaining nine abstracts was added one other, found by opportunistic searching. Complete manuscripts for each paper were retrieved and reviewed.

Cost-effectiveness models

Five of the 10 papers reported full economic evaluations.²⁶⁰⁻²⁶⁴ Oncology was the most frequently studied speciality, with three of the five papers modelling the cost-effectiveness of using prognostic information in cancer detection or treatment.^{260,263,264} Intensive care provided the setting for the remaining two studies.^{261,262} A brief synopsis of each study is provided below. For a more detailed description, see *Table 26*.

Oncology

Calvert and colleagues²⁶⁰ used a Markov process to model the cost-utility of using a prognostic marker (DNA ploidy) to select treatment for men with moderately differentiated prostate cancer. For patients with a diploid test result (indicative of a better prognosis), the treatment modelled was observation (watchful waiting). Diploid absent patients thought to have more aggressive cancers underwent radical prostatectomy. Other treatment policies modelled were observation for all and prostatectomy for all. The incremental cost per QALY of marker-driven care over observation for all was estimated to be £12,068 (price base 2000-1). A policy of radical surgery for all was dominated by watchful waiting.

TABLE 26 Details of the five full economic evaluations identified by filtering search 2

Study (country) (perspective) (time horizon) (currency, year) (discount rate)	Study aim	Methods	Role of prognostic information	Results	Conclusions
Littrup et al., 1994 ²⁶³ (USA) (Not specified) (1 year) (US \$, no year detailed) (Not required)	To examine the cost-effectiveness of incorporating the use of prognostic indicators into screening, biopsy and follow-up decisions for prostate cancer Main screening strategies modelled involve use of PSA levels or age-related PSA levels with or without DRE results For positive screens two biopsy policies were compared – biopsy for all (systematic) versus tailored biopsy using further prognostic information	Information considered prognostic indicators in early prostate cancer detection and measured from a cohort of 2900 men enrolled in the American Cancer Society National Prostate Cancer Detection Project was used as input parameters in a retrospective decision analytic model, built to examine the cost-effectiveness of different common screening and biopsy strategies Costs and cancers detected by each screening/biopsy combination were estimated No incremental analyses were performed	Absolute PSA is a prognostic marker for prostate cancer risk. Elevated levels of PSA may be indicative of cancer; however the test is known to return many false-positive results due to benign prostate enlargement in the screening population. Relating PSA to increasing patient age was examined for an increase in screening specificity as was the inclusion of DRE To avoid unnecessary biopsy for men with false-positive PSA or age-related PSA results, further prognostic information may be used to select out patients considered low risk. Such a tailored policy was modelled using prostate gland volume adjusted PSA, or suspicious DRE and transrectal ultrasound (TRUS) results	Average cost per cancer detected (CER) (calculations for which are unclear) for different screening (S) and biopsy (B) combinations: S CER DRE T \$4,511 PSA3 T \$3,645 PSA4 T \$3,414 PSA4 + DRE T \$4,320 PSA4 + DRE all +ve T \$4,752 Age PSA T \$3,144 Age PSA All +ve T \$3,382 Age PSA + DRE T \$4,259 Age PSA + DRE All +ve T \$4,525 (T, tailored biopsy) PSA3/PSA4 – levels indicative of a positive screen 3 and 4 ng/ml, respectively Age PSA, age-related PSA	By separating screening from biopsy decisions using prognostic information, it is possible to select out lower risk men for conservative follow-up rather than biopsy

continued

TABLE 26 Details of the five full economic evaluations identified by filtering search 2 (cont'd)

Study (country) (perspective) (time horizon) (currency, year) (discount rate)	Study aim	Methods	Role of prognostic information	Results	Conclusions
Calvert <i>et al.</i> , 2003 ²⁶⁰ (UK) (NHS) (Lifetime or 40 years) (UK £, 2000–1) (6% cost and QALY)	To model the cost–utility of using a prognostic marker (DNA ploidy) to select treatment for men with moderately differentiated prostate cancer Policies compared were: 1. marker-driven care (prostatectomy or observation conditional upon DNA ploidy results) 2. prostatectomy for all 3. observation for all	Transition probabilities, treatment costs, survival probabilities and utility scores were obtained from published literature and local data sources and used to populate a decision analytic Markov model Baseline ploidy sensitivity and specificity values were assumed. Such values were varied extensively during sensitivity analysis Lifetime costs and QALYs were modelled for each policy Observation for all was presented as the comparator	The DNA ploidy test provides information on a cell's chromosomal content. Patients returning a diploid test result are thought to have better differentiated tumours, more balanced oncogenes, tumour suppressor genes and therefore a better prognosis. For these patients, an observational strategy is modelled Non-diploid patients are assumed to have more aggressive disease, a poorer prognosis and are selected for treatment by radical prostatectomy	Results based on modelling a cohort of 1000 60-year-old men with moderately differentiated prostate cancer Costs: Observation (all) £1,659,051 Ploidy marker £2,948,177 Prostatectomy (all) £6,389,507 QALYS: Observation (all) 7095 Ploidy marker 7202 Prostatectomy (all) 7068 Incremental cost per QALY: Ploidy marker £12,068 Prostatectomy (all) Dominated	Selecting patients for treatment on the basis of a prognostic marker would be cost-effective provided that test specificity levels remain at or above 80%
Glance <i>et al.</i> , 1998 ²⁶¹ (USA) (Provider) (Not clear) (US \$, no year detailed) (Not required)	To determine the cost-effectiveness of continuing intensive care treatment for patients predicted using a prognostic scoring system to have >90% chance of death after 48 h in the intensive care unit (ICU) Alternative policy modelled is withdrawal of care on the basis of the prognostic scoring system	A decision analytic framework was used to model costs and survival to hospital discharge of continuing care for patients identified using a prognostic model to be medically futile cases The model is populated mainly using data from a cohort of 4106 patients treated in the ICU unit of a single hospital over a 9-year period. Performance of the prognostic scoring system is taken from a published study Results presented as incremental cost per death prevented	Prognostic scoring index based on daily Acute Physiology and Chronic Health Evaluation (APACHE) III scores was used as a tool to identify patients expected to have >90% mortality after 48 h in ICU APACHE scores were developed using multivariate logistic regression. Explanatory variables include those relating to a patient's physiology, chronic health conditions, age and disease category	Incremental cost per death prevented of continuing ICU care for patients predicted to have >90% mortality after 48 h is estimated at \$263,700 Sensitivity analysis demonstrated that results are extremely sensitive to small changes in the specificity of the scoring system Loss of predictive power arising when applying the scoring system outside the database on which it was developed suggests that its use in informing the withdrawal of care from individuals in the ICU is unlikely to be cost-effective	

continued

TABLE 26 Details of the five full economic evaluations identified by filtering search 2 (cont'd)

Study (country) (perspective) (time horizon) (currency, year) (discount rate)	Study aim	Methods	Role of prognostic information	Results	Conclusions
Mason, 1999 ²⁶⁴ (UK) (Provider) (3.3 years) (UK £, 1998–9) (5% cost and life-years)	To examine the cost-effectiveness of using the differential staining cytotoxicity (DiSC) assay to guide the treatment of patients with chronic lymphocytic leukaemia (CLL) For the cost-effectiveness analysis the two main strategies modelled were assay-driven care versus unexploited sensitivity (patients assessed as having some sensitivity to drugs but not to those administered)	Prospective cohort of 178 CLL diagnosed patients undergoing DiSC assay provides the main source of data on survival and prognostic variables The DiSC assay was not used to guide treatment. Patients categorised as follows: 1. resistant to any drug(s) tested or administered 2. sensitive to some drug(s) tested but not to those administered 3. sensitive to drug(s) given Other than the DiSC assay per se, treatment costs are assumed not to differ between patients modelled as receiving and not receiving DiSC assay-guided treatment. Survival is the modelled outcome	The DiSC assay can be used to identify the lowest concentration of a drug required to eradicate 90% of tumour cells <i>in vitro</i> (LC ₉₀). Patients with a drug LC ₉₀ value less than or equal to a predefined threshold are considered to have useful drug sensitivity Based on the cohort of 178 patients, a Cox proportional hazards model was used to predict the probability of surviving out to 3.3 years post-treatment. Independent variables comprised age, gender, stage and treatment sensitivity status Cost-effectiveness results for eight prognostic groups are presented	Overall incremental cost per life-year gained (ICER) of using DiSC assay to guide treatment for CLL patients is estimated at £1470 For different prognostic groups: Cancer: Sex Stage Age ICER M A, B 50 £1150 M A, B 70 £1850 M C 50 £860 M C 70 £1520 F A, B 50 £4640 F A, B 70 £7040 F C 50 £3140 F C 70 £4980	Potential for results to change both in favour and against DiSC assay-guided treatment when assumptions relating to time horizon, improvements in survival and cost are altered DiSC assay-guided treatment for patients with CLL appears to be cost-effective but should be more widely evaluated, perhaps within the setting of an RCT
Hamel et al., 1997 ²⁶² (USA) (Provider) (Lifetime) (US \$ 1994) (3% cost and QALY)	Cost-utility analysis comparing dialysis and continued aggressive care with a 'do nothing' approach for seriously ill hospitalised adults with renal failure requiring dialysis	Prospective cohort of 490 seriously ill patients receiving dialysis for renal failure provides the main source of data on resource use, utilities, survival and prognostic variables. Life expectancy beyond 6 months is extrapolated using an exponential function Assumption that 'do nothing' will result in almost immediate death and has zero associated costs and QALYs	Prognostic model developed on larger patient cohort using Cox proportional hazards method, recalibrated for 490 patients in study. Model used to predict probability of surviving 6 months post-treatment. Independent variables include age, diagnosis, comorbidity and 11 physiological variables. Patients are stratified into 5 groups and incremental cost per QALY calculated	Overall incremental cost per QALY gained (ICUR) of initiating dialysis and continuing aggressive care for seriously ill adults is \$128,200 For different prognostic groups: Proportion surviving 6 months (%) ICUR (\$) 0–10 274,100 11–20 122,600 21–30 85,800 31–40 80,000 41–60 61,900	Results did not change substantially in sensitivity analysis Dialysis for seriously ill patients with renal failure and an average prognosis is not cost-effective when compared with other interventions

In an attempt to reduce low-yield biopsies in prostate cancer detection, Littrup and colleagues²⁶³ used decision analysis to evaluate the cost-effectiveness of using various combinations of prognostic markers for screening, biopsy and follow-up decisions. Main screening strategies compared were use of prostate-specific antigen (PSA) (a marker for prostate cancer risk) with or without digital rectal examination (DRE), versus age-adjusted PSA with or without DRE. Following an abnormal screen, universal versus cancer risk tailored biopsy strategies were compared. Results were presented for each strategy using average cost-effectiveness ratios, calculations of which are unclear. No incremental analyses were presented.

In a 1999 publication entitled 'The DiSC assay', Mason²⁶⁴ modelled the likely impact on costs and survival of using assay guided chemotherapy for different patients with acute lymphocytic leukaemia. *In vitro* testing of tumour cells using a differential staining cytotoxicity (DiSC) assay can be used to determine a patient's sensitivity to chemotherapy drugs, thus allowing the selection of therapies more likely to be effective. Unexploited drug sensitivity was modelled as a comparator intervention and the incremental cost per LYG of assay-guided therapy was estimated for different prognostic groups. Average incremental cost-effectiveness was estimated to be £1470 (price base 1998–9).

Intensive care therapy

Glance and colleagues²⁶¹ examined the cost-effectiveness of continuing intensive care treatment for patients predicted, using a prognostic model, to have a >90% risk of death 48 hours after admission to the unit. Decision trees incorporating Markov nodes were used to model provision versus withdrawal of care on the basis of this prognostic scoring system. Discarding the prognostic model and continuing care for cases predicted to be medically futile resulted in slightly higher survival to hospital discharge (87.2 versus 86.85%) at a slightly higher cost (\$30,100 versus \$29,200). The incremental cost per death prevented was estimated at \$263,700 (price base unknown), but the results were shown to be extremely sensitive to small changes in the specificity of the scoring system. The authors concluded that unless it could be proven that the scoring system would retain the same predictive power outside the database on which it was developed, the use of such a system would not be a cost-effective use of scarce healthcare resources.

Hamel and colleagues²⁶² used a prognostic model to group seriously ill hospitalised patients with renal

failure according to predicted 6-month survival and then estimated the cost-utility of initiating dialysis and aggressive care for each group. The relevant policy alternative modelled was to withhold dialysis. Costs and QALYs of this policy were set to zero. Estimated across all patients, the average incremental cost per QALY of initiating dialysis for such patients was \$128,200 (price base 1994). The incremental per QALY across the five prognostic groups modelled did not fall below \$61,900.

Study quality

The quality of these five studies was assessed using an established checklist for health economics papers²⁶⁵ (Table 27). None of the studies conformed to all relevant requirements stipulated in the checklist. However, flaws were generally confined to presentational rather than methodological issues. Only one study failed to undertake an incremental analysis and was unclear about methods used to calculate average cost-effectiveness ratios.

Other studies

Of the remaining five studies, two were considered to be partial evaluations in that costs and outcomes were presented for prognosis-driven patient management, but not for a relevant comparator.^{266,267} A further study, by Tonnaire and colleagues,²⁶⁸ examined the cost-effectiveness of using various techniques to identify prognostic factors for acute leukaemia but did not model the cost-effectiveness of using such prognostic information to guide therapeutic decisions and treatment. The potential of ECG-based prognostic models to identify groups of patients with acute myocardial infarction and unstable angina for whom intensive therapy may be beneficial was examined by Kent and colleagues.²⁶⁹ Although the possible economic consequences were discussed, the cost-effectiveness of using such prognostic information to inform patient management was not explicitly modelled. Finally, in formulating guidelines pertaining to the use of six colorectal cancer markers and seven breast cancer markers for screening, treatment and surveillance, ASCO searched the literature for evidence on the cost-effectiveness of marker-guided cancer care.²⁷⁰ Literature reviews across all 13 markers appear to reveal only one published cost-effectiveness study,²⁷¹ modelling the cost-utility of using carcinoembryonic antigen (CEA) to monitor patients following definitive surgery for colorectal cancer. Accordingly, it is assumed that no other published cost-effectiveness data were available.

Details of each of these five studies are contained in Table 28.

TABLE 27 Quality assessment of five identified economic evaluations

Study	Checklist questions																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
Calvert et al., 2003 ²⁶⁰	Yes	Yes	Yes	No	Yes	No	Yes	No	NA	NA	Yes	Yes	Yes	NA	No	No	Yes	Yes	
Littrup et al., 1994 ²⁶³	Yes	Yes	No	Yes	Yes	No	Yes	Yes	NA	NA	Yes	NA	NA	NA	No	No	No	No	
Mason et al., 1999 ²⁶⁴	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	NA	NA	Yes	NA	NA	NA	No	Yes	No	Yes	
Glance et al., 1998 ²⁶¹	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	NA	NA	Yes	NA	NA	NA	No	No	Yes	No	
Hamel et al., 1997 ²⁶²	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	NA	NA	Yes	Yes	Yes	NA	No	Yes	Yes	Yes	
Study	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	35	
Calvert et al., 2003 ²⁶⁰	Yes	Yes	Yes	Yes	Yes	No	NA	NA	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Littrup et al., 1994 ²⁶³	No	Yes	Yes	Yes	NA	NA	No	NA	No	No	No	Yes	No	Yes	Yes	Yes	Yes	Yes	
Mason et al., 1999 ²⁶⁴	NA	Yes	Yes	Yes	Yes	No	NA	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Glance et al., 1998 ²⁶¹	No	Yes	Yes	No	NA	NA	No	NA	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Hamel et al., 1997 ²⁶²	Yes	Yes	Yes	Yes	Yes	No	NA	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	
NA, not applicable.																			
Study design																			
1. The research question is stated.																			
2. The economic importance of the research question is stated.																			
3. The viewpoint(s) of the analysis are clearly stated and justified.																			
4. The rationale for choosing the alternative programmes or interventions compared is stated.																			
5. The alternatives being compared are clearly described																			
6. The form of economic evaluation used is stated.																			
7. The choice of form of economic evaluation is justified in relation to the questions addressed.																			
Data collection																			
8. The source(s) of effectiveness estimates used are stated.																			
9. Details of the design and results of effectiveness study are given (if based on a single study).																			
10. Details of the method of synthesis or meta-analysis of estimates are given (if based on an overview of a number of effectiveness studies).																			
11. The primary outcome measure(s) for the economic evaluation are clearly stated.																			
12. Methods to value health states and other benefits are stated.																			
13. Details of the subjects from whom valuations were obtained are given.																			
14. Productivity changes (if included) are reported separately.																			
15. The relevance of productivity changes to the study question is discussed.																			
16. Quantities of resources are reported separately from their unit costs.																			
17. Methods for the estimation of quantities and unit costs are described.																			
18. Currency and price data are recorded.																			
19. Details of price adjustments for inflation or currency conversion are given.																			
20. Details of any model used are given.																			
21. The choice of model used and the key parameters on which it is based are justified.																			
Analysis and interpretation of results																			
22. Time horizon of costs and benefits is stated.																			
23. The discount rate(s) is stated.																			
24. The choice of rate(s) is justified.																			
25. An explanation is given if costs or benefits are not discounted.																			
26. Details of statistical tests and confidence intervals are given for stochastic data.																			
27. The approach to sensitivity analysis is given.																			
28. The choice of variables for sensitivity analysis is justified.																			
29. The ranges over which the variables are varied are stated.																			
30. Relevant alternatives are compared.																			
31. Incremental analysis is reported.																			
32. Major outcomes are presented in a disaggregated as well as aggregated form.																			
33. The answer to the study question is given.																			
34. Conclusions follow from the data reported.																			
35. Conclusions are accompanied by the appropriate caveats.																			

TABLE 28 Details of the five partial evaluations identified by filtering search 2

Study	Study aim ^a	Methods	Role of prognostic information	Results	Conclusions
Mikhail <i>et al.</i> , 1997 ²⁶⁶	To examine the cost-effectiveness of mandatory stress testing within a dedicated chest pain centre (CPC) for management of low to moderate risk patients presenting at hospital with acute chest pain Comparator intervention was identified as routine hospital inpatient evaluation, however only the costs of such care were estimated	Prospective observational study of 502 chest pain patients transferred from hospital to a dedicated CPC. Patients without a diagnosis of end-stage coronary artery disease and for whom acute myocardial infarction (AMI) was ruled out were scheduled for CPC evaluation with further stress testing Costs and final diagnoses were obtained directly from the patient cohort. A retrospective cohort of patients was used to calculate the cost of conventional inpatient evaluation	CPC-based stress testing was used to diagnose patients with ischaemic heart disease (IHD) who, despite not having had an AMI, are at risk of further coronary events and have a 6–24-month prognosis similar to that of patients surviving AMI	Only average cost-effectiveness ratios for stress testing were available 24 patients admitted to hospital from the CPC were diagnosed with IHD. The total cost of stress testing was estimated to be \$75,000 Cost of stress testing per IHD case detected \$3125 Cost of stress testing per patient undergoing PTCA or CABG \$10,714 Simple calculation assuming that 16% of the 24 IHD diagnosed patients would have incurred zero costs and died without CPC stress testing produces an incremental cost per year of life saved of \$1502	Paper concludes that stress testing is safe, reliable and cost-effective for the evaluation of low to moderate risk patients in dedicated chest pain centres
Orr <i>et al.</i> , 1999 ²⁶⁷	To determine the cost-effectiveness, for women with advanced ovarian cancer, of directing chemotherapy using an <i>in vitro</i> drug resistance assay No relevant comparator intervention was included in the study	Prospective observational study. <i>In vitro</i> assays for drug resistance were carried out on a cohort of 66 women with advanced ovarian cancer who had undergone cytoreductive surgery and were used to select appropriate chemotherapy regimens Primary end-point for the study was 3-year survival. Secondary end-point was reported to be the cost-effectiveness of using an <i>in vitro</i> assay to select appropriate chemotherapy regimens for women	<i>In vitro</i> assays used to ascertain resistance to chemotherapy agents and thus to match women with drugs more likely to be effective in improving survival	Only average cost-effectiveness of implementing all assay-guided chemotherapy is reported and is estimated to be \$9768 per patient surviving 3 years Comparisons are made with average cost-effectiveness ratios (calculated minus the assay cost) for patients receiving assay-guided platinum/paclitaxel (TP) and platinum/cyclophosphamide (CP) chemotherapy	Paper concludes that by taking into account costs avoided by reducing ineffective treatments, needless toxicity and loss of quality of life, the estimated cost-effectiveness of assay-guided treatment relative would improve when compared with conventional treatment decisions

continued

TABLE 28 Details of the five partial evaluations identified by filtering search 2 (cont'd)

Study	Study aim ^a	Methods	Role of prognostic information	Results	Conclusions
Tonnaire et al., 1998 ²⁶⁸	To undertake a cost-effectiveness analysis comparing cytogenetic and molecular analysis techniques for detecting prognostic markers (i.e. chromosomal abnormalities) in acute lymphoid leukaemia (ALL) and acute myeloid leukemia (AML)	Consecutive series of 107 patients with acute leukaemia (22 ALL, 85 AML) were tested for chromosomal abnormalities using both conventional cytogenetic and molecular biological (PCR) techniques. Six different diagnostic strategies using single and different combinations of the two tests were modelled	Genetic anomalies are prognostic markers for patients with acute leukaemia. They can be used to select poor prognosis patients for aggressive therapy or to avoid intensive therapy and associated side-effects for patients with good prognoses	For each strategy, cost, anomalies detected and average cost per anomaly detected are reported (US \$1995)	Paper concludes that when detecting genetic anomalies for use as prognostic markers in acute leukaemia, PCR alone is the cost-effective strategy for ALL, whereas for AML, cytogenetic examination should be used in isolation
	Neither technique identified as the relevant comparator	Direct test costs and effects (rate of detection of clinically relevant true positive anomalies) were estimated for each strategy and each leukaemia type	For ALL, the following anomalies were tested for using cytogenetic and PCR techniques: translocations t(9;22)(q34;q11), t(4;11)(q21;q23) and t(1;19)(q23;p13)	For ALL, all six strategies detected the same number of clinically relevant anomalies (i.e. seven). The cost-effective alternative was therefore the least costly strategy (i.e. PCR testing alone at \$11,815)	It is acknowledged by the authors that the value of such prognostic information lies in its potential to inform patient management decisions
	No incremental analyses were performed		For AML anomalies tested for were translocations t(15;17), t(8;21), inv(16), del(5q)/-5, del(7q)/-7 and trisomy 8	For AML, the strategy concluded to be economically efficient was that with the lowest average cost-effectiveness ratio (cytogenetic examination alone at \$1888 per anomaly detected)	
Kent et al., 2000 ²⁶⁹	To determine how CG-based prognostic models can be used to identify patients who are likely and unlikely to benefit from intensive treatment	Paper talks about heterogeneity amongst patients suffering acute myocardial infarction (AMI) with respect to mortality risk and potential to benefit from treatment	Variables included in the previously published model used to predict mortality with and without thrombolytic therapy include age, systolic blood pressure, history of diabetes, heart rate, AMI size, AMI location, right bundle branch block, thrombolytic therapy, time from symptom onset to ECG	Thrombolytic therapy – most likely to benefit in terms of reduced mortality are patients classified as high mortality risk who present soon after MI. This is further emphasised when treatment side effects are considered	Conclusion that treatment of high-risk patients is more likely to be effective and cost-effective applies to both types of therapies examined
	To examine the potential benefits for different patients of reperfusion therapy for AMI, a validated ECG-based prognostic model predicting mortality with and without thrombolytic therapy was applied to 921 patients prescribed thrombolytic drugs at 28 hospitals across the USA	Potential benefits of administering antithrombotic therapy to unstable angina/non-q-wave MI patients was also examined using prognostic models	No details provided on prognostic models developed to predict adjusted mortality risk and risk of MI for patients with unstable angina/non-q-wave MI	Antithrombotic therapy – figure presented to show that a 10% RR reduction (typical of that demonstrated in trials) is much more beneficial for high-risk than for low-risk patients	

continued

TABLE 28 Details of the five partial evaluations identified by filtering search 2 (cont'd)

Study	Study aim ^a	Methods	Role of prognostic information	Results	Conclusions
ASCO Bast et al., 1996 ²⁷⁰	To formulate clinical practice guidelines relating to the use of six colorectal cancer markers and seven breast cancer markers in the prevention, screening, treatment and surveillance of each disease	Guidelines were formulated primarily using evidence gathered from the published literature. Reviews were carried out and information relating to the impact of such markers on five main outcomes was collated: <ol style="list-style-type: none"> overall survival disease-free survival quality of life lesser toxicity cost-effectiveness In the absence of relevant published literature, guidelines were based on expert clinical opinion	Where evidence permitted, the following prognostic markers were examined for evidence on each of the five main outcomes, in the areas of prevention, screening, treatment and surveillance. <p>Colorectal cancer</p> <ol style="list-style-type: none"> Serum CEA Serum lipid-associated sialic acid (LASA) Serum CA 19-9 DNA flow cytometric ploidy (DNA Index) DNA flow cytometric proliferation index (% S phase) p53 tumour suppressor gene ras oncogene <p>Breast cancer</p> <ol style="list-style-type: none"> Serum CA 15-3 Serum CEA ER and PR DNA flow cytometric ploidy (DNA Index) DNA flow cytometric proliferation index (% S phase) p53 tumour suppressor gene c-erbB2 Cathepsin D 	Focusing upon cost-effectiveness evidence only Across all prognostic markers, only one published cost-utility analysis appears to have been identified. The study in question by Kievit and van de Velde ²⁷¹ ref to be changed modelled the cost effectiveness of serial CEA monitoring following surgical resection for colorectal cancer. The comparator modelled was diagnosis of recurrence at symptom presentation Cost-effectiveness results ranged from \$34,688 per life-year saved for a young man with Stage CI cancer, to \$210,333 per life-year saved for a 75-year-old	CEA should be measured preoperatively for colorectal cancer patients if likely to change surgical management. It should also be monitored every 2–3 months for 2 or more years if resection of liver metastasis is indicated. For all remaining colorectal cancer markers, a lack of published data precluded the formulation of recommendations concerning routine use ER and PR should be measured on every primary breast cancer specimen. For all remaining breast cancer markers, a lack of published data precluded the formulation of recommendations concerning routine use

^a Studies described as cost-effectiveness analyses are inappropriately titled. They do not conform to the conventional definition of economic evaluation as published by Drummond and Jefferson.²⁶⁵

Conclusions

Despite the substantial number of hits returned by literature search 2 for question E, closer inspection of such papers reveals very few formal studies examining the cost-effectiveness of using prognostic information to select appropriate treatment or care for patients. The 0.2% (5/2441) of papers that were identified were of varying

quality and highlight the fact that economic evaluation in this area appears still to be in its infancy. At a time when there is much clinical interest in the use of prognostic markers to determine a patient's capacity to benefit from treatment, every effort should be made to assess the cost-effectiveness and the clinical effectiveness of using such information in routine practice.

Chapter 9

Modelling the cost-effectiveness of using prognostic information in the adjuvant treatment of breast cancer

Introduction

Although cost-effectiveness analysis is fundamentally concerned with making choices about which patients or groups of patients should be offered which interventions, little explicit attention has been paid by health economists to the use of PIs in making treatment choices, although these potentially provide a reasonable quantitative prediction of health outcome and therefore are likely to be associated with cost-effectiveness. As Chapter 8 reported, a structured review of published cost-effectiveness studies of PIs identified only five such studies in any therapeutic or disease area. PIs raise at least two issues from an economic perspective: first, what are the incremental costs and effects of making treatment choices for patients or groups of patients on the basis of a PI, compared with some other decision rule such as treating all patients?; and second, given that individual prognostic factors or composite PIs have varying accuracy and that resources may be involved in using them, can a cost-effectiveness framework provide a method of determining how much is worth spending in acquiring prognostic information?

In this chapter, we set out a **framework** for estimating the effectiveness and cost-effectiveness (from an NHS perspective) of adjuvant therapy for individuals or groups of patients with specified prognostic characteristics. In the light of findings from the systematic reviews and the survey presented in previous chapters, construction of a definitive model was considered unfeasible for a number of reasons.

Firstly, as Chapters 4 and 6 revealed, studies systematically reviewing the prognostic and predictive ability of individual factors were generally of poor quality, failed to use individual patient data and did not undertake meta-analysis. Specification of a definitive set of prognostic markers was therefore difficult.

Second, and as Chapter 5 explained, a requirement of a survival (or prognostic) model is that it combines and estimates the collective impact on prognosis of factors which individually are significant. Furthermore, for such techniques to be useful in a clinical setting, that is, to enable a clinician to estimate survival prospects for individual patients, and hence the likely effectiveness afforded by administering adjuvant therapy, it has been noted that the baseline hazard function from such a prognostic model is required. The review of prognostic models presented in Chapter 5 again raised concerns about the quality of study methodology and reporting. Very few variables were shown to feature in more than one or two of the models identified, and with the exception of the NPI¹⁹³ (based on tumour stage, grade and size), few models have been subject to independent validation.

Whether prognosis prediction is maximised by using the NPI is unclear: several researchers have claimed that the addition of other variables such as ER and PR indicators to the index could improve its prognostic ability. The survey presented in Chapter 7 also revealed that other variables are often used in the clinical decision-making setting to supplement the NPI. As these 'NPI variant models' have not yet been externally validated, the NPI *per se*, which was generated using the Cox proportional hazards technique, remains the preferred model with which to estimate the relative survival of different groups of patients.

Prognosis prediction for individual patients using the NPI, however, is problematic. As noted in Chapter 5, the baseline hazard function from a Cox model, which is required for patient-level prognosis, cannot be simply specified, and this probably explains why for the NPI it has never been reported. Increasingly, new web-based prediction tools such as Adjuvant! (www.adjuvantonline.com) (which uses tumour size, stage, grade, ER status, age and co-morbidities to

predict 10-year survival for a patient with and without therapy) are being used by clinicians. For a decision model aiming to estimate the impact of adjuvant treatment on the lifetime survival or quality-adjusted survival of individual patients with differing prognostic characteristics, however, the NPI is not useful, and the time horizon covered by Adjuvant! is just 10 years. Parametric survival models offer a solution to these problems; however, as Chapter 5 revealed, no clinically accepted and validated parametric prognostic models exist in the area of early breast cancer.

Given the difficulties involved in identifying a definitive set of prognostic variables, and the absence of a prognostic model easily capable of facilitating patient-level prognosis, construction of a definitive cost-effectiveness model based on existing published prognostic evidence was not possible. A separate and substantial exercise to develop and validate a parametric model is clearly needed, but equally clearly would be unfeasible within the time and financial constraints of this research project. Nevertheless, given the obvious benefits to be gained by establishing prognosis and treatment effectiveness for individual patients, we considered it important to develop and report a method for incorporating patient-level predictions from a parametric survival model alongside data on costs and quality of life, all within a decision analytic framework.

Such data could then be used to illustrate the way in which patient-level gains in survival and quality-adjusted survival from administering adjuvant therapy could be estimated. In addition, such an approach would illustrate the way in which the incremental cost per LYG and per QALY gained could be calculated for various prognosis-based decision rules that could be used to guide adjuvant treatment decisions for individuals or groups of patients with specified prognostic characteristics. The general approach presented here should be applicable to other types of prognostic information, interventions and disease areas.

Our analysis is based on a patient-level data set that was obtained from the Medical Oncology Unit at the Churchill Hospital, Oxford, with information on patient characteristics, treatments and outcomes over a minimum 5-year and maximum 15-year follow-up period. All of the modelling and analysis reported in this chapter make use of these data. We present three sections detailing different stages of our analysis. The first describes the data and details the process of developing a prognostic index – referred to here as the Oxford Prognostic

Index (OPI) – based on a parametric survival model for the data. The second section describes how a model can be created around the OPI to describe the full lifetime experience of a woman with breast cancer in terms of both costs and (quality-adjusted) survival. Using this model, the third section describes how alternative decision algorithms for choosing how to treat women with breast cancer could be evaluated, including the commonly employed NPI.

We begin by describing the nature of the data set obtained from the Medical Oncology Unit, Churchill Hospital, Oxford. We then go on to describe the development of the OPI based on the linear predictor of an exponential survival model for breast cancer death.

The patient data set

A patient-level data set was obtained from the Medical Oncology Unit at the Churchill Hospital, Oxford, containing details of breast cancer patients treated at this centre from 1986 onwards. Patients were selected who had operable breast cancer and who had at least 5 years of follow-up information subsequent to initial surgery. Patients with ductal or lobular carcinomas *in situ* were omitted from the analysis, in line with previous prognostic modelling in breast cancer.^{193,272} This gave an initial total of 1174 patients potentially available for analysis, but with some items of information missing for some patients, as shown in *Table 29*.

The proportion of missing values was low in almost all variables except ER levels, which were not available for 342 (29.1%) of the 1174 cases treated mainly in the earlier years of the dataset. Given the potential prognostic importance of this variable (patients whose cancers have ER-positive receptors tend to have a better prognosis than patients whose cancers do not have these receptors, but only if treated with tamoxifen) and given that cancers with ER-negative receptors are also more likely to respond to chemotherapy, it was decided to impute a value for ER level where it was missing, but to drop all other cases where at least one variable had a missing value. This gave a final data set of 1058 observations. The imputation process is described below, followed by descriptive information on the final dataset.

Imputation of ER status

ER levels are measured in a number of ways, but commonly as a continuous measure of

TABLE 29 Initial variable list and missing information

Variable	Valid observations	No. (%) missing
Patient identifier	1174	0 (0)
Date of birth	1173	1 (0.1)
Age in years at operation	1173	1 (0.1)
Date of first operation	1174	0 (0)
ER level	832	242 (29.1)
Grade (ductal) using the Bloom and Richardson system ¹⁹⁵	1173	1 (0.1)
Tumour size	1117	57 (5.1)
Number of nodes sampled	1165	9 (0.8)
Number of nodes positive	1145	29 (2.5)
Adjuvant chemotherapy used	1167	7 (0.6)
Adjuvant radiotherapy used	1170	4 (0.3)
Adjuvant hormone therapy used	1168	6 (0.5)
Relapse	1174	0 (0)
Time to death or censored	1171	3 (0.3)

TABLE 30 A generalised linear model for predicting ER level (gamma distribution family with log link)

Explanatory variable	Coefficient	SE	p-Value
Age in years at operation	0.0273	0.0065	0.0000
Tumour size	0.1468	0.0558	0.0090
Tumour stage	0.2155	0.1107	0.0520
Tumour grade	-0.2679	0.0857	0.0020
Adjuvant chemotherapy used	-0.8759	0.1867	0.0000
Adjuvant hormone therapy used	0.4362	0.1480	0.0030
Adjuvant radiotherapy used	-0.1124	0.1576	0.4760
Logged time to death or censored	0.5718	0.1144	0.0000
Did patient die?	0.5156	0.2192	0.0190
Year of initial operation	0.0742	0.0350	0.0340
Cancer screen detected	0.1956	0.1415	0.1670
Relapse experienced	-0.1857	0.1667	0.2650
Constant	-150.1005	70.0740	0.0320

ferntomoles of ER per milligram of cytosol protein or by immunohistochemistry grading on an eight-point scale. It is then commonly dichotomised around a cut-off value to give an indication of ER positivity. For the imputation process it was considered appropriate to predict a continuous ER value and then to apply the usual procedure for turning this into a dichotomous variable (this would also permit later examination of the cut-off point if necessary). The available data showed that ER status followed an extremely skewed distribution ranging from 0 to 742 with a mean [standard deviation (SD)] of 70 (111) and a median (interquartile range) of 24 (5–80). For this reason, consideration was given to developing a prediction model for a transformation of the ER variable.

However, the need to transform predictions back to the original scale is problematic, since straightforward back-transformation of predictions

on the transformed scale would be biased. For this reason, the imputation modelling was performed within the class of generalised linear models where the assumption is that it is the expected values rather than the data themselves that are transformed, hence the issue of bias in back-transformation is avoided. Generalised linear models represent an extremely flexible class of models whereby different transformation link functions can be combined with different distribution families. We compared a number of different possible models before deciding on a gamma distribution family with a log link function as the basis of the imputation equation. *Table 30* reports the generalised linear model used to impute ER value. On the basis of this model, predicted values of ER status were generated for all patients in the dataset. Where the ER level of the patient was missing, the predicted ER level from the generalised linear model in *Table 30* was employed instead.

Employing this imputation process effectively increased the sample size for the subsequent prognostic modelling by 27%. Ideally, multiple imputation should be performed to incorporate uncertainty in the prediction process into the final results. This was not undertaken for this analysis for the sake of simplicity in presenting the general framework. Furthermore, since ER status is just one explanatory variable in the overall prognostic model, it is not expected that uncertainty in the prediction process will have major consequences for the prognostic model results.

Descriptive information on dataset

Table 31 shows the distribution of patients by year of initial operation. The earliest year for which information was available was 1986, and no women treated after 1996 were included to ensure that at least 5 years of follow-up information were available. About 50% of the sample was treated in the years 1993–6.

The mean age of women at initial treatment was 56.6 years, with a range from 25 to 90 years. Table 32 shows the frequency distribution of women by age.

As noted above, it was necessary to impute ER values for 285 (27%) of the 1058 patient records. For the observed ER values, the mean was 70 (SD 111), with a range from 0 to 742. The mean value of imputed ER values was 63 (SD 43), with a range from 3 to 217. Figure 6 shows the cumulative percentage of women by ER level (observed and imputed). A total of 291 (28%) of women in the sample were below the cut-off point of 10 fmol/mg frequently used to define women as ER negative

TABLE 31 Calendar year in which primary treatment occurred

Year	No. of patients	%
1986	32	3.0
1987	53	5.0
1988	48	4.5
1989	67	6.3
1990	81	7.7
1991	89	8.4
1992	132	12.5
1993	143	13.5
1994	135	12.8
1995	145	13.7
1996	133	12.6
Total	1058	100.0

TABLE 32 Age groups of patients at first treatment

Age group (years)	No of patients	%
<35	30	2.8
≥ 35 and <45	155	14.7
≥ 45 and <55	311	29.4
≥ 55 and <65	276	26.1
≥ 65 and <75	231	21.8
75+	55	5.2
Total	1058	100.0

or ER poor, and it is evident from the distribution of values that small changes in the cut-off point will have a large impact on the proportion categorised as ER positive.

Table 33 shows the distribution of patients by grade (ductal), using the Bloom and Richardson system.¹⁹⁵ About 20% of women were in the good,

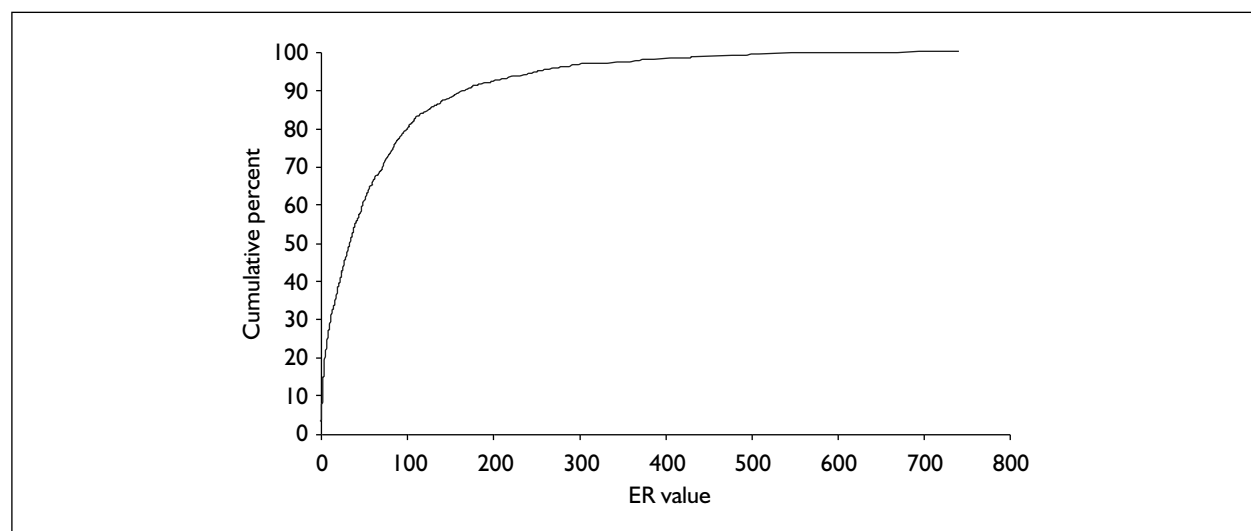


FIGURE 6 Cumulative percentage of women by ER value

TABLE 33 Distribution of women by grade of tumour

Grade	No. of patients	%
1 – Good	210	19.8
2 – Moderate	484	45.7
3 – Poor	364	34.4
Total	1058	100.0

TABLE 34 Frequency distribution by tumour size (cm)

Size (cm)	No. of patients	%
< 1	67	6.3
≥ 1 and < 2	367	34.7
≥ 2 and < 3	337	31.9
≥ 3 and < 4	167	15.8
≥ 4 and < 5	56	5.3
5+	64	6.0
Total	1058	100.0

TABLE 35 Frequency distribution by number of nodes positive

No. of nodes	No. of patients	%
0	625	59.1
1	151	14.3
2	92	8.7
3	57	5.4
4	36	3.4
5	20	1.9
6	22	2.1
7	11	1.0
8	11	1.0
9	6	0.6
10+	27	2.7
Total	1058	100.0

46% in the moderate and 34% in the poor category. The mean tumour size was 2.4 cm, with a range from 0.1 to 11 cm. *Table 34* shows tumour size arranged by category.

Table 35 shows the degree of nodal involvement in the sample. A total of 625 (59%) of women were node negative, 300 (28%) had 1–3 nodes positive and 133 (13%) had ≥ 4 nodes positive.

Figure 7 shows the number of women in the sample receiving different adjuvant therapies. Of the 1019 patients receiving some form of adjuvant therapy, 273 (27%) received adjuvant chemotherapy alone or in combination, 885 (87%) received adjuvant radiotherapy alone or in combination and 781 (77%) received adjuvant hormone therapy alone or in combination.

A total of 312 (29%) of patients were recorded as having a relapse during the period of follow-up. Mean time to relapse was 3.24 years and the median time to relapse was 2.34 years. A total of 311 patients died during the follow-up period, with a mean time to death of 4.16 years and a median time to death of 3.56 years. The mean follow-up time to death or censoring was 7.63 years.

The original dataset recorded cause of death as ‘progressive disease’, ‘other’ or ‘not known’. Of the 47 recorded as ‘not known’, 29 were known to have had a relapse and 18 were recorded as not having a relapse. The cause of death for the 29 who had a recorded relapse was therefore reclassified as ‘progressive disease’ and the cause

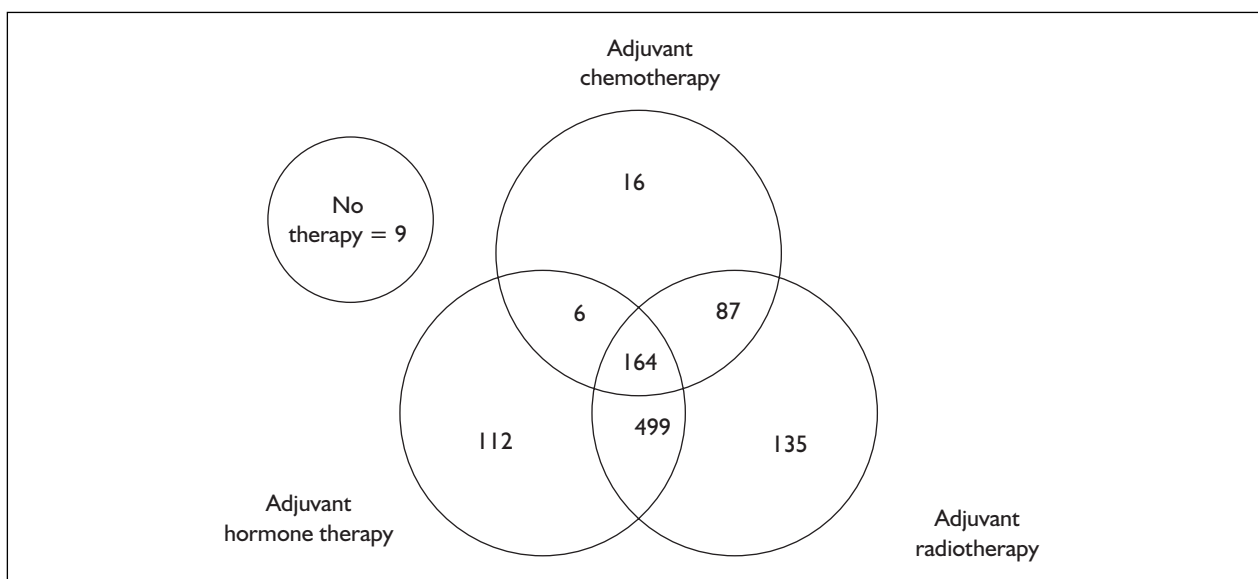


FIGURE 7 Number of women receiving adjuvant chemotherapy, radiotherapy and hormone therapy

TABLE 36 Cross-tabulation of relapse and death

		Died from breast cancer?		Total
		No	Yes	
Relapse?	No	746 (100%)	0 (0%)	746 (100%)
	Yes	100 (32%)	212 (68%)	312 (100%)
Total		846 (80%)	212 (20%)	1058 (100%)

of death for the 18 who had no recorded relapse was reclassified as ‘other’. Following this, 212 of the 311 deaths in the dataset were in the ‘progressive disease’ category and 99 were in the ‘other’ category.

Table 36 shows a cross-tabulation of all patients who experienced a relapse and all patients who died from breast cancer. A total of 100 (32%) of the 312 patients who had a relapse did not die during the follow-up period. On the basis of the information recorded on each patient, it was possible to calculate the NPI for each patient using the standard formula provided by Haybittle and colleagues:¹⁹³ $0.2 \times \text{tumour size in centimetres plus tumour grade plus tumour stage}$ (where 1 = no nodal involvement, 2 = 1–3 nodes positive and 3 = 4 or more nodes positive).

Figure 8 shows the frequency distribution of NPI scores. The mean value of the NPI was 4.12 and the median was 4.26. The overall pattern is similar to that reported by Todd and colleagues,¹⁹⁶ with the size component primarily producing a distribution within each integer value of the index. Using the most commonly used cut-off points for the NPI (<3.4 = Good, 3.4–5.4 = Moderate, >5.4 = Poor), 367 (35%) of patients in the sample fell into the Good category, 520 (49%) into the Moderate category and 171 (16%) into the Poor category. Figure 9 shows Kaplan–Meier survival curves for each prognostic group.

Estimating a survival model: the OPI

The NPI is formulated using an adapted linear predictor from a Cox proportional hazards survival model.¹⁹³ The possibility was explored of using parametric survival analysis on the patient dataset described above to create a PI that would allow the illustration of the use of patient level prognostic information in a cost-effectiveness framework. This index has been called the Oxford Prognostic Index (OPI) to prevent confusion with the validated NPI.

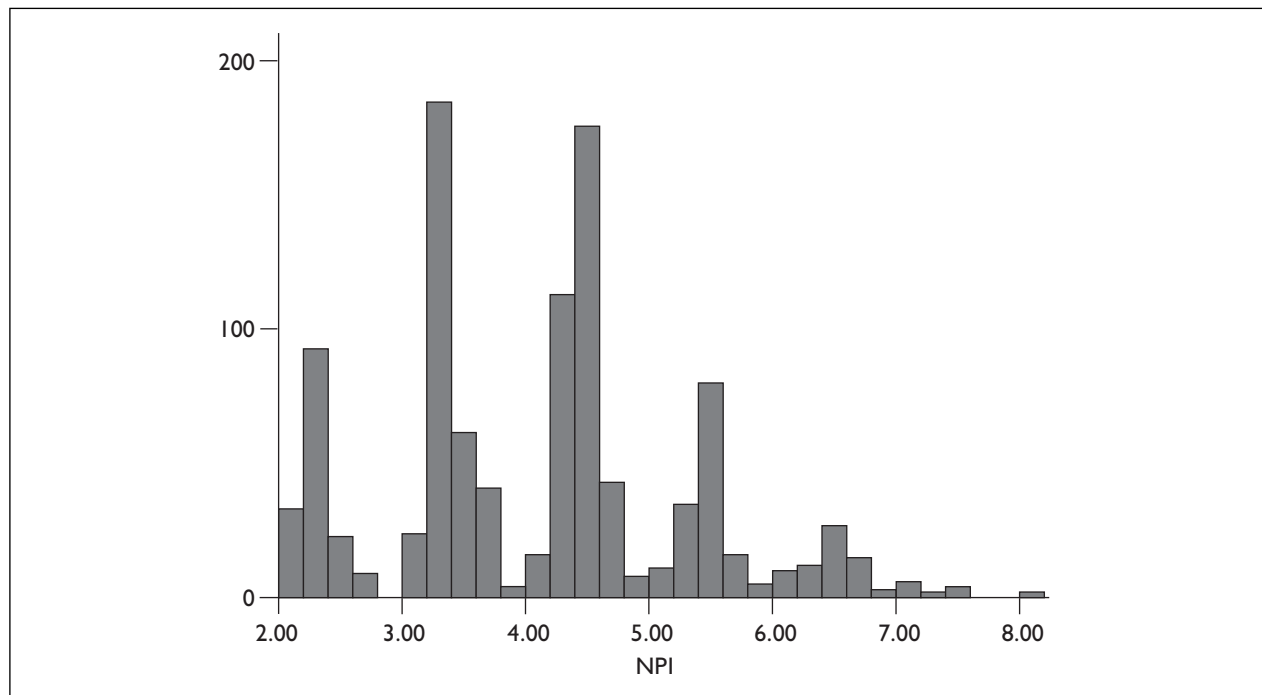


FIGURE 8 Frequency distribution of NPI scores

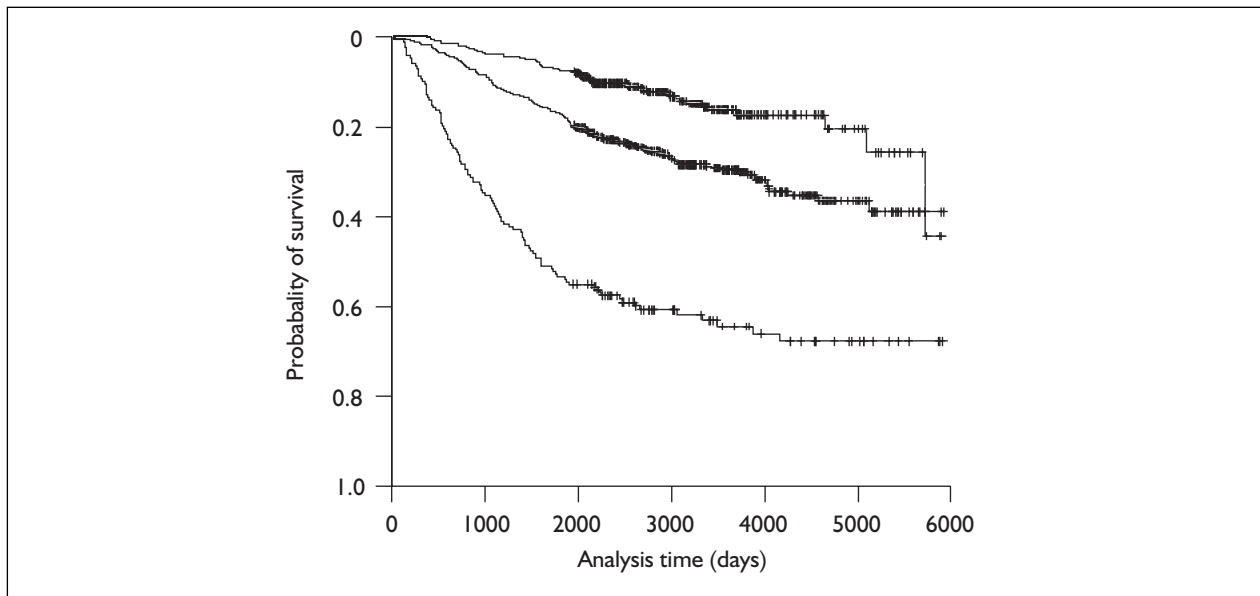


FIGURE 9 Kaplan–Meier survival curves by NPI grouping. NPI categories: top line, Good; middle line, Moderate; bottom line, Poor prognosis. +, Censored observations.

Survival analysis was employed using patients' characteristics described in the section above as potential prognostic factors for time to breast cancer death in this patient group. Although non-parametric methods are commonly employed in medical statistics, the survival function was parameterised since the survival analysis would form the basis of the economic model (see below), which requires the baseline hazard function to be known. Three potential parametric models for survival time were considered, exponential, Weibull and Gompertz, which are capable of modelling constant, monotonically increasing or monotonically decreasing hazard functions. Other common functional forms were not considered, such as log-normal or log logistic, since the lack of monotonicity can sometimes lead to unrealistic

hazard functions when extrapolated. Both the Weibull and Gompertz models nest the exponential distribution as a special case and tests revealed that within the data there was no evidence for moving away from a simple exponential model of constant hazard. The final model that we selected is reproduced in *Table 37*.

Note that dummy variables for treatment with chemotherapy and hormone treatment were included. The purpose of this is not to estimate treatment effects, since it is well known that observational data such as these suffer selection biases in this regard. Rather, the purpose was to adjust for treatment in order to generate a predicted survival in the absence of treatment (achieved by setting the treatment dummy

TABLE 37 Coefficients estimated from survival analysis using a simple exponential model of constant hazard

Explanatory variable	Coefficient	SE	p-Value
Age at operation	0.031791	0.0060357	<0.001
Tumour size	0.185175	0.0369811	<0.001
Tumour Stage II	0.577657	0.1500816	<0.001
Tumour Stage III	1.361642	0.1860969	<0.001
Tumour Grade II	0.379655	0.225727	0.093
Tumour Grade III	0.933166	0.2287583	<0.001
Adjuvant chemotherapy used	-0.10918	0.1808838	0.546
Adjuvant hormone therapy used	-0.20215	0.2235965	0.366
Year of operation 1996	-0.92151	0.3670285	0.012
ER status	-0.44553	0.2348257	0.058
Predicted ER	-0.0857	0.2691269	0.75
Constant	-11.55	0.486934	<0.001

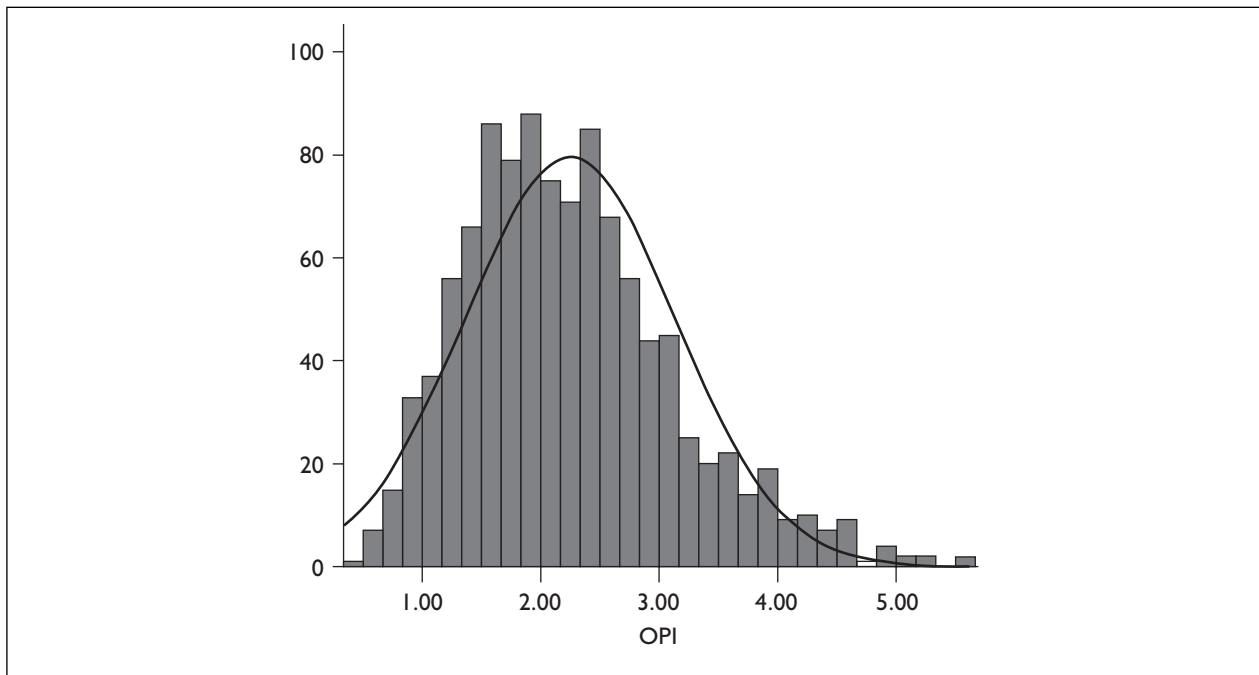


FIGURE 10 Frequency distribution of OPI scores

variables to zero). In order to generate the OPI, we first calculated the linear predictor (the cross product of the coefficients in *Table 37* with the prognostic factors) for each patient in the dataset, first setting the treatment dummy variables to zero. This generated a variable that ranged from -11.5 to -6.4 , to which 12 was added in order to generate a PI on the range 0.5–5.6, where 0.5 represents a good and 5.6 a poor prognosis. The distribution of the OPI, for our patient data set, is shown in *Figure 10* and can be compared to the distribution of the NPI in *Figure 8*.

It should be clear from the description above that the OPI has a direct relationship with the linear predictor in an exponential survival analysis. Nevertheless, it is instructive to consider grouping the OPI into prognostic groups and presenting the survival curves for such groups. To determine the index scores that best discriminate on the basis of prognosis, various ways of classifying women into prognostic groups were examined. Issues considered when constructing groups were loss of information about differences between individuals,²⁷³ degree of separation between Kaplan–Meier survival curves and the number of groups considered practical for use in a clinical setting.

Table 38 shows the loss of information (L) attributable to grouping women in various ways according to their OPI score. Here L is calculated

as the weighted average of the variance across the chosen groups divided by the variance of the sample as a whole. Grouping ‘like’ individuals together results in smaller within-group variances, thus retaining the information about differences between women. When women are ungrouped, $L = 1$ and there is said to be a complete loss of information about differences between individuals.²⁷³

Table 38 shows that least information is lost with five groupings constructed using normal distribution percentages. In *Figure 11*, showing the separation between the Kaplan–Meier survival curves for the five groups, it is clear, however, that

TABLE 38 Loss of information from grouping

Method of grouping	L
Whole sample	1
Equally sized groups:	
Two	0.386829
Three	0.231677
Four	0.158414
Five	0.116141
Groups using normal distribution percentages:	
Three groups 27, 46, 27%	0.206799
Four groups 16.4, 33.6, 33.6, 16.4%	0.124592
Five groups 10.9, 23.7, 30.7, 23.7, 10.9%	0.061408
Integer split:	
Six groups $\leq 1, \leq 2, \leq 3, \leq 4, \leq 5, > 5$	0.093138

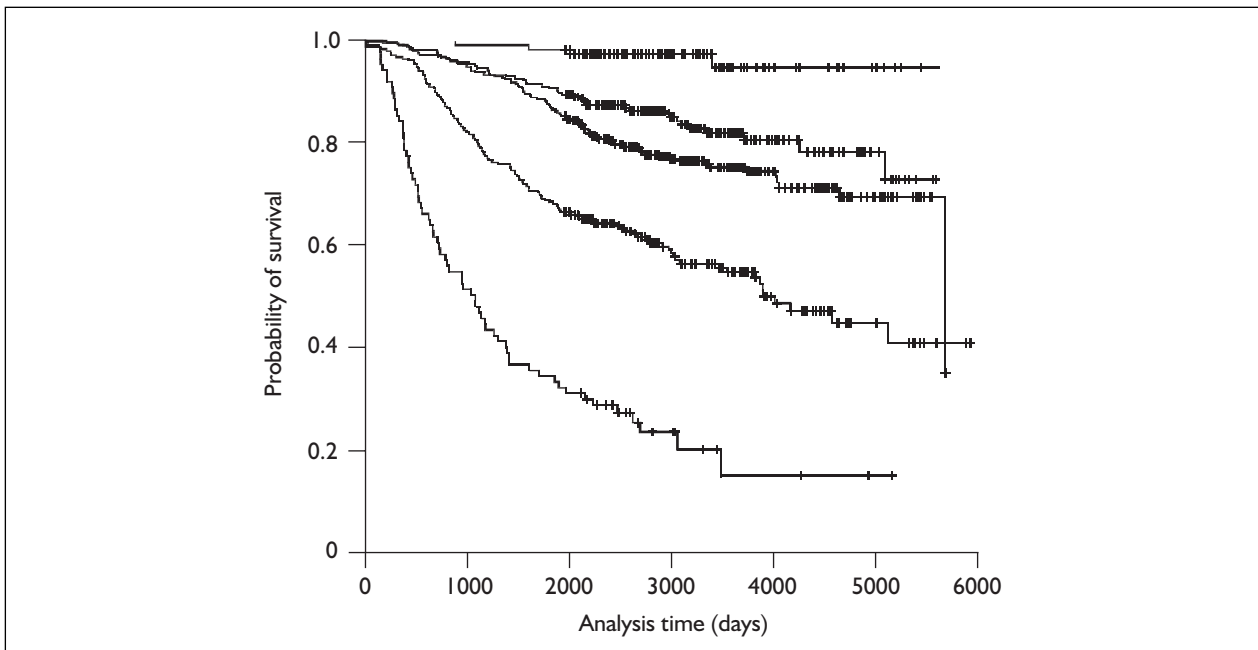


FIGURE 11 Kaplan–Meier survival curves for five groups based on the normal distribution percentages. OPI categories: top line, lowest 10.9% scores; second top line, middle lower 23.7% scores; middle line, middle 30.7% scores; second bottom line, middle upper 23.7% scores; bottom line, highest 10.9% of scores. +, Censored observations.

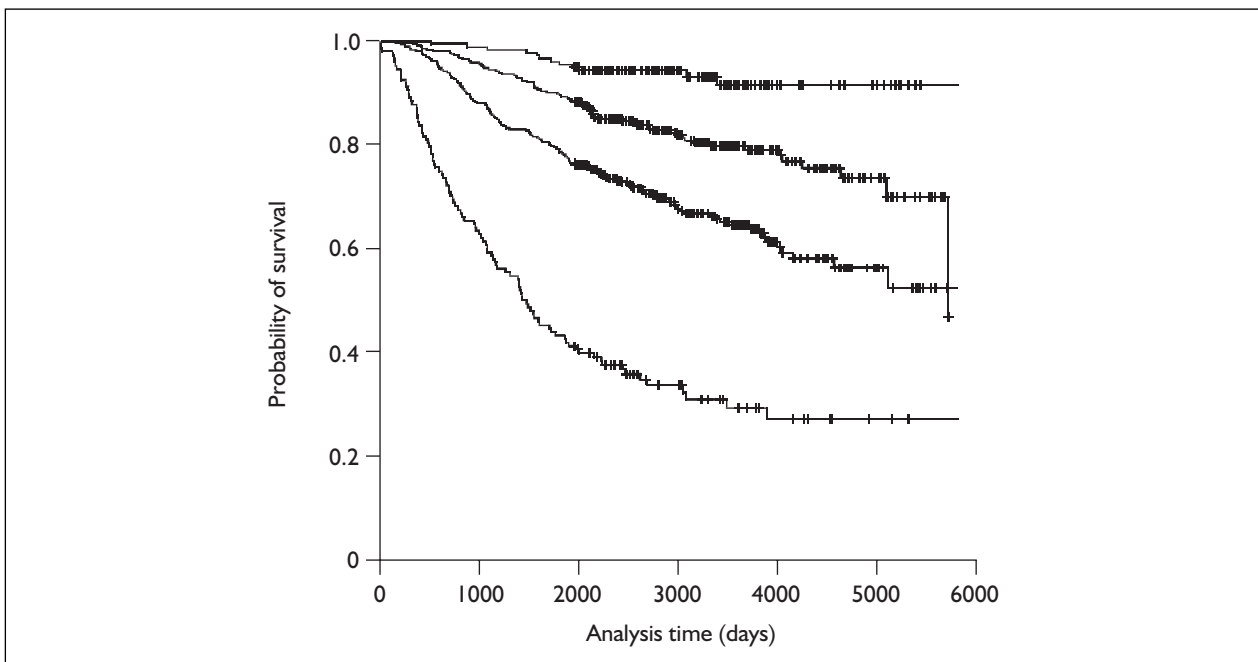


FIGURE 12 Kaplan–Meier survival curves for four groups based on the normal distribution percentages. OPI categories: top line, lowest 16.4% scores; middle top line, middle lower 33.6% scores; middle bottom line, middle upper 33.6% scores; bottom line, highest 16.4% of scores. +, Censored observations.

the separation based on such groupings is poor between the middle 30.7% and lower middle 23.7% of OPI scores. Four groups based upon normal distribution percentages give a better discrimination (see *Figure 12*) and this is thought to be a practical number of prognostic

groups for clinicians to work with. Such groupings are therefore used from this point onwards in the analysis of data and presentation of results. The associated index scores defining the boundaries between each group are shown in *Table 39*.

TABLE 39 OPI values and prognostic groups

Index value	Prognosis
OPI < 1.37	Excellent
1.37 ≤ OPI < 2.15	Good
2.15 ≤ OPI < 3.05	Moderate
OPI > 3.05	Poor

Figure 12 should be interpreted with caution – these survival curves are indicative of the survival experience for members of each group. However, the best estimate of (expected) survival remains the survival curve based on the linear predictor (or equivalently the OPI) for each individual patient.

A model for the cost and (quality-adjusted) survival of breast cancer patients

By combining methodologies used in determining prognosis with those used in health economic evaluation, it is possible to illustrate an approach for simulating the effectiveness and cost-effectiveness associated with the decision to treat individual women or groups of women with different prognostic characteristics.

Estimating overall survival

The survival analysis for time to breast cancer death formed the basis of the cost-effectiveness model. In order to estimate overall survival, the breast cancer survival model was combined with estimates of non-breast cancer death rates obtained from standard life tables.²⁷⁴ Although the survival analysis performed in the preceding section found no evidence of time dependence in the hazard function, the mean follow-up of

7.63 years is relatively modest and it is reasonable to consider that the risk attenuates for patients having survived 10 years free from breast cancer recurrence. To capture this and avoid discontinuity in the overall survival curve, the predicted breast cancer death hazard was reduced between years 11 and 15 by a factor of 1 (full effect) to 0 (no effect) in increments of 0.2. In other words, breast cancer risk is assumed to be the full risk estimated by the survival model for years 1–10, after which it attenuates in steps of 20% of the full risk until by year 15 there is assumed to be only the background risk that exists in the general population.

Estimates of treatment effect

To ensure that treatment effects could be examined within the model, indicator variables for adjuvant chemotherapy and adjuvant hormone therapy were included when estimating the exponential survival model (as described in the preceding section), and these were set to zero when the linear predictor was calculated in order to describe the baseline prognosis of patients in the absence of treatment. The effect of deciding to give any patient chemotherapy or hormone therapy in the model was then estimated from a systematic review and meta-analysis undertaken by the Early Breast Cancer Trialists’ Collaborative Group.^{12,13} Table 40 reports the values used in the model: in summary, it is assumed that hormone treatment has no effect on ER-negative women and has an RR of 0.72 for women who are ER positive. For chemotherapy, the RR reduction is related to age, with an RR of 0.73 for those aged less than 50 years and 0.92 for those over 60 years.

Adding in cost and quality adjustments

Table 41 summarises the parameter values used in the cost-effectiveness model. Given the

TABLE 40 Effect of adjuvant chemotherapy and hormone treatment on the RR of breast cancer death

Group	Hormone treatment	Control	RR of treatment compared with control
ER negative	0.408	0.374	1
ER positive	0.221	0.28	0.72
	Chemotherapy treatment	Control	RR of treatment compared with control
Age <50 years	0.323	0.394	0.73
Age 50–59 years	0.352	0.385	0.86
Age >60 years	0.357	0.38	0.92

Source: EBCTCG, 1998.

TABLE 41 Cost, utility and probability parameters used in cost-effectiveness model, with sources of information

Variable	Value	Description	Source
Discounting:			
cDR	0.03	Cost discount rate	UK Treasury
oDR	0.03	Outcome discount rate	UK Treasury
Treatment cost:			
hormC	136.26	Hormone treatment cost	BNF 44. Cost based on 20-mg daily dose/duration 5 years
chemoC	2000	Adjuvant chemotherapy cost	Clinical opinion/Karnon and Brown ²⁷⁵
Side-effect probabilities:			
hormSEp	0.05	Probability of side-effects from hormone Tx	Clinical opinion/Karnon and Brown ²⁷⁵
chemoSEp	0.3	Probability of side-effects from chemotherapy Tx	Clinical opinion/Karnon and Brown ²⁷⁵
Side-effect utility decrements:			
hormSEud	0.20	Utility decrement associated with having side-effects from hormone Tx	Weighted average of Grade III/IV toxicity and major toxicity utility decrements from Karnon and Brown ²⁷⁵
chemoSEud	0.17	Utility decrement associated with having side-effects from chemotherapy Tx	Weighted average of Grade III/IV toxicity and major toxicity utility decrements from Karnon and Brown ²⁷⁵
Side-effect treatment costs:			
hormSEc	987	Cost of treating side-effects of hormone Tx	Weighted average of tamoxifen Grade III/IV toxicity and major toxicity costs from Karnon and Brown ²⁷⁵
chemoSEc	462	Cost of treating side-effects of chemotherapy Tx	Weighted average of tamoxifen plus chemotherapy Grade III/IV toxicity and major toxicity costs from Karnon and Brown ²⁷⁵
Relapse:			
relapseUD	0.35	Utility decrement associated with breast cancer relapse	Assumes locoregional relapse from Karnon and Brown ²⁷⁵
relapseCost	10520	Cost associated with breast cancer relapse	Cocquyt <i>et al.</i> ²⁷⁶
Tx, treatment.			

illustrative nature of the framework, a single source was used to inform the input values for most parameters. Health service costs associated with treatments, namely adjuvant chemotherapy (such as CMF) and adjuvant hormone therapy (tamoxifen), were included. The probabilities of experiencing side-effects with these treatments, and of consequent costs and utility decrements associated with those side-effects, were obtained from the published literature. Within the model, it was assumed that each death from breast cancer was preceded by a breast cancer recurrence with associated quality of life decrements and costs. For simplicity, non-fatal breast cancer recurrences were not separately modelled. The cost of these (fatal) relapses together with the utility decrements experienced by the patients were also obtained from previously published studies. All costs are expressed in 2002 prices and, along with outcomes, were discounted at an annual rate of 3%.

Illustrative predictions from the cost-effectiveness model

Using the modelling approach set out above, it was then possible to estimate the lifetime costs and outcomes (life expectancy and quality-adjusted life expectancy) of each patient in the dataset, in the presence or absence of adjuvant chemotherapy treatment or hormone treatment. For simplicity, the main analyses focused on the chemotherapy decision and it was assumed that all ER-positive patients and no ER-negative patients received hormone treatment.

Figure 13 shows the way in which the model predicts survival following breast cancer surgery for three hypothetical women: (a) a 45-year-old woman with a tumour of size 1 cm, Stage I and Grade I, who is ER positive (NPI good prognosis category with a score of 2.2), (b) a 45-year-old woman with a tumour of size 2 cm, Stage II and

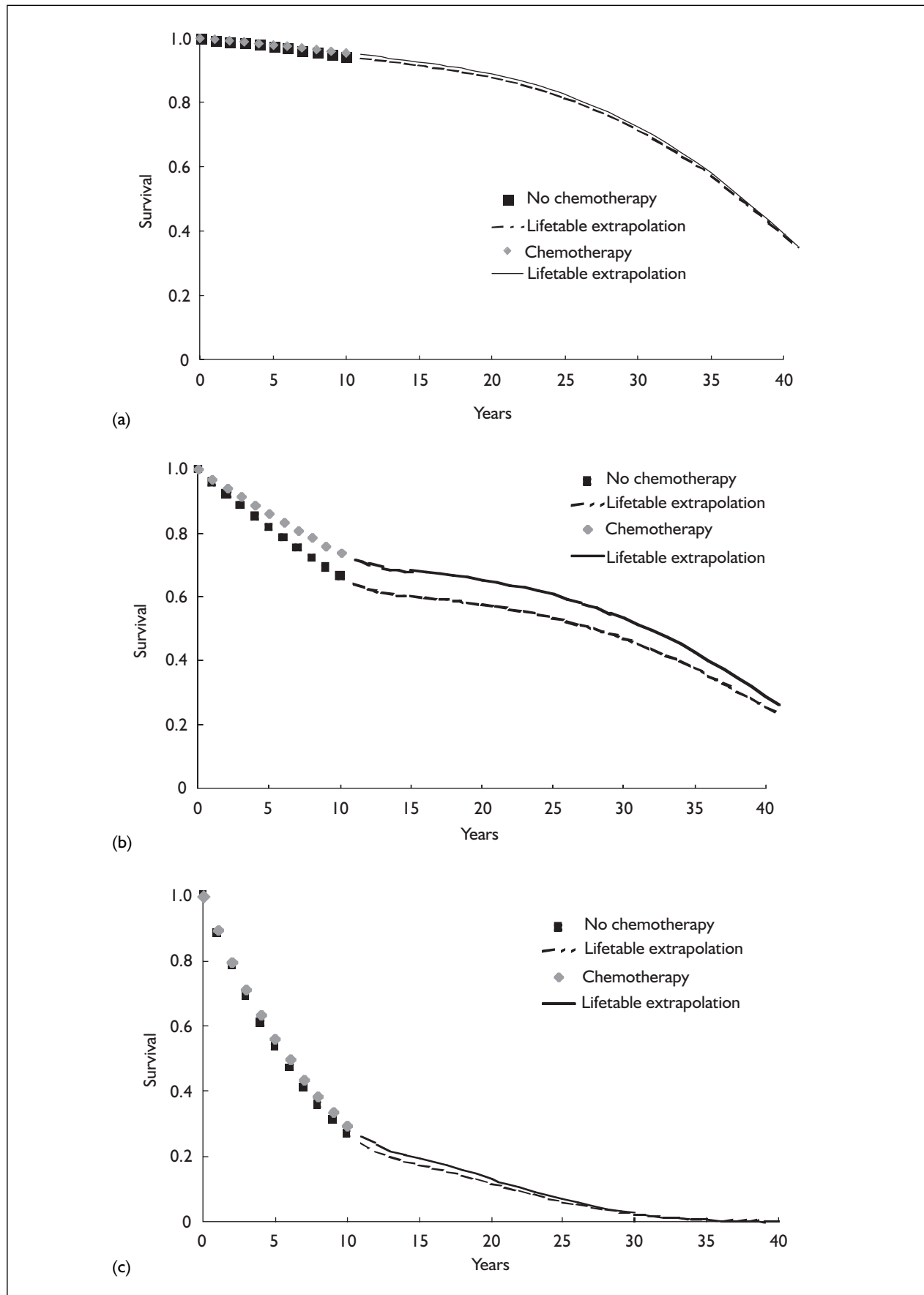


FIGURE 13 Predicted survival following breast cancer surgery assuming treatment or no treatment with adjuvant chemotherapy

Grade III, who is ER negative (NPI moderate prognosis category with a score of 5.4) and (c) a 65-year-old woman with a tumour of size 3 cm, Stage III and Grade II, who is ER negative (NPI poor prognosis category with a score of 5.6).

In *Figure 13(a)*, the woman's characteristics indicate excellent prognosis, and this is evident in the figure: the 10-year probability of survival is 0.94 if adjuvant chemotherapy is not given and 0.95 if it is given, and quality-adjusted life expectancy (undiscounted) is 26.62 years if no adjuvant chemotherapy is given and 26.81 years if adjuvant chemotherapy is given; the (discounted) cost-effectiveness of treatment compared with no treatment is £24,059 per QALY gained. In *Figure 13(b)*, the woman's characteristics indicate moderate prognosis: the 10-year survival probability is 0.66 if adjuvant chemotherapy is not given and 0.74 if it is given; quality-adjusted life expectancy (undiscounted) is 18.85 years if no adjuvant chemotherapy is given and 20.82 years if adjuvant chemotherapy is given; the (discounted) cost-effectiveness of treatment compared with no treatment is £1229 per QALY gained. In *Figure 13(c)*, the woman's characteristics indicate poor prognosis: the 10-year survival probability is 0.27 if adjuvant chemotherapy is not given and 0.30 if it is given; quality-adjusted life expectancy (undiscounted) is 5.58 years if no adjuvant chemotherapy is given and 5.88 years if adjuvant chemotherapy is given, and the (discounted) cost-effectiveness of treatment compared with no treatment is £8768 per QALY gained.

Presentation of results – the risk table approach

Given the survival model estimated, there were 648 possible combinations of prognostic variables for which the effectiveness and cost-effectiveness of adjuvant therapy could be modelled. A risk or look-up table format similar to those generated from Framingham was used to present the results. Each cell of the table was used to present the modelled effectiveness (in terms of quality-adjusted survival) of adjuvant chemotherapy for a woman with a specific combination of prognostic variables, the idea being that in addition to viewing survival graphs such as those above, clinicians could also consult such tables to gain an estimate of the likely overall health consequences of adjuvant therapy for any presenting woman. For ease of interpretation, cells within the tables presenting the effectiveness results were shaded according to whether adjuvant chemotherapy appeared effective (QALYs gained were positive) or not. *Figure 14* provides an example of a risk table.

In this instance, it is evident that for women aged 85 years or over, the benefit of adjuvant chemotherapy is outweighed by the side-effects of the treatment.

Patient selection using different prognostic criteria: costs and outcomes

A further reason for developing the model described in the preceding section was to provide a basis for comparing alternative criteria for deciding how patients are treated with respect to adjuvant chemotherapy. The review of unit

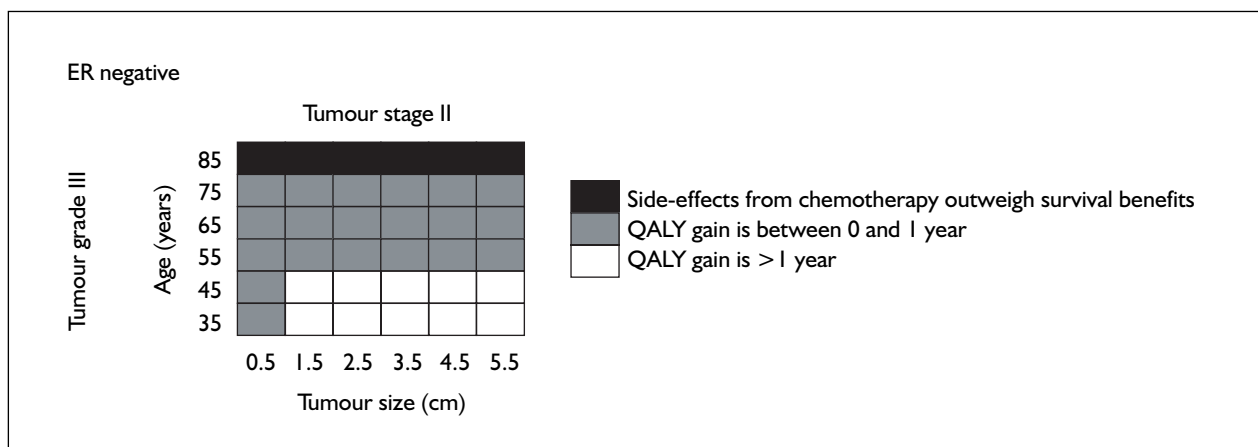


FIGURE 14 Presentation of discounted QALY results using a risk table format

protocols presented in Chapter 7 has already demonstrated that UK breast care units make differential use of the same set of prognostic factors when selecting patients to undergo adjuvant systemic therapy. This section describes the predicted impact of alternative selection criteria on the overall costs and outcomes, using the Oxford dataset to illustrate the method used.

Potential decision criteria

A series of possible decision criteria were identified that could potentially be employed to decide whether patients would be selected to receive or not receive adjuvant chemotherapy:

- The current (2002) treatment protocol of the Medical Oncology Unit at the Churchill Hospital, Oxford.
- Cut-off values using the NPI.
- Cut-off values using the OPI.
- A net health gain rule, whereby treatment is offered to a patient if the model predicts a gain in quality-adjusted life expectancy (that is, survival benefit outweighs disutilities from side-effects and/or recurrence).
- A cost-utility rule, whereby treatment is offered to a patient if the model predicts that the cost per QALY gained of treatment over no treatment is less than £30,000.²⁷⁷
- Actual treatment choice.
- Treat all patients.
- Treat no patients.

Once patients had been allocated to treatment or no treatment under each decision rule, total costs and effects were summed across all patients and the results of each strategy were plotted on a cost-effectiveness plane. The current protocol was taken as the comparator to calculate incremental cost-effectiveness results. Sensitivity, specificity and positive and negative predictive values for each

strategy were also calculated, taking the cost-effectiveness approach as the 'gold standard'.

Predicted costs and outcomes from applying the decision criteria

Table 42 shows descriptive results of applying different prognostic criteria when deciding which women to select for adjuvant chemotherapy. In practice, 26% of women in the sample were given adjuvant chemotherapy, and the estimated mean discounted quality-adjusted life expectancy of all patients based on this selection criterion was 11.12 years. If the current protocol had been applied to all patients in the database, 35% would have been given adjuvant chemotherapy with a slight increase in overall quality-adjusted life expectancy. Using a cost-effectiveness ceiling of £30,000 per QALY gained to select patients, 636 would have been given adjuvant chemotherapy, whereas treating if there was any anticipated net gain in quality-adjusted life expectancy would result in 91% of patients being given adjuvant chemotherapy. Setting the NPI at a threshold of 4.4, 39% of patients would have been selected for treatment, and setting the OPI at a threshold of 2.5 would result in 34% of patients being treated.

Table 43 and Figure 15 show these data in terms of incremental cost-effectiveness ratio (ICER), with the current protocol set as the comparator.

Shown on the *x*-axis in Figure 15 is the difference between alternative decision criteria and the current protocol in terms of the total number of QALYs generated by the patient cohort. On the *y*-axis is the difference in total costs. In comparison with the current protocol (which is placed at the origin), using the NPI with a cut-off value of 4.4 or the OPI at a cut-off value of 2.5 would give less health benefit at higher cost, whereas using the actual treatment decisions taken is estimated to

TABLE 42 Average costs and effects per patient of different prognostic criteria (sample = 1058 patients)

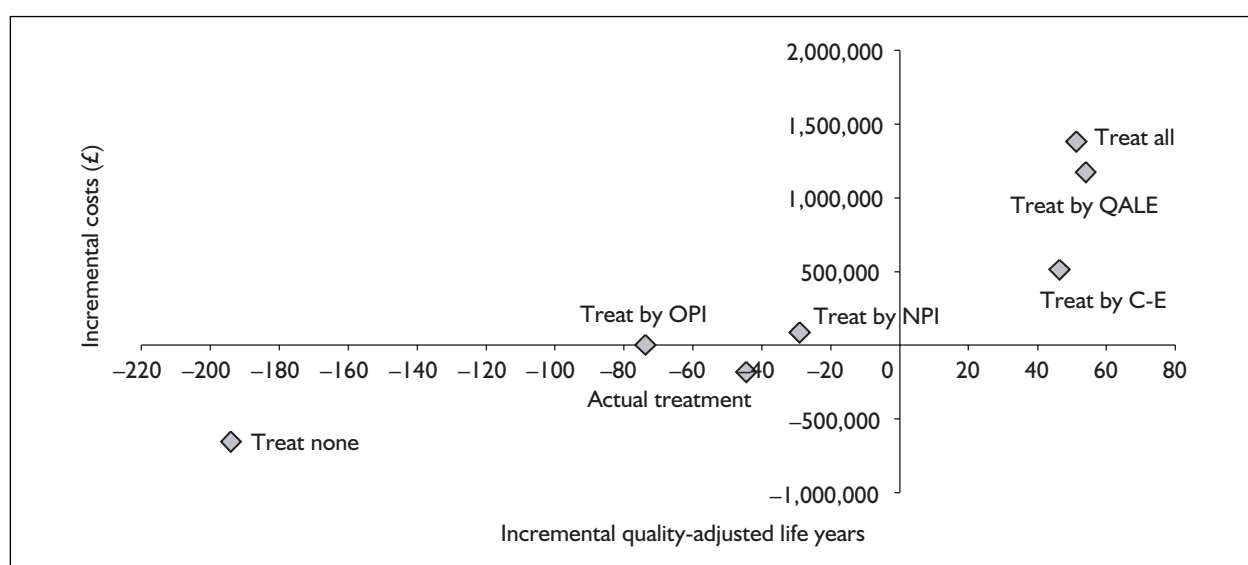
	No. (%) of patients selected for treatment	Average discounted quality-adjusted life expectancy (years)	Average costs (£)
Treat if NPI >4.4	411 (39)	11.13	2952
Treat if OPI >2.5	359 (34)	11.09	2873
Treat if QALE >0	958 (91)	11.21	3979
Treat if C-E <£30,000	636 (60)	11.20	3355
Treat all	1058 (100)	11.21	4176
Treat none	0 (0)	10.97	2254
Actual treatment	273 (26)	11.12	2700
Current protocol	371 (35)	11.16	2864

C-E, cost-effectiveness ceiling; QALE, quality-adjusted life expectancy.

TABLE 43 Incremental costs, effects and cost-effectiveness of different prognostic criteria compared with current protocol

	Incremental effect (QALYs)	Incremental cost (£)	ICER (£)
Treat by NPI	-29.31	93,058	-
Treat by OPI	-74.13	9,459	-
Treat by QALE	53.66	1,178,671	21,965
Treat by C-E	46.06	518,811	11,265
Treat all	51.03	1,387,932	27,197
Treat none	-194.31	-646,307	-
Actual treatment	-44.75	-173,750	-

C-E, cost-effectiveness ceiling; QALE, quality-adjusted life expectancy.

**FIGURE 15** Incremental costs and effects of different prognostic methods compared with current protocol on the cost-effectiveness plane

give less health benefit but at lower cost. Treating everyone would have given more health benefit than the current protocol at an ICER of over £27,000 per QALY gained. Treating only if a net gain in quality-adjusted life expectancy is anticipated gives the greatest health benefit of all options considered, at an ICER of £22,000 per QALY gained. Finally, making adjuvant chemotherapy decisions on the basis of anticipated cost-effectiveness, with a ceiling of £30,000 per QALY gained, gives more health benefit than the current protocol and has a relatively low ICER of just over £11,000 per QALY gained.

Variations in cut-off values used to make treatment selections

Figure 16 shows the effect of altering the threshold values of the NPI between 3 and 5, the OPI between 1 and 3.5 and cost-effectiveness ceiling between £0 and £40,000 per QALY when deciding who to offer adjuvant chemotherapy, and of

increasing the threshold for net quality-adjusted life expectancy gain from 0 to 0.5 (i.e. treat only if net gain in health is at least 0.5 QALYs). In each case the incremental costs and effects are plotted on the cost-effectiveness plane against the comparator of the current protocol.

The OPI gives a net gain in health outcome compared with the current protocol when the threshold value is lowered to approximately 1.85 from the baseline of 2.5. The NPI gives a net gain in health outcome compared with the current protocol when the threshold value for treatment is reduced from the baseline value of 4.4 to 4.2 or lower, but the incremental cost-effectiveness is high. When the cost-utility ceiling is reduced to £9300 per QALY it becomes cost-saving relative to the current protocol, but below £7250 per QALY it gives less health benefit. Altering the quality-adjusted life expectancy gain threshold traces the same points as the cost per QALY sensitivity: when

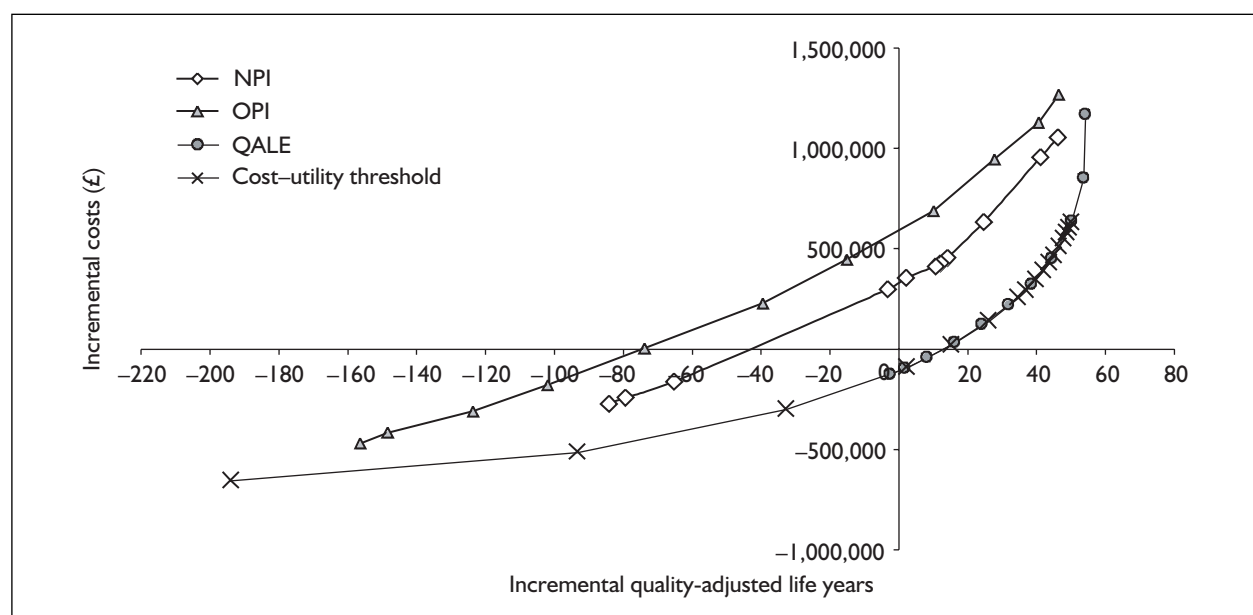


FIGURE 16 Sensitivity analysis of incremental costs and effects of different prognostic methods compared with current protocol when threshold/ceiling values are changed

the decision rule is to treat as long as any health gain is obtained, this is equivalent to an infinite willingness to pay. Treating only when at least 0.5 QALYs are expected to be gained is equivalent to setting a cost-effectiveness threshold at approximately £6750 per QALY gained.

Predictive performance of different criteria

Table 44 reports the sensitivity and specificity of different treatment selection criteria compared with cost-effectiveness criteria, using which (at a ceiling value of £30,000 per QALY gained) a total of 636 patients would have been selected for treatment. The actual treatment decisions made for these patients resulted in the lowest number being selected for chemotherapy: 273 of the 1058 patients. Both actual treatment decisions and the current Oxford protocol had a high specificity and high positive predictive value. Using the net improvement in quality-adjusted life expectancy

criteria (equivalent to infinite willingness to pay for any health gain) resulted in the highest number of patients being selected for treatment: 958.

Further research

Given the apparent uncertainty as to which patients should and should not receive adjuvant chemotherapy in the UK, we plan to develop the decision analytic approach illustrated here to assess more formally the cost-effectiveness of using these and other UK breast cancer protocols to select women for such treatment. For this to be achievable, several developments are required in order to move the method presented here from its present function as an illustrative framework to that of a definitive model. First, the parametric survival model requires external validation using other datasets. Second, refinements to the structure of the decision model will be carried out,

TABLE 44 Sensitivity, specificity and positive (PPV) and negative predictive value (NPV) of different treatment selection criteria compared with cost-effectiveness criteria

	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	No. treated
NPI	49.1	76.5	75.9	49.9	411
OPI	36.8	70.4	65.2	42.5	359
QALE	100.0	23.7	66.4	100.0	958
Actual	41.4	97.6	96.3	52.5	273
Current protocol	57.7	99.1	98.9	60.8	371

QALE, quality-adjusted life expectancy.

for example to take a more realistic account of relapse – at present it is simply assumed that relapse only occurs prior to a breast cancer death. Third, the model does not currently allow for the costs of obtaining the prognostic information for use in decision-making. Costs associated with measurement of ER status, assessment of lymph node involvement and tumour grade and size measurement therefore need to be included. Finally, the model will be developed so as to adhere to well-established guidelines for decision analytic modelling; for example, data from systematic reviews will be used to inform parameter estimates and uncertainty surrounding these estimates will be examined and reported using probabilistic sensitivity analysis.

Conclusions

In this chapter, a framework or methodological approach has been presented for incorporating prognostic information within a decision analytic model. It has been demonstrated that output from such a framework may be useful at two different decision-making levels. First, for the clinician, the approach facilitates patient-level prognosis prediction with and without adjuvant therapy, and second, for the policy maker, it permits exploration of the cost-effectiveness of using alternative prognosis-based decision criteria to select women for treatment.

It is expected that in developing a definitive model we will produce a useful tool for use by both of these parties, in a field where research into

and use of prognostic information are increasing. Modification of the framework to incorporate new prognostic variables or accommodate changes to routine practice over time (for example, the use of new hormonal therapies) would be relatively straightforward.

Conference presentations

Versions of the work presented in this chapter have been presented at both UK and European conferences:

Campbell HE, Gray AM, Briggs AH, Harris AL. Cost-effectiveness of using different prognostic information and decision criteria to select women with early breast cancer for adjuvant systemic therapy. Presented at The Health Economists' Study Group, Canterbury, Kent, July 2003.

Campbell HE, Briggs AH, Gray AM, Altman D, Harris AL. A framework for modelling and presenting the cost-effectiveness of adjuvant systemic therapy for women with early breast cancer. Awarded best rated abstract at the EORTC European Conference on the Economics of Cancer, Brussels, September 2003.

Campbell HE, Briggs AH, Gray AM, Harris AL. A framework for modelling and presenting the cost-effectiveness of adjuvant systemic therapy for women with early breast cancer. Presented at the Breast Cancer Research Group Meeting, Institute of Molecular Medicine, John Radcliffe Hospital, Oxford, February 2004.

Chapter 10

Summary, discussion and conclusions

Each of the preceding chapters can to a certain extent stand alone with a description of the goals, methods, results and conclusions. This chapter acts as a summary of the individual systematic reviews presented in the preceding chapters and cross-links information and draws final conclusions and implications for clinical practice and research.

Although this report addresses a specific question confined to the use of prognostic and predictive factors in deciding which patients with early-stage breast cancer may benefit from adjuvant systemic therapies, the lessons learnt may well be useful when considering how prognostic and predictive factors are used in medicine generally.

Evidence-based medicine has concentrated its efforts on systematically reviewing the evidence from RCTs.²⁷⁸ This has largely been because RCTs are the most reliable sources of evidence. However, treatment decisions cannot be taken in isolation from diagnostic and prognostic studies. The question posed in this review is an extremely good example of the important relationship between diagnostic and/or prognostic factors, predictive factors and choice of therapy.

It has been known for the past 100 years that some women presenting with apparently localised breast cancer can be cured by an operation to remove the tumour or breast. The importance of selection of patients was appreciated early on; indeed, it seems likely that Haagensen and Cooley's²⁷⁹ good results with radical surgery relied more on meticulous patient selection than on the nature of the surgery. In the last quarter of the twentieth century we learnt more about the nature of treatment failure in women whose breast cancer was destined to relapse despite excision of the primary localised cancer. It became clear that systemic dissemination of breast cancer often occurs early in the development of the disease, even when the cancer appears localised and is of modest size.²⁸⁰ Halsted,²⁸¹ a century ago, was wrong to assume that metastatic spread of breast cancer occurred in a progressive, temporal and predictable fashion – from tumour, through the breast to lymph nodes and only then via blood vessels to the rest of the body.

Our current concept is that many breast cancers spread locally and into lymphovascular channels right from the outset – and are therefore often systemic very early on. In this scenario, biological aspects of the cancer related to its ability to invade blood vessels and to metastasise become much more important and rank alongside the clinical extent of local spread of the cancer.²⁸⁰

Realisation that, if this theory was true, systemic therapies could be used to treat disseminated breast cancer when it was at a micro-metastatic phase led to a large series of RCTs of various systemic therapies in the last 25 years. These have been systematically reviewed by the Early Breast Cancer Trialists' Collaborative Group.^{7–14,282,283} Their findings have been consistent in that they have been able to show that a modest proportion of patients benefit and have increased long-term survival. Subsequent RCTs and reviews have refined the questions and subgroups likely to benefit, but we are still left with the problem that we need to treat many patients for a modest number to benefit.

For 100 women with breast cancer where the risk of recurrence and death from breast cancer is fairly high (50 will die of their disease), a difficult decision faces these women and their clinical advisors:

- 50 were never destined to relapse and will therefore not benefit from adjuvant systemic therapy – but we cannot identify these women.
- 50 are destined to relapse and might benefit from such therapy.
- In most instances the reduction in risk of death is about 25%.

Hence, in this example, only 12 or 13 of the 100 women will have had benefit from adjuvant systemic therapy.

This would not be a major problem if the physical, emotional and financial costs of the adjuvant therapy were minimal. However, this is not the case – chemotherapy is toxic and lasts for some months. Hormone therapy, such as tamoxifen, is less toxic, but still has unpleasant side-effects and some long-term risks and continues for 5 years.

In addition, each of these treatments is relatively expensive for healthcare providers.

Early RCTs of adjuvant therapy in breast cancer concentrated on women at moderately high risk of recurrence (those with one to four involved axillary lymph nodes). Later RCTs have included increasing proportions of women who are axillary node negative and at a much lower risk of recurrence. The dilemma about when to use adjuvant therapy is heightened in this situation. If the risk of death from breast cancer is 20%, we can reduce this by 25%. This means that 100 women will be treated, five of whom will show a survival benefit – or, to put it another way, 95 of 100 women will be treated without any benefit.

The crux of this report is whether it is possible to identify (a) women not needing adjuvant therapy and (b) those who are destined to relapse without adjuvant therapy and who will benefit from such therapy.

Ideally, adjuvant therapy should be avoided in those destined not to relapse. Adjuvant therapy should be given to those destined to relapse and who may well benefit from current adjuvant therapy. Those destined to relapse despite current adjuvant therapy will be suitable candidates for novel approaches being tested in a new generation of RCTs.

Approach to the problem

There is a vast literature on prognostic and predictive factors in early breast cancer. Because of this, we have chosen to lay the foundation of the report by concentrating on writing a series of overviews of published reviews on various aspects of the problem. These have been targeted at the following issues:

- prognostic models in breast cancer
- predictive factors in breast cancer
- reviews of prognostic information in breast cancer
- quality assessment of prognostic studies in general
- the clinical use of prognostic information in breast and other cancers
- quality of life, cost and cost-effectiveness studies relevant to modelling.

Extensive literature searches were carried out for each question with the support of an information specialist. Potential reviews were examined for their suitability for inclusion according to

prespecified criteria and those selected for inclusion were then assessed for the quality of the methods used in preparation. Quality was assessed using prespecified check-lists. Each of these processes was carried out independently by at least two of the authors. In none of the systematic reviews was it appropriate to carry out any quantitative data synthesis. Instead, narrative summaries of the reviews are presented together with a smaller number of narrative summaries of individual studies in selected situations.

This information has been integrated with a survey of the patterns of use of prognostic and predictive factors in UK cancer centres and units. This asked about the availability and use of prognostic and predictive factors in each clinical setting when choosing whether to use systemic therapy in early breast cancer. A further overview asked what information was available on the cost-effectiveness of prognostic factors in medicine in general. All of the information from these various overviews and the survey were then used to model the cost-effectiveness of the use of prognostic and predictive factors in women with early breast cancer who were being considered for adjuvant systemic therapy.

Assessing the methodological quality of prognostic studies

This proved to be a very difficult area to search, as there were few discriminating keywords. No useful reviews were found in these searches, despite 5897 abstracts being found. Seventeen relevant papers were found in other searches carried out for the report. Other papers found in an *ad hoc* fashion were also used in preparing this report.

This topic is presented in detail in Chapter 3, but a uniform characteristic of the papers identified was the lack of empirical evidence to support the importance of particular study features affecting the reliability of study findings and the avoidance of bias. Despite this, Altman,^{43,48} building on the work of others, has described a list of methodological features (*Table 2*) that are likely to be important for internal validity.

Poor reporting of primary studies is a difficulty commonly encountered in systematic reviews. It is a greater problem for prognostic studies than randomised trials, especially when survival times are analysed, for many of which the results are presented only graphically (unadjusted effects). Methods exist to estimate lnHR (and SE) from

Kaplan–Meier graphs,⁶³ but these methods cannot always be used and they make unverifiable assumptions (especially when estimating the SE). Also, the methods provide unadjusted estimates, and are therefore much more suited to randomised trials than observational studies.

Because of poor-quality research, prognostic markers often remain under investigation for many years after initial studies without any resolution as to whether they are useful. Multiple small, separate, uncoordinated and often unvalidated studies often delay the process of defining the role of prognostic markers.

Because of this, systematic reviews of published studies reveal a confusing picture with many studies poorly done and poorly reported. They often fail to answer questions, although they are helpful as they draw attention to the paucity of good-quality evidence and the need to improve the quality of future research in this area. Cooperation from the outset between different research groups could lead to clear results emerging more rapidly than is commonly the case, especially if such efforts are put into prospective studies or retrospective studies based on individual data from carefully assembled databases and/or tissue banks.⁸²

The main problems encountered in systematically reviewing evidence from these types of studies are summarised below:

- Difficulty of identifying all studies – there is often a lack of keywords.
- Negative (non-significant) results may not be reported (publication bias) and researchers may report a small number of significant results from a larger number of factors examined (selection bias). The dangers of such biases seem likely to be larger than with RCTs.
- Inadequate reporting of methods used in the original reports.
- Variations in study design.
- Most studies are retrospective.
- Variations in inclusion criteria between different studies.
- Lack of recognised criteria for quality assessment of prognostic/predictive studies.
- Different assays/measurement techniques and storage of samples between studies of the same factor.
- Variations in methods of analysis.
- Differing methods of handling continuous variables (some data dependent, such as optimal cut-points).

- Different statistical methods of adjustment.
- Adjustment for different sets of variables.
- Inadequate reporting of quantitative information on outcome.
- Lack of adequate follow-up time.
- The power available to detect a predictive effect is markedly lower than for detecting a prognostic effect – roughly four times the sample size is necessary to have equivalent power.
- Few analyses of predictive factors are prespecified, but are rather based on exploratory analyses and possibly also selected from among several such analyses.
- Variation in presentation of results (e.g. survival at different time points).

These difficulties are serious enough, in that they hinder systematic review of evidence. More serious is the knock-on consequence of failing to identify which factors are really useful in clinical practice or delaying this process by many years.

Systematic review of reviews of studies of prognostic factors

Systematic review methods were used to examine the most reliable reviews of prognostic factors. There is an enormous literature reporting studies of prognostic and predictive factors in breast cancer. Because of the constraints of the brief for this report, we decided to review reviews of the subject rather than attempt to review individual reports.

An extensive literature review, not restricted to the English language, was carried out and at least two of the authors assessed reviews for suitability for inclusion and then rated included reviews by prespecified quality criteria.

Quality assessment used a standard data collection form that had been piloted and included the following domains:

- extent of searching (including databases, and any language or time restrictions)
- description of the eligibility criteria for the review
- comparability of the included studies
- assessment of publication bias
- assessment of heterogeneity
- conduct of sensitivity analyses.

In this section, we identified reviews that were considered to be systematic in their nature. Data

extracted from these were used to prepare narrative summaries for prognostic factors where evidence was of sufficient quality. The number of systematic reviews found for each prognostic factor varied between one and six.

The most commonly used, and currently the most effective, prognostic factors in breast cancer have been little studied in the context of systematic reviews. This is perhaps similar to the lack of systematic reviews and randomised trials, for interventions whose effectiveness is striking enough to be beyond doubt. The prognostic factors to which this applies include nodal status and age at diagnosis.

The reviews that we identified in this report are of factors for which the evidence is less robust. In many of these reviews, the importance given to nodal status and age at diagnosis above is apparent, the focus of the review being a factor within a group of women whose breast cancer has already been categorised using one of the above (most commonly nodal status, and in particular to investigate additional prognostic factors for women with node-negative breast cancer). *Table 45* shows the prognostic factors for which we identified systematic reviews.

Chapter 4 includes a detailed narrative review of each of these using a standard structure. This section condenses this information by type of factor to bring together the main findings.

TABLE 45 Prognostic factors and identified systematic reviews

Prognostic factor	No. of reviews
Cathepsin D	3
Epidermal growth factor	2
HER2 ERBB2	6
Urokinase and its receptors	3
p53	5
p21	2
bcl2	1
C-MYC amplification	2
Proliferation indices	5
pS2	1
nm23	1
Aneuploidy or DNA ploidy	2
Tumour size	1
Grade	1
ER status	1
Vascular invasion	1
Bone marrow micrometastases	2
Body size	1

Oestrogen receptor pathways

The ER is critical in breast cancer, both as a cause of growth of breast cancer when stimulated by oestrogen and as a therapeutic target to inhibit growth. When activated, the ER switches on the PR and many other genes, including a secreted factor pS2. Thus the latter two genes act as markers for active ER signalling. If the ER is too low to detect, PR may still be detectable and measurement of this can provide evidence of an active ER signalling pathway warranting antioestrogen therapy.

This project found only one systematic review⁸³ of ER focusing on the relationship between ER-negative status and disease-free survival and overall survival. It is concluded that the association between ER status and prognosis in women with node-negative breast cancer has not been confirmed or refuted by the seven studies in the review.

pS2

One review was found of studies of the prognostic nature of pS2. This was within a review of the prognostic relevance of biological markers in general.⁸⁴ This review contained only two studies. Although this review concluded that pS2 was associated with poorer disease-free survival and overall survival, it was not possible to assess the quality of the review itself or whether the few studies identified are a complete sample of the evidence 10 years ago. Because of this, the findings of the review must be treated with caution and more research is needed to confirm or refute any association between pS2 and prognosis.

Growth factor pathways

About one-third of breast cancers do not express the ER and are regulated by other growth factor pathways. Many of these signal via transmembrane receptors, the extracellular domain binding the growth factor and the intracellular domain activating a signalling cascade. EGFR is a member of the HER2 (human EGF receptor family 2) family, EGFR being the first to be described. Both tend to be reciprocally expressed with ER although all combinations of expression occur. These growth factor receptors can be inhibited by drugs or antibodies and are an important target for therapy in breast cancer. The gene coding HER2 is often amplified (i.e. there are multiple copies in the cancer cells) and this genetic change is strong evidence for a major role in the tumours that express it. Both receptors are associated with stimulation of growth and invasion in breast cancer cell lines, hence the interest in whether

they have a direct role in behaviour of breast cancer.

Up-regulation of transcription factors and nuclear proteins that regulate the cell cycle is another mechanism of transformation and the gene for C-MYC is commonly amplified in breast cancer.

Epidermal growth factor receptor

Two reviews were found of studies of the prognostic nature of EGFR. One of these was part of the presentation of the detailed results for a series of patients together with 16 prior series, of which nine had survival data⁸⁵ and the other was a review of 40 studies of EGFR.⁸⁶

There are quality issues in both of these reviews. However, they are both of the opinion that firm conclusions cannot be drawn about the association between EGFR and survival – we agree that the association is unproven and that further research is needed.

Human EGF receptor family 2 (HER2)

This project identified six reviews of studies of the prognostic nature of HER2. These included three standalone reviews^{87–89} and one within a general discussion of prognostic and predictive factors.⁹⁰ A fifth review investigated HER2 within a wide-ranging review of prognostic factors in node negative breast cancer.⁸³ The Porter-Jordan review⁸⁴ is superseded by the more recent reviews and is not discussed further here. Although the Nunes review⁸⁹ is the most recent, it appears highly selective in the literature that it includes and for this reason has not been considered further.

All of these studies show a significant association between HER2 and shorter disease-free or overall survival, in either univariate or multivariate analyses. All of the reviews include large numbers of studies (13–97) and patients (4996–22,616) although some are restricted to specific subgroups of women with breast cancer.

No details of the eligibility criteria or assessment process are given for the Henderson,⁹⁰ Ross⁸⁸ or Révillion⁸⁷ reviews. Although the eligibility criteria for the Mirza review⁸³ were stated briefly, no details were given of how these criteria were applied.

The Henderson review⁹⁰ shows the univariate *p*-value for disease-free survival and/or overall survival for each study. Most of the studies showed a statistically significant association between worse outcome and HER2 level. The Ross review⁸⁸ adopts a vote counting approach and concludes

that the associations between HER2 and prognosis are “somewhat controversial”. They point out that, for node-positive patients, the status of HER2 as an independent prognostic factor was seen in some, but not all, studies, and that the studies with longer follow-up were less likely to find an association. For node-negative patients, they point out that “most studies did not find HER2 to be a prognostic indicator”.

In the Mirza review⁸³ of node-negative breast cancer, all 13 studies reported a univariate analysis of the relationship between HER2 and disease-free survival and/or overall survival. Eight of these showed a significant association and five did not. Six studies reported a multivariate analysis: HER2 was significant for disease-free survival in two but not significant for both disease-free and overall survival in the other four. On the basis of their findings, Mirza and colleagues⁸³ conclude that there is only “limited association” between HER2 and prognosis in these patients.

The suggestion of a likely association between HER2 and poorer prognosis in women with node-positive breast cancer appears to be justified based on the weight of evidence in these reviews. There is less evidence for an association among node-negative patients.

C-MYC amplification

Two reviews were found of studies of C-MYC amplification as a prognostic factor. One of these was a review of C-MYC amplification, with meta-analysis, in women with breast cancer.⁹¹ The other investigated C-MYC amplification within a broader review of biological markers in general.⁸⁴ The Deming review⁹¹ was more comprehensive and up-to-date and further data were not presented for the Porter-Jordan review.⁸⁴

The Deming review⁹¹ was a well-conducted systematic review in which the authors found a relationship between C-MYC amplification and poor outcomes. However, they point out weaknesses in methodology and suggest that more rigorous studies, with consistent methodology, are needed to verify the association. Although based on the evidence in this review, it does appear that the presence of C-MYC amplification is an indicator for poor prognosis.

Cell proliferation markers

There are many other growth factors and receptors involved in breast cancer, so a final common pathway, proliferation, is often measured. Proliferation can be assessed by the number of

cells dividing (the mitotic index) or the expression of key molecules involved in regulating the cell cycle. S-phase fraction is a measure of the percentage of cells in the tumour that are making new DNA. Ki-67 is a nuclear protein that is up-regulated during the cell cycle and can be scored on tumour histology sections. It provides another marker for the number of proliferating cells.

Three useful systematic reviews were found of studies of the prognostic nature of cell proliferation markers. One was a review of S-phase in women with breast cancer⁹³ and one examined four proliferation indices: thymidine-labelling index (TLI-BrdULI), S-phase fraction, mitotic count and Ki-67-MB-1.⁹⁴ A third review investigated S-phase fraction, mitotic index and Ki-67 within a review of prognostic factors in node-negative breast cancer.⁸³

Wenger and Clark⁹³ found 20 studies that investigated the relationship between S-phase fraction and overall survival using univariate analysis, of which 18 found that high S-phase fraction was associated with decreased overall survival. We concluded that a high S-phase fraction has clinical utility for patients with breast cancer and that high S-phase fraction is associated with poor prognosis. The Daidone and Silvestrini review⁹⁴ found eight studies (half of the women were in one study), which investigated the relationship between thymidine-labelling index and survival using multivariate analysis.

Five of these studies found that high thymidine-labelling index was significantly associated with worse disease-free or overall survival. They reported on 22 trials of S-phase fraction, of which 15 found that this was associated with statistically significant shorter disease-free or overall survival. The review found nine studies of mitotic count. Five of these found this was a significant prognostic factor in multivariate analyses. They also reported on four studies that investigated the relationship between Ki-67-MB-1 and disease-free or overall survival using multivariate analysis. Two of these found that high Ki-67-MB-1 was associated with decreased disease-free survival. They did not combine the studies they identified in a meta-analysis, but conclude that the four proliferation indices they investigated are each associated with poor prognosis.

Three of the five studies in the Mirza review⁸³ of S-phase fraction conducted a univariate analysis of the relationship with disease-free or overall survival. It was a significant positive prognostic

factor in two studies, but negative in the other. Four studies used a multivariate analysis. S-phase fraction was a prognostic factor for overall survival in three of these studies and was not a prognostic factor for disease-free survival in the fourth. Based on these findings, Mirza and colleagues⁸³ conclude that S-phase fraction is a “useful” prognostic factor.

Overall, it is felt that there is a consistency in the literature favouring the conclusion that markers of a high proliferation index are prognostic of a poor outcome.

Genetic instability and checkpoints

Some of the genes that transform normal breast cells into cancer (oncogenes) can induce genetic instability by BRCA1, ataxia telangiectasia and BRCA2. In addition, as cells proliferate, genetic damage occurs and uncontrolled proliferation without proper ‘check-points’ on duplicating DNA results in losses and gains of chromosomes. A downstream final end-point of many of these changes is whether there are the normal number of chromosomes and amount of DNA per cell or abnormal amounts. This is assessable by measuring ‘ploidy’ of tumours, where being aneuploid is having an abnormal amount of DNA. It may be expected that those with the most abnormality can generate variants with more aggressive behaviour.

Cell cycle ‘check-points’ include a nuclear protein p53 that stops cells dividing if the DNA is damaged or in response to other stresses. One of the proteins that does this is p21, itself up-regulated by p53. High expression may indicate the ability of the cancer cells to respond to stress.

Aneuploidy or DNA ploidy

Two reviews were found of studies of the prognostic nature of DNA ploidy or aneuploidy, both being in the context of broader reports on prognostic factors. The older review is part of a general discussion of various prognostic factors⁹⁵ but remains relevant because it covers a period before that assessed by the second review. The more recent review is a review of prognostic factors in node-negative breast cancer.⁸³

There were quality issues in both reviews, although the older review was probably weaker. The results of the two reviews and constituent studies were inconsistent and overall there was insufficient evidence to draw firm conclusions on a relationship between ploidy and survival.

p53

Four reviews were found of studies of the prognostic nature of p53. Three of these focused entirely, or almost entirely, on p53^{96,98,258} and one investigated p53 within a review of prognostic factors in node-negative breast cancer.⁸³

These four reviews approach the association between the p53 gene and prognosis in different ways and reach slightly different conclusions. However, it is concluded that the strength of the evidence does support an association between alterations to the p53 gene and poor prognosis.

p21

Two reviews were found of studies of the prognostic nature of p21. One of these was part of the presentation of the results for a series of patients.¹⁰² The other considered p21 alongside a more detailed review of p53.²⁵⁸

It is concluded that more research is needed to confirm or refute any association between p21 and prognosis.

Apoptosis

As a tumour expands, it may outgrow its blood supply, become hypoxic and have insufficient metabolites for growth – leading to areas of cell death. Although there is an increase in cell division, often it is so abnormal that the cell dies. Both of these processes contribute to a high death rate of cancer cells. The mechanism of cell death requires specific biochemical pathways leading to ‘apoptosis’ – programmed cell death. Certain genes can protect cells against apoptosis, e.g. bcl-2, and therefore cancers which express these genes may have a better chance of survival and growth. bcl-2 is a protein that stabilises mitochondria against release of toxic proteins and metabolites that activate apoptosis.

p53 (discussed above) is also important in producing apoptosis in response to stress, thus ensuring that damaged cells do not carry on to produce abnormal daughter cells. Mutations blocking this effect are common in most cancers, allowing cancer cells a survival advantage.

bcl-2

This project identified one review of studies of the prognostic nature of bcl-2.¹⁰³ Although Zhang and colleagues¹⁰³ conclude that patients with high bcl-2 immunostaining have a better clinical outcome than those with low/negative bcl-2 expression, we felt that this is likely to be explained by the relationship between bcl-2

protein and differentiation. It is difficult to judge the quality of this review in the absence of clearer information on how the reviewers conducted their search, appraised the reports they found and determined what should be included in their review. Overall, it was not felt possible to determine if their conclusion of a definite relationship between prognosis and bcl-2 is justified by the studies that have been done, as opposed to the studies that they have included.

Metastases

For tumour cells to spread from the primary tumour to distant sites (metastasis), destruction of the extracellular matrix around the cells is necessary. This is also necessary to allow new blood vessels to supply the tumour and for the circulating tumour cells to invade distant organs. Initially they need to invade the local blood vessels (vascular invasion) or lymphatics (lymphatic invasion), often scored as lymphovascular invasion. There are many pathways involved in normal degradation and turnover of the extracellular matrix, including multiple proteases and heparinases. Cathepsin D was one of the earliest to be studied because it was found to be oestrogen regulated, but many others are important, including urokinase and its receptor, which binds the enzyme to the surface of invading cells, several metalloproteases and stromelysins. Another pathway regulating metastasis is the enzyme nm23. This suppresses metastasis and loss of its expression is associated with disease spread, so it is a tumour suppressor gene.

One way of measuring the final overall metastatic potential of cell division is to look for small clusters of cancer cells in the bone marrow or the circulation by highly sensitive immunochemistry and molecular techniques which can detect minimal residual disease.

Cathepsin D

This project identified three reviews of studies of the prognostic nature of cathepsin D. One was a review of cathepsin D in women who had node-negative breast cancer.¹⁰⁴ The others investigated cathepsin D within broader reviews of prognostic factors in node-negative breast cancer⁸³ and biological markers in general.⁸⁴

All three reviews concluded that high expression of cathepsin D is related to poorer prognosis in women with node-negative breast cancer. Consideration of women with node-positive disease was restricted to the Porter-Jordan review.⁸⁴ A brief mention was made in this review to a non-statistically significant

relationship between cathepsin D and prognosis in such women, but this is based on a citation to a single study.

Although there are potential flaws in the reviews identified for this project, the consistency of the findings across the reviews and the quantity of evidence identified justify the conclusion that high expression of cathepsin D is related to poorer prognosis in women with node-negative breast cancer. There is insufficient evidence to draw any reliable conclusions about this relationship in women with node-positive disease.

Urokinase-type plasminogen activator

We found two reviews and one pooled analysis of studies of the prognostic nature of urokinase or its receptors. The urokinase-type plasminogen system comprises at least four proteins, all of which have been studied in reviews. These are the urokinase-type plasminogen activator (uPA), its membrane bound receptor (uPAR) and two inhibitors (PAI-1 and PAI-2); and the identified studies examined these to varying extents.

The two reviews were standalone reviews of uPA, uPAR, PAI-1 and PAI-2¹¹¹ and PAI-1.¹¹² The pooled analyses, done by EORTC-RBG, used a specially compiled dataset to examine uPA and PAI-1.¹¹³

All three reports concluded that a high value of PAI-1 is related to poorer prognosis. The consistency of this finding in the reviews and in the pooled analysis, despite the different data used in each, justifies this conclusion. The reviews and the pooled analyses are also strongly supportive of the same relationship for uPA. There is insufficient data on uPAR and PAI-2 to draw any conclusions.

nm23

This project identified one review of studies of the prognostic nature of nm23. This was within a review of the prognostic relevance of biological markers in general.⁸⁴

There was a lack of information on the methods in this review and it is not possible to assess the quality of the review itself or whether the three studies reviewed are a complete sample of the worldwide evidence from over 10 years ago. Further information is needed.

Bone marrow micrometastases and minimal residual disease

One systematic review was found of studies of the prognostic nature of bone marrow micrometastases

in various types of cancer, with most studies being on breast cancer,¹¹⁵ and one review of minimal residual disease in breast cancer.¹¹⁶

It is felt that the conclusion of both reviews that more evidence is needed to be certain about the possible prognostic nature of bone marrow micrometastases or minimal residual disease is justified. However, based on the evidence in the reviews, it does appear that the presence of micrometastases is an indicator for poor prognosis.

Standard pathology

Other prognostic factors studied include aspects of standard pathology, which always needs to be considered, because of the robust and extensive data and routine use in management. These include size and grade of the tumour, lymphovascular invasion, involvement of surgical margins and spread to regional lymph glands. These are the core on to which new factors can be grafted to see if further refinement is possible.

Tumour size

One review was found of studies of the prognostic nature of tumour size. This was within a broader review of prognostic factors in node-negative breast cancer.⁸³

All of the nine reported studies found a statistically significant association between tumour size and either disease-free survival, overall survival or both. Six of the nine studies reported a univariate analysis of the relationship between tumour size and disease-free survival or overall survival. All of these showed a significant relationship. In addition, four studies included a multivariate analysis, and tumour size was a significant prognostic factor for overall survival in each of these.

We agree with the conclusions of Mirza and colleagues⁸³ that there is good evidence that increasing tumour size is associated with poorer prognosis in women with node-negative breast cancer.

Tumour grade

One review was found of studies of the prognostic nature of tumour grade, within a general review of prognostic factors in node-negative breast cancer.⁸³

Five of six studies reported included a univariate analysis of the relationship between grade and disease-free or overall survival. All five showed a significant relationship. Three studies used a

multivariate analysis and Mirza and colleagues⁸³ highlight the importance of systemic therapy as a variable in such analyses. In the study in which none of the patients received systematic therapy, there was a significant association between grade and overall survival. There were two studies in which some patients received systemic therapy. One study that included treatment in the multivariate analysis concluded that grade was a significant prognostic factor. The other study did not include treatment in its model and showed no significant association.

Mirza and colleagues⁸³ conclude that there is an association between grade and prognosis for women with node-negative breast cancer. Since the association was seen in almost all the analyses in the studies in the review, this conclusion appears justified. However, the possibility of publication bias and the lack of detail on how studies were assessed as eligible for the review mean that it should still be treated with caution. In addition, the review provides no information on the association between tumour grade and prognosis in women with node-positive breast cancer, although this may not have been studied since an association between poor grade and poor outcome has been assumed in women with node-positive breast cancer.

Vascular invasion

One review was found of studies of the prognostic nature of lymphovascular invasion. This was within a broader review of prognostic factors in node-negative breast cancer.⁸³

Two of the five studies reported a univariate analysis of the relationship between vascular invasion and disease-free survival and overall survival. Both showed a significant relationship. One of these studies and the other three in the review used a multivariate analysis. Vascular invasion was a statistically significant prognostic factor for overall survival in all of these.

Mirza and colleagues' conclusion⁸³ that the presence of vascular invasion is associated with poorer prognosis in women with node-negative breast cancer appears justified, although the lack of information on how any of the five included studies were identified or judged to be eligible suggests that some degree of caution is warranted. Although no systematic review of tumour grade in node-positive breast cancer was found, there is no reason to believe that the effect of grade would be confined to node-negative patients. Vascular invasion is one

of the histopathological cornerstones of the NPI and it may not have been reported in a systematic review since the association was assumed already.

Body size

Because larger or fatter people may have differences in the amount of hormones they make or growth factors, body habitus has also been studied for its effect on prognosis and is potentially important as it may be controllable by diet.

One review was found of studies of the prognostic nature of body size.¹²² The authors conclude that body size appears to exert a modest effect on prognosis in breast cancer that persists after adjustment for the effects of other prognostic factors. Although some aspects of this review are well done, notably the detailed assessment of the quality of the reports of the studies, its conclusion is based on vote counting and its reliance on English language original articles leaves it open to publication bias. We conclude that the moderate association between body size and prognosis may be an overestimate – as suggested by the authors.

Overall conclusions

The lack of good-quality systematic reviews and well-conducted studies of prognostic factors in breast cancer is a striking finding. There are no registers of studies of prognostic factors or of reviews of prognostic studies. Searching is difficult and the risk of publication and selection bias is an ever-present problem.

Many of the reviews used weak methods, scoring poorly on issues identified in *Table 2*. The same applies to primary studies where there was poor methodology and reporting of results. In addition, there is much variation in patient populations, assay methods, analysis of results, definitions used and reporting of results. Most studies appear to be retrospective and some use inappropriate methods likely to inflate outcomes such as optimising cut points and failing to test the results in an independent population. *Table 46* summarises findings from systematic reviews of various prognostic factors.

Very few reviews used meta-analysis to conduct a pooled analysis and to provide an estimate of the average size of any association. Instead, most reviews relied on vote counting.

TABLE 46 Summary of findings from a systematic review of reviews of prognostic factors in breast cancer

Cathepsin D	++ for node negative
Epidermal growth factor	+/-
HER2/c-ERBB2	+
Urokinase and its receptors	PAI-1 ++
UPA	-/+ +
UPAR and PAI-2	+/-
p53	++
p21	+/-
bcl-2	+/-
C-MYC amplification	+
Proliferation indices	S-phase fraction ++
pS2	+/-
nm23	+/-
Aneuploidy or DNA ploidy	+/-
Grade	+
Tumour size	++
ER status	+/-
Vascular invasion	+
Bone marrow micrometastases	+
Body size	+/-

–, No relationship between factor and survival;
+/-, insufficient evidence to identify a relationship between factor and survival; +, evidence of a relationship between factor and survival; ++, clear evidence of a relationship between factor and survival.

Prognostic models in breast cancer

Statistical models for predicting patient outcome are termed prognostic models. They are widely used in cancer for investigating patient outcome in relation to multiple patient and disease characteristics. Such models may allow classification of patients into two or more groups with different prognoses. This information can be used to influence therapy and in selecting diagnostic tests. It can also be used to estimate the prognosis of individual patients.

With regard to this report, such models can be used to identify women who have such a good prognosis that the benefit of adjuvant therapy would be too small to justify the side-effects or cost. Conversely, there may be groups of patients whose prognosis is so poor that improved survival is too unlikely to justify treatment with conventional approaches. In this circumstance, no adjuvant therapy or adjuvant therapy with an experimental approach may be most appropriate. At present, many patients with breast cancer fall into a middle category where there is moderate to modest risk of recurrence and where adjuvant therapy seems appropriate. Ideally, prognostic models should be better at discriminating in this group those who might benefit from adjuvant therapy – identifying

more accurately those destined to relapse and who may benefit from adjuvant treatment.

This subject is considered in detail in Chapter 5. Our main interest was in identifying prognostic models for which there has been an evaluation of how successfully the models have been used in a different setting, that is, models which have been validated externally.^{124,125,130}

Both clinical and statistical issues are of major importance when thinking about prognostic models (Table 7). Wyatt and Altman¹²⁹ suggested the following prerequisites for clinical credibility:

- All clinically relevant patient data should have been tested for inclusion in the model – in practice, models are often developed in retrospective studies using those variables that happen to have already been collected for other reasons. Because of this, many studies omit potentially valuable variables as they do not have data for them.
- It should be simple for doctors to obtain all the patient data required, reliably and without expending undue resources, in time to generate the prediction and guide decisions. Data should be obtainable with high reliability, particularly in those patients for which the model's predictions are most likely to be needed.
- Model builders should try to avoid arbitrary thresholds for continuous variables.
- The model's structure should be apparent and its predictions should make sense to the doctors and patients who will rely on them – this argues against 'black-box' methods such as neural networks.
- It should be simple for doctors to calculate the model's prediction for an individual patient.

Most studies of models have been too small. Harrell and colleagues¹³⁸ suggested that the number of Events should be at least 10 times the number of potential Prognostic Variables investigated, a value (EPV) supported by a simulation study.¹³⁹ Feinstein¹⁴⁰ suggested that a minimum EPV of 20 is safer and Schumacher and colleagues suggest 10–25.¹³³ In most cases the minimum EPV is less than 10, which is likely to be a major source of unreliability in their findings, especially when they have used some stepwise algorithm for selecting the model.

Prospective studies are preferred as the data sets are likely to be much more complete than in retrospective studies using clinical databases collected for other purposes. Incomplete data is a

common and often serious problem for studies developing prognostic models. Thus while the sample size may be large, patients missing one or more variables will generally need to be excluded from a modelling exercise. The obvious effect will be to reduce power, but a much more serious possibility is the risk of introducing bias. The alternative to excluding patients with missing data is to impute the missing values. Various strategies are available,^{143,144} but they make some fairly strong and unverifiable assumptions about the reasons for the incompleteness in the data.

An important aspect of modelling is how to handle continuous variables. The choice is primarily between keeping such variables continuous, usually leading to the specification of a linear relation between the variable and lnHR, or creating categories and thus largely avoiding the problem of model specification. There are considerable advantages in keeping variables continuous.^{146,147} However, categorisation is extremely common in oncology – indeed, splitting into two groups (dichotomisation) may be considered the norm.

Chapter 5 discusses the disadvantages of dichotomisation further and the dangers in different methods of choosing cut-points – particularly the use of optimised cut-points.

The development of models often uses techniques such as Cox regression. Unfortunately, the results of the widely used stepwise regression analyses are likely to be misleading.^{40,59} The regression coefficients in the final selected model may be biased, being on average too large. This effect is the result of the inclusion and exclusion of variables based on their association with outcome. Significance tests associated with these inflated coefficients are not strictly valid and the *p*-values are too small. It is common practice to include all variables significant at an arbitrary level of significance of 0.05 in the final model. The selection of variables on this basis has no direct relationship to clinical importance. Also, the classification of certain variables as important misrepresents the fact that models based on very different sets of variables may predict equally well. All of these difficulties are exacerbated when there are few events per variable. The wide use of such methods is an argument supporting the need to carry out a validation study before claiming that a model is useful. Other approaches to modelling, including ANNs, are discussed in Chapter 5.

With one or more continuous variables, it is usually necessary to calculate a PI and divide the

range of values into bands representing different levels of risk. There is no consensus on how many groups should be created, or on how to choose the cut-points.¹⁴⁶

Reviewing current models

Papers were sought that presented new prognostic models for patients with operable breast cancer or which evaluated a previously published model (validation study), or both of these. A total of 4791 abstracts were initially identified. The final number of papers reviewed was 78, 17 of which were eventually excluded as ineligible. Methods for study selection and data extraction are covered in Chapter 5.

The characteristics of the 61 included studies are shown in *Table 9*. About 89% were cohort studies and 79% were retrospective. None justified the sample size. Inclusion criteria were only explicit in 38% of studies. Different definitions and end-points were frequently used. Length of follow-up was not stated in 38% of reports and in 77% there was no clear statement on how patients lost to follow-up were dealt with.

Many studies had fewer than 10 events per candidate variable. An extreme case is the study of Kaufmann and colleagues,¹⁸⁶ who investigated 11 variables in a data set of 57 patients with only four events.

Table 10 presents the characteristics of 54 prognostic models developed for breast cancer. The striking findings are the variety of different analytical methods used and the variability and often failure to report key features that would allow readers to assess the validity of the model.

Assessment of published prognostic models

The NPI and derivatives

The NPI is one of the oldest indices proposed for breast cancer patients. It is one of the relatively few such indices that is actually used in clinical practice. Its use is particularly common in the UK (Chapter 7). Likely explanations for its wide uptake include its simplicity, clinical credibility and especially the demonstration that it performs well in different populations (validation).

The Nottingham group have published a series of papers refining the model.^{193,196,197} The third of

these papers¹⁹⁷ examined the survival of a cohort of 1629 patients, including those already analysed. The index was again shown (graphically) to produce three risk groups with well-differentiated survival, similar to that seen in the original study. In this larger series the proportions in the three risk groups were 29, 54 and 17%. These authors made the important suggestion that lymph node stage could be replaced in the NPI by the number of involved nodes. They suggested using groups of 0, 1–3 and 4+ involved nodes, although they did not present any analyses to show the impact of this change. Despite some deviations from what is now common statistical practice, the model clearly has very good discrimination both in the original sample and in subsequent evaluations elsewhere, as described below.

At least six studies have examined the validity of the NPI. Balslev and colleagues²³⁶ evaluated the NPI in 9149 patients. This was a large, high-quality study in which patients were all enrolled in prospective, protocol-led studies performed by the Danish Breast Cancer Group. Using three risk groups, they found that the survival was broadly similar in the Danish cohort to that in the original Nottingham cohort, although 10-year survival was rather better in the poor prognostic group in the Danish series. This similarity was despite some differences in the details of assessing stage and grade from the original Nottingham study.

Collett and colleagues²³⁷ evaluated the NPI in 1223 patients and compared their results with those of Balslev and colleagues.²³⁶ They used the modified NPI using number of involved nodes (0, 1–3, 4+). They split the NPI into three groups and reported 10-year survival rates similar to Balslev and colleagues,²³⁶ although slightly better in the lowest two groups. From an analysis of about 700 women, they concluded that adding ER and PR to the NPI gave additional prognostic information, but they did not present a model with all those variables included. These authors showed that the prognostic separation achieved by the NPI was weaker in the second 5 years than the first 5 years. Such a difference is not unexpected, but hardly any papers have considered this issue.

Sundquist and colleagues²³⁸ evaluated the NPI in 608 Swedish patients and confirmed the high degree of prognostic separation between groups defined by the index. They also fitted a new model using the same variables, which they called the KPI. The KPI was similar to the NPI, but tumour size and especially grade had higher coefficients in the KPI model. The authors

‘normalised’ their model by multiplying by 0.78 so that the two indices had the same means. The two models agreed well and gave similar discrimination. The results were very similar for all-cause mortality and cancer mortality.

Other models

Despite considerable heterogeneity of clinical characteristics, the variables studied and the statistical approach to deriving a model, some clear features in other models can be seen. A relatively small number of variables feature in more than one or two of the models. The most commonly included variables are nodal status, tumour size and grade. Other variables often included are age, ER status and PR status. A number of variables feature in just one published model. In many cases, these studies were the only ones to investigate those particular factors, and presumably their inclusion reflected a particular research interest of that group. Although age featured fairly often for the end-points of death and recurrence, it was rarely important in models for predicting cancer death.

We have not attempted to summarise the discriminatory ability of the many models, partly because of a lack of a standard metric used in all papers and partly because such measures could be influenced by major variations in case mix. By definition, in almost all cases the variables in the models were shown to be statistically significant in the study samples, although many of the studies were small enough for concerns about over-optimism.

Only three papers^{187,239,284} reported validation studies of two models other than the NPI, and only one was not by the group that developed the model. The study by Collan and colleagues,¹⁸⁷ of just 120 women, represents the only independent evaluation of any prognostic model other than the NPI.

Remarkably few published prognostic models have been re-examined by independent groups in independent settings. Most validation studies have been by the investigators themselves. The few validation studies have been carried out on ill-defined samples, sometimes of smaller size and short follow-up, and authors in general are unclear about how to summarise the performance beyond showing Kaplan–Meier plots of survival. Some authors used a different patient outcome when validating a model than the one used to develop it, as if these outcomes were interchangeable. A few of the investigators suggested modifications to

the original model in the light of the validation process. Any such changes would, of course, themselves need validation.

Overall, the only clear message from the published validation studies is support for the prognostic value of the NPI. Chapter 5 discusses some of the methodological flaws seen when prognostic models are developed. No new prognostic factors have been shown to add substantially to those identified in the 1980s. As Haybittle²¹⁸ noted, "Any improvement [on the NPI] in prediction must now depend on finding factors which are as important as, but independent of, lymph node, stage and pathological grade." The NPI remains a useful clinical tool, although additional factors may enhance its use. Such factors have proved surprisingly elusive,¹⁶⁹ as is evidenced by the continued widespread use of the NPI in clinical practice (Chapter 7) after 20 years.

Further, no other prognostic model has emerged that is clearly superior to the NPI. That said, it seems clear that there is a small set of prognostic variables, perhaps especially ER, that may usefully add to the variables (grade, tumour size and positive lymph nodes) included in the NPI.

Predictive factors in breast cancer

In general, factors examined for their ability to predict response have been the same ones tested for prognostic value. This section will not include factors where the treatment requires an intact receptor or target [ER for tamoxifen and HER2 for trastuzumab (Herceptin)]. Rather, it will concentrate on a series of factors suggested to be of predictive value, but where a mechanism of action for the factor and drug is not clearly evident.

Cathepsin D

Two studies were found. One²⁵⁰ was a moderate sized regional study of cathepsin D and the other²⁵¹ a larger regional study of cathepsin D and plasminogen activator inhibitor-1. Both studies were retrospective and of reasonable quality. Both studies show that cathepsin D may be predictive of response to tamoxifen in (different) subgroups of patients. They provide weak evidence that cathepsin D can be used to predict response to hormone therapy (tamoxifen). In view of the potential for reporting bias in this literature and the relative weakness of the evidence, we conclude that there is no good evidence to support the use

of cathepsin D as a predictor of response to systemic adjuvant therapy.

HER2/neu

Five systematic reviews were found. Each review containing a series of different and overlapping trials.^{87,254-257} The methodology used in each study was of variable quality.

There is a consistency in the results of the reviews. Despite suggestive evidence of predictive effects of over-expression of HER2 (anthracyclines more likely to be beneficial and tamoxifen less useful), the overall consensus of the reviews is that there is insufficient evidence to develop guidelines on the use of HER2 to select specific hormonal or chemotherapy for individual patients. We would concur with this. It is, of course, essential when selecting patients for treatment with Herceptin.

p53

Two systematic reviews^{258,259} of the predictive value of p53 were found and this overview is confined to these reviews. Their quality was variable and some of the papers included patients with advanced disease. Although both stressed the need for further research, they also both concluded that there was no evidence to support the use of p53 as a predictive factor for tamoxifen or chemotherapy.

Survey of UK cancer centres and units

There is little systematic evidence about the patterns of use of adjuvant therapy for early breast cancer in the UK or elsewhere. The evidence that is available suggests that there are major variations in practice and that decisions are not always made on the basis of complete clinical information. The aims of the survey were threefold: first, to ascertain what prognostic and predictive factors were available to and used by clinicians when selecting adjuvant systemic therapy, second, to identify the patterns of use of adjuvant therapy arising as a result of using prognostic and predictive information and third, to ensure that subsequent analyses, undertaken within the overall project were appropriate, practical and could be implemented within UK breast cancer units.

Details of the methods used to pilot and carry out the survey are given in Chapter 7. The lead clinician for breast cancer and responsible histopathologist for each cancer centre and unit were invited to complete the survey.

Twelve of the 230 units indicated that they did not treat breast cancer, giving overall response rates for clinicians of 77% (168/218) and for histopathologists of 90% (196/218).

Clinicians' questionnaire

When deciding whether to offer adjuvant chemotherapy for women 50 years and younger, clinicians cited the most important clinical factors as nodal status (pathological) (96%), tumour grade (95%), tumour size (86%) and ER status (79%).

These same factors were cited when selecting systemic adjuvant chemotherapy for newly diagnosed women aged over 50 years with regionally localised breast cancer, 96, 93, 86 and 82%, respectively.

Factors deemed to be the least clinically important in both patient populations were margins positive with invasive cancer (50 years and younger, 14%; over 50 years, 13%); clinical nodal status (13% both age groups); proliferation index (5 and 6%, respectively) and bcl-2 (2% both age groups).

The factors rated as most important by clinicians [tumour grade, size, nodal status (pathological), ER status and menopausal status] were widely available to clinicians. Age, vascular invasion and histological subtype were also reported as being very available, yet were not identified as being the most clinically useful factors when selecting adjuvant therapies for this patient population.

The two factors that were least available (proliferation index and bcl-2) were also cited as the least clinically important across both treatment and age subgroups. Margins positive with invasive cancer and nodal status (clinical) were available to 96 and 83% of clinicians, respectively, yet both appear to be of minimal clinical importance across treatment and age subgroups.

PR was the main factor cited as being clinically important (particularly when selecting hormone therapy) that was not always available to the clinician. Other factors that were deemed to be clinically important but not always available were physiological age (particularly in the chemotherapy, aged over 50 years subgroup), HER2 (principally for the chemotherapy treatment groups) and a calculated PI.

About 38% of clinicians stated that there were factors that were not currently available to them that they would like to have available. All factors identified were biological or molecular in nature. Of these, HER2 was the most common factor.

Cathepsin D and bcl-2 were identified as being unavailable.

Some 42% of clinicians stated that there were one or more factors other than those listed that they used when deciding on systemic adjuvant therapy for newly diagnosed women with regionally localised breast cancer – the two most quoted factors were patient preference (44 instances) and co-morbidity (14 instances).

There was a striking variation in the numbers of patients being offered adjuvant therapy – this was particularly so for chemotherapy. Clinicians were asked what proportion of newly diagnosed patients received adjuvant chemotherapy, adjuvant hormonal therapy or a combination of both. This question was not responded to by a number of clinicians – this was more common in units than centres: 18% for adjuvant chemotherapy, 15% for adjuvant hormone therapy and 25% for a combination of adjuvant hormone and adjuvant chemotherapy. Amongst clinicians who did complete the question, there was a wide range of responses for each treatment option. The most commonly cited proportions for adjuvant chemotherapy were 30–59%, accounting for 70% of responding centres. The most commonly cited proportions for adjuvant hormone therapy were 60–100%, accounting for 85% of responding centres.

Protocols (45 analysed)

Of the 79% of clinicians who indicated that there was a written protocol in their unit for making decisions on adjuvant systemic therapy, 53% stated that this protocol was based on a PI. The NPI was the most commonly cited index. Other indices mentioned were the Manchester Scale (by three SUs) and the Mount Vernon Index (by one SU). No other index was cited. The majority of remaining responses indicated that a number of factors were used to create a 'unit-specific index'. A number of these were derivatives of the NPI, whereby other factors were added to the original NPI.

Protocols were based in the first instance upon pathological nodal status, tumour grade and tumour size, those factors identified by clinicians during the survey as being the most clinically important when selecting adjuvant systemic therapy for women with early breast cancer. Approximately half of the 45 (22/45, 49%) units combined these factors to generate the NPI, whereas the remaining 23 units (23/45, 51%) recognised the importance of such factors but

chose to develop treatment protocols without the use of a single index score.

Use of the NPI

Amongst 22 units using the NPI, there were five different ways of using the index; the number of prognostic groupings into which women could be classified on the basis of their NPI score ranged from five down to three.

- Ten of the 22 protocols (45%) advocate using the NPI to split patients into five subgroups.¹⁶⁹
- Five protocols (23%) divide patients into three groups using the NPI, each using different index cut-off values. Only one uses the original three NPI groupings.¹⁹³
- One protocol uses integer scores to group women.¹⁹⁶

It is striking that there are so many variations on this one simple index, some of which use their own unvalidated cut-off points.

Age/menopausal status

Many centres use age or menopausal status in addition to the NPI. Once again there was marked variation in practice. Of 20 centres using NPI, 14 used age and six menopausal status. For those using age, there were five different categories used.

Hormone receptor status and other factors

ER status was required in all 22 protocols using the NPI, reflecting findings from the clinicians' survey that ER status is also one of the most important clinical factors when determining suitability for adjuvant systemic therapy. In addition to ER status, nine protocols also consider the PR status of women within each prognostic group to be important, and two recommend that HER2 should also be considered if available. One unit suggested that lymphovascular invasion and performance status should be included.

Centres not using the NPI

Although 23 (51%) of the 45 available protocols choose not to use the NPI to combine formally tumour size, pathological nodal status and grade information, 21 of these still consider the three variables when making a decision on adjuvant therapy. One protocol considers only two of the three to be important (nodal status and grade) and the other appears to consider hormonal status only.

Age and/or menopausal status are considered to be important in 22/23 protocols. Women are

assessed according to their menopausal status in 14 and age is considered in the remaining eight, though age classification varied by protocol. All 23 protocols consider ER status as an important factor when selecting women for adjuvant therapy. PR status is also considered by six units, HER2 by one and 10 protocols include lymphovascular invasion as a criterion to be considered.

Guidelines for prescribing adjuvant therapy

Tamoxifen

There were 45 protocols that included guidelines about when to prescribe adjuvant hormone therapy. There were no less than 15 different guidelines. Only two were commonly used, one in 13 protocols (prescribe for all ER- or PR-positive women) and one in 10 protocols (prescribe for all ER- or PR-positive women with the exception of those identified as having a favourable prognosis – but they define favourable differently).

Two protocols recommend tamoxifen in all ER/PR-positive women but they also support the use of tamoxifen for all women with the exception of those who are ER- or PR-negative. It is assumed therefore that ER- or PR-poor women at these two units may be considered for adjuvant hormone therapy.

Seventeen of the remaining 20 protocols include broadly similar guidelines in that the majority advocate use of tamoxifen for ER- or PR-positive pre- and postmenopausal women. These 17 differ, however, in that they also contain treatment recommendations for additional cohorts of women considered likely to benefit from adjuvant hormone therapy and/or for women for whom therapy may be considered unnecessary or ineffective.

Four units (who appear to have based their protocols on the 1994 *BMJ* papers by Richards and colleagues^{285,286}) use tamoxifen for all patients irrespective of hormone receptor status) with the exception of those presenting with favourable tumour characteristics.

Two protocols suggest tamoxifen for ER-positive women with an NPI score of 2.4–3.3, for all women with an NPI score of 3.4–4.3 and for no women with an NPI score exceeding 4.4. At one centre tamoxifen is administered to all women with an NPI score <3 and only ER-positive women with an NPI score <3, whereas

at another tamoxifen is indicated for all Stage I, II and III women. One protocol suggests that tamoxifen may be beneficial for some very low-risk women (NPI score <2.4), should be given to all postmenopausal women and ER-positive premenopausal women with low or moderate risk and is not beneficial for any woman with a high or very high risk.

One centre recommends adjuvant hormone therapy for non-ER-negative premenopausal women and all postmenopausal women below the age of 65 years, with the exception of those having an extremely favourable prognosis (size <1 cm, Grade I, Stage I tumours). Another supports the use of tamoxifen for all ER-positive premenopausal women without a favourable prognosis (defined as size <2 cm, Grade I, Stage I tumours) and all postmenopausal women.

Overall, although there are common themes, there are marked variations in the ways in which women are offered adjuvant hormone therapy.

Adjuvant chemotherapy

This information was available in 18 protocols. Seven (39%) of the protocols suggest chemotherapy for all patients (irrespective of age and hormone receptor status) with an NPI score of >5.4 . A further nine (50%) recommend administering adjuvant chemotherapy to all women with an NPI score >5.4 provided they are below a certain age threshold – but use a variety of ages. Of the remaining two protocols, one recommends chemotherapy for all women except those who are postmenopausal and ER or PR positive. Here chemotherapy should be discussed. The other suggests chemotherapy for all women except those who are ER positive and between the ages of 60 and 69 years.

Seven (39%) protocols continue to indicate use of chemotherapy for all patients (irrespective of age and hormone receptor status) with NPI scores of 4.4–5.4. A further seven suggest chemotherapy for all women with NPI scores falling within the same range provided they are below 70 or 69 years old. Among the remaining four protocols, one selects out all women between the ages of 50 and 75 years who are not ER negative, two recommend chemotherapy for all women except those who are postmenopausal and ER or PR positive, for whom chemotherapy should be discussed, and one chose not to support chemotherapy for any woman aged 60–69 years or for women with ER-positive tumours.

Just five (28%) protocols advocate the use of adjuvant chemotherapy in all women (irrespective of age and hormone receptor status) with an NPI score between 3.4 and 4.4. One other recommends treatment for any woman below the age of 70 years with such NPI scores. Of the four centres offering adjuvant chemotherapy to women below the age of 50 years, two also discuss the benefits to be gained by treatment of ER-negative women aged 50–70 years. Three protocols provide treatment for ER-negative women below the age of 69 years and will discuss chemotherapy for ER-positive women only below the age of 50 years.

One protocol advocates adjuvant chemotherapy for all women with the exception of postmenopausal ER- or PR-positive women for whom it recommends treatment should be discussed, whereas another considers ER- or PR-positive women over the age of 35 years to be unsuitable for such treatment. Two protocols suggest that chemotherapy should be offered to all women below the age of 70 years with an NPI score of ≥ 4.4 . However, once below this threshold score, emphasis is placed on discussing rather than on offering chemotherapy to such women. Finally, implementation of guidelines from one protocol would see no women categorised as having this ‘moderate’ prognosis being offered adjuvant chemotherapy.

Within the NPI range 2.4–3.4, half (nine) of the protocols contain guidelines, which vary further with NPI scores. At four centres chemotherapy is not routinely provided for any women with an NPI score between 2.41 and 2.5. Its use is supported in all premenopausal and ER-negative postmenopausal women with NPI scores between 2.5 and <3.4 , and it is available to all women with an NPI score of 3.4. Two protocols specify that adjuvant chemotherapy should not be given to women less than 70 years of age, with NPI scores between >2.4 and ≤ 3.3 and should be discussed for all women less than 70 years of age with an NPI score equal to 3.4.

According to the one protocol using integer score values to create prognostic groups, no woman with an NPI score between >2.4 and 3 is prescribed chemotherapy, whereas all women with a score >3 are offered treatment. Of the remaining nine protocols, one supports the use of chemotherapy only in ER-negative women below the age of 35 years, another will discuss therapy only for ER-negative women below the age of 70 years, and a third does not advocate chemotherapy for any women except those who are below the age of

50 years and who are ER negative – for these women treatment should be discussed. Finally, six centres (four specifying an upper age limit of 69 years) do not support the use of chemotherapy for any woman falling into this prognostic group.

Only one protocol recommends the use of adjuvant chemotherapy for women with an NPI score of <2.4. These women are required to have an ER-negative tumour and to be below the age of 35 years.

Where the NPI is not used, adjuvant treatment is largely determined by nodal status, with tumour size, grade and ER status playing major roles. Once again there are major variations in the recommendations of specific guidelines (see Chapter 7, *Tables 21 and 22*).

Comparison of protocols using NPI and those not using NPI

For women with a prognosis described as ‘excellent’ (i.e. NPI score of ≤ 2.4), published literature suggests that survival may not be significantly different from that of an age-matched population, and the probable benefit from adjuvant chemotherapy is negligible. *Table 23* shows that these findings are reflected in the guidelines put forward for ‘excellent’ prognosis women by the protocols using the NPI. Between 10 and 55% of the protocols not using the NPI recommend considering or offering adjuvant chemotherapy to premenopausal women in the same prognostic group. Treatment is, however, guided towards women below the age of 35 years, for whom the prognosis appears particularly unfavourable regardless of tumour characteristics.

Amongst the protocols using the NPI, chemotherapy guidelines for women with NPI scores between 2.41 and 3.4 on the whole reflect the fact that these women are generally considered to have a ‘good’ prognosis. Uncertainty does surround the decision about whether to consider or offer adjuvant chemotherapy to ER- or PR-negative premenopausal women with a ‘good’ prognosis. *Table 23* shows that approximately half of the protocols suggest that treatment may be beneficial whereas the other half do not.

For the NPI-based protocols, the total number recommending women be considered for or offered adjuvant chemotherapy remains constant across different grade/node combinations within each prognostic group. No difference exists between the number of protocols offering chemotherapy to women with a positive nodal

status and an NPI score between 2.41 and 3.4 and those women having the same NPI score but achieved without nodal involvement (e.g. 50 and 67% for premenopausal ER-positive and ER-negative women, respectively). In contrast, the requirement for adjuvant chemotherapy is considered by the protocols not using the NPI to vary within each prognostic group according to pathological nodal status, grade or size. *Table 23* shows more clearly the importance that these protocols place upon nodal status when considering adjuvant chemotherapy. As identified in Chapter 7, almost all advocate adjuvant chemotherapy for node-positive women regardless of tumour grade or size. The difference between the proportion of NPI- and non-NPI-based protocols recommending adjuvant chemotherapy for node-positive women is greatest within the ‘good’ prognostic group (NPI score 2.41–3.4). For example, only 6% of protocols using the NPI would consider adjuvant chemotherapy for a node-positive, ER/PR-positive, postmenopausal woman with a Grade I, Stage I, size ≤ 1 cm tumour. The corresponding figure amongst units not using the NPI is 86%. For NPI scores exceeding 3.41, the proportions are not dissimilar. The uncertainty as to how much post-menopausal ER- or PR-positive women can benefit from adjuvant chemotherapy also seems to be shared by both types of protocol across each prognostic group.

It appears that guidelines from protocols using the NPI result in fewer excellent or good prognosis women being considered for or offered adjuvant chemotherapy than do guidelines from non-NPI-based protocols. For high-risk or poor-prognosis women, the opposite is true. More women are likely to be treated as a consequence of guidelines formulated for NPI-based protocols than would be the case for women treated at centres not using the NPI to select candidates for adjuvant chemotherapy.

Conclusions of survey

The survey confirmed pathological nodal status, tumour grade, tumour size and ER status as the most clinically important factors used in the UK when selecting women with early breast cancer for adjuvant systemic therapy. Clinicians reported that such factors were readily available to them, and this was independently verified by histopathologists managing breast cancer pathology within the same centres.

Factors identified as useful by clinicians included PR and HER2, but uncertainty surrounding the extent to which these factors are available suggests

the need for closer communication between pathology laboratories and clinicians. At centres where histopathologists confirmed the availability of PR, only 82% of clinicians said they had access to such information. The corresponding figure for HER2 was even lower at 54%.

Unit protocols allowed us to explore patterns of use of adjuvant therapy resulting from the use of prognostic and predictive information. Reviewing these protocols revealed that although centres appear to be using the same prognostic and predictive factors when selecting women to receive adjuvant therapy, much variation in clinical practice exists between UK breast cancer centres.

Such variation occurs at two levels. First, centres with protocols based on the NPI appear to differ from centres not using the single index score. Second, within NPI and non-NPI users, between-centre variability exists in guidelines for use of adjuvant therapy for women for whom the benefits are uncertain, such as postmenopausal women and women with a negative nodal status and low to intermediate grade tumours. Consensus amongst units appears to be greatest when selecting women for adjuvant hormone therapy – the decision being based primarily on ER or PR status rather than combinations of a number of factors. Guidelines as to who should receive adjuvant chemotherapy, however, have been shown to be much less uniform.

Health economic evaluation

Although it is widely acknowledged that early breast cancer is a clinically heterogeneous disease, and that women with different clinical characteristics have differing prognoses and hence will benefit differently from adjuvant systemic therapy, the variations in clinical practice observed in the survey carried out as part of this research (see Chapter 7) suggest that many clinicians remain uncertain about the benefits to be gained by administering adjuvant chemotherapy to certain groups of women.

The modelling presented in Chapter 9 was developed to demonstrate how a decision-making framework could have the potential to assist clinicians and policy makers in dealing with this uncertainty. By combining methodologies used in determining prognosis with those used in health economic evaluation, it was possible to simulate the effectiveness (in terms of survival and quality-adjusted survival) and the cost-effectiveness

associated with the decision to treat individual women or groups of women with different prognostic characteristics. Initially, it was intended that the systematic reviews of prognostic models and prognostic and predictive factors undertaken as part of this study would inform the different prognostic groups for which results would be generated. However, the acquisition of a fairly large dataset containing patient-level data on prognostic factors, treatments and outcomes of women diagnosed with early breast cancer made it possible to estimate directly a regression-based risk equation. The equation generated estimates of the baseline breast cancer hazard and 10-year risk of breast cancer death for individual women or groups of women with all possible combinations of the prognostic factors identified. Including all-cause mortality alongside breast cancer mortality, and using data from the early breast cancer overviews on likely treatment effect to adjust the baseline hazard downwards, it was then possible to generate estimates of lifetime survival with and without adjuvant chemotherapy. Extending the model to incorporate treatment costs and the costs and reductions in health-related quality of life associated with the likely side-effects of treatment and relapse allowed simulation of the costs and QALYs with and without adjuvant chemotherapy for all possible combinations of prognostic factors and the incremental cost-effectiveness of treatment.

As predicted, results from the model showed that effectiveness and cost-effectiveness of adjuvant systemic therapy have the potential to vary substantially depending on prognosis. For some women therapy may prove very effective and cost-effective, whereas for others it may actually prove detrimental (i.e. the reductions in health-related quality of life outweigh any survival benefit).

It is suggested that the outputs from models constructed using the methods described in Chapter 9 have the potential to be useful at two decision-making levels: first at the patient level, where a clinician must determine whether net benefits can be obtained from administering adjuvant therapy for any presenting woman, and second at the policy-making level, where decision-makers must be able to determine the total costs and outcomes associated with different treatment protocols which are based to differing degrees on prognostic information. To assist clinicians, various ways of presenting the model's outputs were examined, including a risk table format where clinicians could look up a patient's prognostic factors to determine the likely benefits

(survival and quality-adjusted survival) from administering therapy. For policy makers, it was demonstrated that the model's output could be used to evaluate the cost-effectiveness of different

treatment protocols based on prognostic information. The framework should also be valuable in evaluating the likely impact and cost-effectiveness of new potential prognostic factors.



Acknowledgements

We thank Dr Jin Ling Tang for translating a paper from Chinese. We also thank Sarah Moore for her help in organising the articles for the systematic review of reviews of prognostic factors.

Contribution of authors

Helen Campbell (Senior Researcher) analysed and wrote the section on UK adjuvant therapy treatment protocols, reviewed and reported on the studies identified as assessing the cost-effectiveness of using prognostic information to inform treatment decisions and assisted in the development and reporting of the health economic modelling framework presented in Chapter 9. Adrian Harris (Consultant Medical Oncologist) suggested the main prognostic and predictive pathways to investigate, contributed to the review of the literature of major prognostic studies and reviews, wrote the section on prognostic factors and reviewed the entire manuscript. He also contributed to the clinical practice assessment of how markers were applied in Oxford and was involved in analysis of prognostic profiles on this set of patients. He was also involved in the design of the questionnaire about markers used in practice. Andrew Briggs (Lindsay Professor of Health Policy and Economic Evaluation) was involved in supporting all of the health economic components of the review. This included designing the economic modelling, undertaking statistical analysis, drafting and commenting on the economic chapters and handling responses from referees on these chapters. Alistair Gray (Professor of Health Economics) oversaw design and supervision of economic analysis and contributed to economic analyses and preparation of presentations, papers and final report. Chris Williams (Consultant Medical Oncologist) took part in the conception and design of the project and contributed to

developing the searches, selection of papers, abstraction of data, conduct of the questionnaire, analysis of the results and preparation of the final report. Doug Altman (Professor of Statistics in Medicine) took part in the design of the project and contributed to selection of papers, abstraction of data, development of the survey of clinical practice and preparation of the final report. He took primary responsibility for the review of the methodological quality of prognostic studies and the review of published prognostic models. Julie Glanville (Associate Director of CRD) prepared and tested the search strategies for the review, documented the searches and wrote up the search process for the report. Mike Clarke (Professor of Clinical Epidemiology) took part in the conception and design of the project. He contributed to developing the searches, selection of papers, abstraction of data, preparation of the questionnaire, analysis of the results and preparation of the final report. He had special responsibility for the systematic review of reviews of studies of prognostic factors (Chapter 4) and the section on the methods for the systematic reviews (Chapter 2). Susan Brunskill (Senior Information Scientist) undertook the questionnaire and study selection components of the project. She contributed to the questionnaire design, analysed the results and was involved in the preparation of the questionnaire chapter. She contributed to the development of the searches, ran the search strategies, screened references for eligibility and drafted the study selection sections for each chapter. Kathy Johnston (Economic Adviser) took part in the conception and design of the project, and contributed to developing the searches, selection of papers and abstraction of data. Mark Lodge (Convenor, Cochrane Cancer Network) helped develop the concept, supported development of the searches and helped coordinate the preparation of the report.



References

1. Jemal A, Thomas A, Murray T, Thun M. Cancer statistics, 2002. *CA Cancer J Clin* 2002;**52**:23–47.
2. DOH 2002 NHS Cancer Screening Programmes. [Cancer Screening Programmes]. URL: www.cancerscreening.nhs.uk. Accessed 6 September 2002.
3. Bray F, Sankila R, Ferlay J, Parkin DM. Estimates of cancer incidence and mortality in Europe in 1995. *Eur J Cancer* 2002;**38**:99–166.
4. American Cancer Society. [Breast cancer facts and figures]. URL: www.cancer.org. Accessed 6 September 2002.
5. Forbes JF. Breast cancer. In Williams CJ, editor. *Evidence-based oncology*. London: BMJ Books; 2003.
6. Brown P. UK death rates from breast cancer fall by a third. *BMJ* 2000;**321**:849.
7. Early Breast Cancer Trialists' Collaborative Group. Effects of adjuvant tamoxifen and of cytotoxic therapy on mortality in early breast cancer: an overview of 61 randomised trials among 28,896 women. *N Engl J Med* 1988;**319**:1681–92.
8. Early Breast Cancer Trialists' Collaborative Group. *Treatment of early breast cancer, volume 1: worldwide evidence 1985–1990*. Oxford: Oxford University Press; 1990.
9. Early Breast Cancer Trialists' Collaborative Group. Systematic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy: 133 randomised trials involving 31,000 recurrences and 24,000 deaths among 75,000 women. *Lancet* 1992;**339**:1–15, 71–85.
10. Early Breast Cancer Trialists' Collaborative Group. Effects of radiotherapy and surgery in early breast cancer: an overview of the randomised trials. *N Engl J Med* 1995;**333**:1444–55.
11. Early Breast Cancer Trialists' Collaborative Group. Ovarian ablation in early breast cancer: overview of the randomised trials. *Lancet* 1996;**348**:1189–96.
12. Early Breast Cancer Trialists' Collaborative Group. Polychemotherapy for early breast cancer: an overview of the randomised trials. *Lancet* 1998;**352**:930–42.
13. Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer: an overview of the randomised trials. *Lancet* 1998;**351**:1451–67.
14. Early Breast Cancer Trialists' Collaborative Group. Effect of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomized trials. *Lancet* 2005;**365**:1687–717.
15. Haagensen CD, Stout AP. Carcinoma of the breast: criteria of operability. *Ann Surg* 1969;**118**:859–70, 1032–51.
16. Crile G. In Stoll BA, editor. *Breast cancer management: early and late*. London: Heinemann; 1977.
17. Saeki T. What is the Japanese consensus on adjuvant chemotherapy in breast cancer? *Breast Cancer* 2003;**10**:15–20.
18. Ravdin PM. How can prognostic and predictive factors in breast cancer be used in a practical way today? *Recent Results Cancer Res* 1998;**152**:86–93.
19. St Gallen Consensus Conference. *J Clin Oncol* 2001;**19**:3817–27.
20. NIH Consensus Statement 114. Adjuvant therapy for breast cancer. November 1–3 2000;**17**:1–23.
21. American Society of Clinical Oncology Expert Panel. Clinical practice guidelines for the use of tumour markers in breast and colorectal cancer: Report of the American Society of Clinical Oncology Expert Panel. *J Clin Oncol* 1996;**14**:2843–77.
22. American Society of Clinical Oncology. 2000 update of recommendations for the use of tumour markers in breast and colorectal cancer: clinical practice guidelines of the American Society of Clinical Oncology. *J Clin Oncol* 2001;**19**:1865–78.
23. Borghouts JA, Koes BW, Bouter LM. The clinical course and prognostic factors of non-specific neck pain: a systematic review. *Pain* 1998;**77**:1–13.
24. Carson CA, Fine MJ, Smith MA, Weissfeld LA, Huber JT, Kapoor WN. Quality of published reports of the prognosis of community-acquired pneumonia. *J Gen Intern Med* 1994;**9**:13–19.
25. Cole MG, Bellavance F. Depression in elderly medical inpatients: a meta-analysis of outcomes. *Can Med Assoc J* 1997;**157**:1055–60.
26. Davies HT, Crombie IK. Bias in cohort studies. *Hosp Med* 2000;**61**:133–5.
27. Delgado-Rodriguez M. [Biases in the study of prognostic factors]. *Med Clin* 1999;**112**(Suppl 1): 51–8.
28. Laupacis A, Wells G, Richardson WS, Tugwell P. Users' guides to the medical literature. V. How to

- use an article about prognosis. Evidence-Based Medicine Working Group. *JAMA* 1994;**272**:234–7.
29. Levine MN, Browman GP, Gent M, Roberts R, Goodyear M. When is a prognostic factor useful? A guide for the perplexed. *J Clin Oncol* 1991; **9**:348–56.
 30. Look MP. Pooled analysis of uPA and PAI-1 for prognosis in primary breast cancer patients. EORTC Receptor and Biomarker Study Group. *Int J Biol Markers* 2000;**15**:70–2.
 31. Lucassen PL, Assendelft WJ, van Eijk JT, Gubbels JW, Douwes AC, van Geldrop WJ. Systematic review of the occurrence of infantile colic in the community. *Arch Dis Child* 2001;**84**:398–403.
 32. Marras LC, Geerts WH, Perry JR. The risk of venous thromboembolism is increased throughout the course of malignant glioma: an evidence-based review. *Cancer* 2000;**89**:640–6.
 33. Melnikow J, Nuovo J, Willan AR, Chan BK, Howell LP. Natural history of cervical squamous intraepithelial lesions: a meta-analysis. *Obstet Gynecol* 1998;**92**:727–35.
 34. O'Malley PG, Taylor AJ, Jackson JL, Doherty TM, Detrano RC. Prognostic value of coronary electron-beam computed tomography for coronary heart disease events in asymptomatic populations. *Am J Cardiol* 2000;**85**:945–8.
 35. Randolph AG, Guyatt GH, Richardson WS. Prognosis in the intensive care unit: finding accurate and useful estimates for counseling patients. *Crit Care Med* 1998;**26**:767–72.
 36. Steinberg EP, Eknayan G, Levin NW, Eschbach JW, Golper TA, Owen WF, *et al.* Methods used to evaluate the quality of evidence underlying the National Kidney Foundation–Dialysis Outcomes Quality Initiative Clinical Practice Guidelines: description, findings, and implications. *Am J Kidney Dis* 2000;**36**:1–11.
 37. Vance RP, Elevitch FR. College of American Pathologists Foundation Conference VIII on patient-centered pathology practice–outcomes and accountability. Overview, glossary, and bibliography. *Arch Pathol Lab Med* 1997;**121**:1151–62.
 38. Yarbro JW, Page DL, Fielding LP, Partridge EE, Murphy GP. American Joint Committee on Cancer prognostic factors consensus conference. *Cancer* 1999;**86**:2436–46.
 39. Zarbo RJ. The oncologic pathology report. Quality by design. *Arch Pathol Lab Med* 2000;**124**:1004–10.
 40. Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. *Br J Cancer* 1994;**69**:979–85.
 41. McGuire WL. Breast cancer prognostic factors: evaluation guidelines. *J Nat Cancer Inst* 1991; **83**:154–5.
 42. Altman DG. Systematic reviews of evaluations of prognostic variables. *BMJ* 2001;**323**:224–8.
 43. Altman DG. Systematic reviews of evaluations of prognostic variables. In Egger M, Davey SG, Altman DG, editors. *Systematic reviews in health care. Meta-analysis in context*. London: BMJ Books; 2001. pp. 228–47.
 44. Kernan WN, Feinstein AR, Brass LM. A methodological appraisal of research on prognosis after transient ischemic attacks. *Stroke* 1991;**22**:1108–16.
 45. Hudak PL, Cole DC, Haines AT. Understanding prognosis to improve rehabilitation: the example of lateral elbow pain. *Arch Phys Med Rehabil* 1996; **77**:586–93.
 46. Marx BE, Marx M. Prognosis of idiopathic membranous nephropathy: a methodologic meta-analysis. *Kidney Int* 1997;**51**:873–9.
 47. Straus SE, McAlister FA. A clinician's guide to journal articles about prognosis. *ACP J Club* 1999; **130**:A13–15.
 48. Altman DG. Systematic reviews of evaluations of prognostic variables. *BMJ* 2001;**323**:224–8.
 49. Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, van der Meulen JHP, *et al.* Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;**282**:1061–6.
 50. Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. *Br J Cancer* 1994;**69**:979–85.
 51. Fayers PM, Machin D. Sample size: how many patients are necessary? *Br J Cancer* 1995;**72**:1–9.
 52. Schmoor C, Sauerbrei W, Schumacher M. Sample size considerations for the evaluation of prognostic factors in survival analysis. *Stat Med* 2000;**19**:441–52.
 53. Bentzen SM. Prognostic factor studies in oncology: osteosarcoma as a clinical example. *Int J Radiat Oncol Biol Phys* 2001;**49**:513–18.
 54. Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and related biases. *Health Technol Assess* 2000;**4**(10).
 55. Simon R. Evaluating prognostic factor studies. In Gospodarowicz MK, Henson DE, Hutter RVP, *et al.*, editors. *Prognostic factors in cancer*. New York: Wiley-Liss; 2001.
 56. Shaheen NJ, Crosby MA, Bozyski EM, Sandler RS. Is there publication bias in the reporting of cancer risk in Barrett's esophagus? *Gastroenterology* 2000;**119**:333–8.
 57. Popat S, Matakidou A, Houlston RS. Thymidylate synthase expression and prognosis in colorectal cancer: a systematic review and meta-analysis. *J Clin Oncol* 2004;**22**:529–36.

58. Brocklehurst P, French R. The association between maternal HIV infection and perinatal outcome: a systematic review of the literature and meta-analysis. *Br J Obstet Gynaecol* 1998;**105**: 836–48.
59. Altman DG, Lyman GH. Methodological challenges in the evaluation of prognostic factors in breast cancer. *Breast Cancer Research and Treatment* 1998;**52**:289–303.
60. Riley RD, Abrams KR, Sutton AJ, Lambert PC, Jones DR, Heney D, *et al.* Reporting of prognostic markers: current problems and development of guidelines for evidence-based practice in the future. *Br J Cancer* 2003;**88**:1191–8.
61. Wu YW, Colford JM Jr. Chorioamnionitis as a risk factor for cerebral palsy. A meta-analysis. *JAMA* 2000;**284**:1417–24.
62. Brundage MD, Davies D, Mackillop WJ. Prognostic factors in non-small cell lung cancer. A decade of progress. *Chest* 2002;**122**:1037–57.
63. Parmar M, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Stat Med* 1998;**17**:2815–34.
64. Cohen J. The cost of dichotomization. *Appl Psychol Meas* 1983;**7**:249–53.
65. Morgan TM, Elashoff FA. Effect of categorizing a continuous covariate on the comparison of survival time. *J Am Stat Assoc* 1986;**81**:917–21.
66. Altman DG, Lausen B, Sauerbrei W, Schumacher M. Dangers of using 'optimal' cutpoints in the evaluation of prognostic factors. *J Nat Cancer Inst* 1994;**86**:829–35.
67. Vollmer RT, Seigler HF. Using a continuous transformation of the Breslow thickness for prognosis in cutaneous melanoma. *Am J Clin Pathol* 2001;**115**:205–12.
68. Hilsenbeck SG, Clark GM, McGuire WL. Why do so many prognostic factors fail to pan out? *Breast Cancer Res Treat* 1992;**22**:197–206.
69. Sauerbrei W, Royston P, Bojar H, Schmoor C, Schumacher M, the German Breast Cancer Study Group (GBSG). Modelling the effects of standard prognostic factors in node-positive breast cancer. *Stat Med* 1999;**79**:1752–60.
70. Buettner P, Garbe C, Guggenmoos-Holzmann I. Problems in defining cutoff points of continuous prognostic factors: example of tumor thickness in primary cutaneous melanoma. *J Clin Epidemiol* 1997;**50**:1201–10.
71. Steels E, Paesmans M, Berghmans T, Branle F, Lemaitre F, Mascaux C, *et al.* Role of p53 as a prognostic factor for survival in lung cancer: a systematic review of the literature with a meta-analysis. *Eur Respir J* 2001;**18**:705–19.
72. Greenland S. A critical look at some popular meta-analytic methods. *Am J Epidemiol* 1994;**140**:290–6.
73. Jüni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ* 2001;**323**:42–6.
74. Altman DG, De Stavola BL, Love SB, Stepniwska KA. Review of survival analyses published in cancer journals. *Br J Cancer* 1995;**72**:511–18.
75. Andersen PK. Survival analysis 1982–1991: the second decade of the proportional hazards regression model. *Stat Med* 1991;**10**:1931–41.
76. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. REporting recommendation for tumour MARKer prognostic studies (REMARK). *Br J Cancer* 2005;**93**:387–91.
77. Cuzick J. Interaction, subgroup analysis and sample size. *IARC Sci Publ* 1999;109–21.
78. Brookes ST, Whitley E, Peters TJ, Mulheran PA, Egger M, Davey SG. Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. *Health Technol Assess* 2001;**5**(33).
79. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat Med* 2002;**21**:2917–30.
80. Royston P, Sauerbrei W, Ritchie A. Is treatment with interferon-alpha effective in all patients with metastatic renal carcinoma? A new approach to the investigation of interactions. *Br J Cancer* 2004;**90**:794–9.
81. Schmitz-Dräger BJ, Goebell PJ, Ebert T, Fradet Y. p53 immunohistochemistry as a prognostic marker in bladder cancer. *Eur Urol* 2000;**38**:691–700.
82. Burke HB, Henson DE. Specimen banks for cancer prognostic factor research. *Arch Pathol Lab Med* 1998;**122**:871–4.
83. Mirza AN, Mirza NQ, Vlastos G, Singletary SE. Prognostic factors in node-negative breast cancer: a review of studies with sample size more than 200 and follow-up more than 5 years. *Ann Surg* 2002;**235**:10–26.
84. Porter-Jordan K, Lippman ME. Overview of the biologic markers of breast cancer. *Hematol Oncol Clin North Am* 1994;**8**:73–100.
85. Fox SB, Smith K, Hollyer J, Greenall M, Hastrich D, Harris AL. The epidermal growth factor receptor as a prognostic marker: results of 370 patients and review of 3009 patients. *Breast Cancer Res Treat* 1994;**29**:41–9.
86. Klijn JG, Berns PM, Schmitz PI, Foekens JA. The clinical significance of epidermal growth factor receptor (EGF-R) in human breast cancer: a review on 5232 patients. *Endocr Rev* 1992;**13**:3–17.

87. Révillion F, Bonnetterre J, Peyrat JP. ERBB2 oncogene in human breast cancer and its clinical significance. *Eur J Cancer* 1998;**34**:791–808.
88. Ross JS. The HER-2/neu gene and protein in breast cancer: methods of detection and clinical significance. *J Clin Ligand Assay* 2000;**23**:21–32.
89. Nunes RA, Harris LN. The HER2 extracellular domain as a prognostic and predictive factor in breast cancer. *Clin Breast Cancer* 2002;**3**:125–35.
90. Henderson IC, Patek AJ. The relationship between prognostic and predictive factors in the management of breast cancer. *Breast Cancer Res Treat* 1998;**52**:261–88.
91. Deming SL, Nass SJ, Dickson RB, Trock BJ. C-myc amplification in breast cancer: a meta-analysis of its occurrence and prognostic relevance. *Br J Cancer* 2000;**83**:1688–95.
92. Berns EMJJ, Klijn JGM, van Staveren IL, Portengen H, Noordegraaf E, Foekens JA. Prevalence of amplification of the oncogenes c-myc, HER2/neu, and int-2 in one thousand human breast tumours: correlation with steroid receptors. *Eur J Cancer* 1992;**28**:697–700.
93. Wenger CR, Clark GM. S-phase fraction and breast cancer – a decade of experience. *Breast Cancer Res Treat* 1998;**51**:255–65.
94. Daidone MG, Silvestrini R. Prognostic and predictive role of proliferation indices in adjuvant therapy of breast cancer. *J Nat Cancer Inst Monogr* 2001;**30**:27–35.
95. Sunderland MC, McGuire WL. Prognostic indicators in invasive breast cancer. *Surg Clin North Am* 1990;**70**:989–1004.
96. Barbareschi M. Prognostic value of the immunohistochemical expression of p53 in breast carcinomas: a review of the literature involving over 9000 patients. *Appl Immunohistochem* 1996;**4**:106–16.
97. Elledge RM, Allred DC. Prognostic significance of p53 gene alterations in node negative breast cancer. *Breast Cancer Res Treat* 1998;**52**:79–98.
98. Pharoah PDP, Day NE, Caldas C. Somatic mutations in the p53 gene and prognosis in breast cancer: a meta-analysis. *Br J Cancer* 1999;**80**:1968–73.
99. Silvestrini R, Benini E, Daidone MG, Veneroni S, Boracchi P, Cappelletti V, et al. p53 as an independent prognostic marker in lymph node-negative breast cancer patients. *J Nat Cancer Inst* 1993;**85**:965–70.
100. Elledge RM, Fuqua SA, Clarck GM. Prognostic significance of p53 gene alterations in node negative breast cancer. *Breast Cancer Res Treat* 1993;**26**:225–35.
101. Iacopetta B, Grieru F, Powell B. Analysis of p53 gene mutation by polymerase chain reaction – single strand confirmation polymorphism provides independent prognostic information in node-negative breast cancer. *Clin Cancer Res* 1998;**4**:1597–602.
102. Domagala W, Welcker M, Chosia M. p21/WAF/Cip1 expression in invasive ductal breast carcinoma: relationship to p53, proliferation rate, and survival at 5 years. *Virchows Arch* 2001;**439**:132–40.
103. Zhang G-J, Kimijima I, Tsuchiya A, Abe R. The role of bcl-2 expression in breast carcinomas (review). *Oncol Rep* 1998;**5**:1211–16.
104. Ferrandina G, Scambia G, Bardelli F, Benedetti Panici P, Mancuso S, Messori A. Relationship between cathepsin-D content and disease-free survival in node-negative breast cancer patients: a meta-analysis. *Br J Cancer* 1997;**76**:661–6.
105. Thorpe SM, Rochefort H, Garcia M, Freiss G, Christensen IJ, Khalaf S, et al. Association between high concentrations of Mr 52,000 D and poor prognosis in primary human breast cancer. *Cancer Res* 1989;**49**:6008–14.
106. Tandon AK, Chamness GC, Chirgwin JM, McGuire WL. Cathepsin D and prognosis in breast cancer. *N Engl J Med* 1990;**322**:297–302.
107. Isola J, Weitz S, Visakorpi T, Holli K, Shea R, Khabbaz N, et al. Cathepsin D expression detected by immunohistochemistry has independent prognostic value in axillary node-negative breast cancer. *J Clin Oncol* 1993;**11**:36–43.
108. Namer M, Ramaioli A, Fontana X. Prognostic value of total cathepsin D in breast tumours: a possible role in selection of chemoresistant patients. *Breast Cancer Res Treat* 1991;**19**:85–93.
109. Foekens JA, Look MP, Bolt-de Vries J, Meijer-van Gelder ME, van Putten WL, Klijn JG. Cathepsin-D in primary breast cancer: prognostic evaluation involving 2810 patients. *Br J Cancer* 1999;**79**:300–307.
110. Henry JA, McCarthy AL, Angus B. Prognostic significance of the estrogen regulated protein cathepsin D in breast cancer: an immunohistochemical study. *Cancer* 1990;**65**:265–71.
111. Look MP, Foekens JA. Clinical relevance of the urokinase plasminogen activator system in breast cancer. *Acta Pathol, Microbiol Immunol Scand* 1999;**107**:150–9.
112. Harbeck N, Kruger A, Sinz S. Clinical relevance of the plasminogen activator inhibitor type 1 – a multifaceted proteolytic factor. *Onkologie* 2001;**24**:238–44.
113. Look MP, van Putten WL, Duffy MJ, Harbeck N, Christensen IJ, Thomssen C, et al. Pooled analysis of prognostic impact of urokinase-type plasminogen activator and its inhibitor PAI-1 in 8377 breast cancer patients. *J Nat Cancer Inst* 2002;**94**:116–28.

114. Foekens JA, Peters HA, Look MP, Portengen H, Schmitt M, Kramer MD, *et al.* The urokinase system of plasminogen activation and prognosis in 2780 breast cancer patients. *Cancer Res* 2000; **60**:636–43.
115. Funke I, Schraut W. Meta-analysis of studies on bone marrow micrometastases: An independent prognostic impact remains to be substantiated. *J Clin Oncol* 1998; **16**:557–66.
116. Diel IJ, Cote RJ. Bone marrow and lymph node assessment for minimal residual disease in patients with breast cancer. *Cancer Treat Rev* 2000; **26**:53–65.
117. Singletary SE, Larry L, Tucker S. Detection of micrometastatic tumour cells in bone marrow of breast carcinoma patients. *J Surg Oncol* 1991; **47**:32–6.
118. Diel IJ, Kaufmann M, Costa SD. Micrometastatic breast cancer cells in bone marrow at primary surgery: prognostic value in comparison with nodal status. *J Clin Oncol* 1996; **88**:1652–64.
119. Merkle E, Baehr J, Henke A, Buehner M, Martus P. [Immunocytochemical detection of tumor cells in bone marrow as a prognostic factor in breast carcinoma]. *Geburtshilfe und Frauenheilkunde* 1994; **54**:662–9.
120. Chen Y-Y, Schnitt SJ. Prognostic factors for patients with breast cancer 1 cm and smaller. *Breast Cancer Res Treat* 1998; **51**:209–25.
121. Carter CL, Allen C, Henson DE. Relation of tumour size, lymph node status and survival in 24,740 breast cancer cases. *Cancer* 1989; **63**:181–7.
122. Goodwin PJ, Boyd NF. Body size and breast cancer prognosis: a critical review of the evidence. *Breast Cancer Res Treat* 1990; **16**:205–14.
123. Stewart LA, Clarke MJ, for the Cochrane Collaboration Working Group on Meta-analysis Using Individual Patient Data. Practical methodology of meta-analysis (overviews) using updated individual patient data. *Stat Med* 1995; **14**:2057–79.
124. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med* 2000; **19**:453–73.
125. Clark GM. Do we really need prognostic factors for breast cancer? *Breast Cancer Res Treat* 1994; **30**:117–26.
126. Cooke T, Reeves J, Lannigan A, Stanton P. The value of the human epidermal growth factor receptor-2 (HER2) as a prognostic marker. *Eur J Cancer* 2001; **37**:3–10.
127. Chan CM, Baratta FS, Ozzello L, Ceriani RL. Monoclonal antibody BrE-3 participation in a multivariate prognostic model for infiltrating ductal carcinoma of the breast. *Breast Cancer Res Treat* 1994; **30**:243–61.
128. Burstein AH. Fracture classification systems: do they work and are they useful? *T J Bone Joint Surg Am Vol* 1993; **75**:1743–4.
129. Wyatt JC, Altman DG. Commentary: prognostic models: clinically useful or quickly forgotten? *BMJ* 1995; **311**:1539–41.
130. Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD. Validity of prognostic models: when is a model clinically useful? *Semin Urol Oncol* 2002; **20**:96–107.
131. Tarassenko L, Whitehouse R, Gasparini G, Harris AL. Neural network prediction of relapse in breast cancer. *Neural Comput Appl* 1996; **4**:105–13.
132. Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariate models. *Ann Intern Med* 1993; **118**:210.
133. Schumacher M, Holländer N, Schwarzer G, Sauerbrei W. Prognostic factor studies. In Crowley J, editor. *Handbook of statistics in clinical oncology*. New York: Marcel Dekker; 2001. pp. 321–78.
134. Concato J. Challenges in prognostic analysis. *Cancer (Suppl)* 2001; **91**:1607–14.
135. Burke HB, Henson DE. The American Joint Committee on Cancer. Criteria for prognostic factors and for an enhanced prognostic system. *Cancer* 1993; **72**:3131–5.
136. Redelmeier DA, Lustig AJ. Prognostic indices in clinical practice. *JAMA* 2001; **285**:3024–5.
137. McShane LM, Simon R. Statistical methods for the analysis of prognostic factor studies. In Gospodarowicz MK, Henson DE, Hutter RVP, *et al.* editors. *Prognostic factors in cancer*. New York: Wiley-Liss; 2001. pp. 37–48.
138. Harrell FE Jr, Lee KL, Matchar DB, Reichert TA. Regression models for prognostic prediction: advantages, problems, and suggested solutions. *Cancer Treat Rep* 1985; **69**:1071–7.
139. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. *J Clin Epidemiol* 1995; **48**:1503–10.
140. Feinstein AR. *Multivariable analysis: an introduction*. New Haven, CT: Yale University Press; 1996.
141. Clark TG, Altman DG. Developing a prognostic model in the presence of missing data: an ovarian cancer case study. *J Clin Epidemiol* 2003; **56**: 28–37.
142. Hoppin JA, Tolbert PE, Taylor JA, Schroeder JC, Holly EA. Potential for selection bias with tumor tissue retrieval in molecular epidemiology studies. *Ann Epidemiol* 2002; **12**:1–6.

143. Vach W. Some issues in estimating the effect of prognostic factors from incomplete covariate data. *Stat Med* 1997;**16**:57–72.
144. Schafer JL. Multiple imputation: a primer. *Stat Methods Med Res* 1999;**8**:3–15.
145. Burton A, Altman D. Missing covariate data within cancer prognostic studies: a review of current reporting and proposed guidelines. *Br J Cancer* 2004;**91**:4–8.
146. Altman DG. Suboptimal analysis using ‘optimal’ cutpoints. *Br J Cancer* 1998;**78**:556–7.
147. Bossard N, Descotes F, Bremond AG, Bobin Y, De Saint Hilaire P, Golfier F, *et al.* Keeping data continuous when analyzing the prognostic impact of a tumor marker: an example with cathepsin D in breast cancer. *Breast Cancer Res Treat* 2003;**82**:47–59.
148. Bradburn MJ, Clark TG, Love SB, Altman DG. Survival analysis part II: multivariate data analysis – an introduction to concepts and methods. *Br J Cancer* 2003;**89**:431–6.
149. Sauerbrei W, Schumacher M. A bootstrap resampling procedure for model building: application to the Cox regression model. *Stat Med* 1992;**11**:2093–109.
150. Royston P, Parmar MK. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Stat Med* 2002;**21**:2175–97.
151. Nardi A, Schemper M. Comparing Cox and parametric models in clinical studies. *Stat Med* 2003;**22**:3597–610.
152. LeBlanc M. Tree-based methods for prognostic stratification. In Crowley J, editor. *Handbook of statistics in clinical oncology*. New York: Marcel Dekker; 2001. pp. 457–72.
153. Lemon SC, Roy J, Clark MA, Friedmann PD, Rakowski W. Classification and regression tree analysis in public health: methodological review and comparison with logistic regression. *Ann Behav Med* 2003;**26**:172–81.
154. Erlichman C, Warde P, Gadalla T, Ciampi A, Baskerville T. RECPAM analysis of prognostic factors in patients with stage III breast cancer. *Breast Cancer Res Treat* 1990;**16**:231–42.
155. Ripley RM, Harris AL, Tarassenko L. Neural network models for breast cancer prognosis. *Neural Comput Appl* 1998;**7**:367–75.
156. Rodvold DM, McLeod DG, Brandt JM, Snow PB, Murphy GP. Artificial neural networks for physicians: a technology introduction. In Gospodarowicz MK, Henson DE, Hutter RVP, *et al.*, editors. *Prognostic factors in cancer*. New York: Wiley-Liss; 2001. pp. 57–70.
157. Biganzoli E, Boracchi P, Daidone MG, Gion M, Marubini E. Flexible modelling in survival analysis. Structuring biological complexity from the information provided by tumor markers. *Int J Biol Markers* 1998;**13**:107–23.
158. Sargent DJ. Comparison of artificial neural networks with other statistical approaches: results from medical data sets. *Cancer* 2001;**91**:1636–42.
159. Schwarzer G, Vach W, Schumacher M. On the misuses of artificial neural networks for prognostic and diagnostic classification in oncology. *Stat Med* 2000;**19**:541–61.
160. Schwarzer G, Schumacher M. Artificial neural networks for diagnosis and prognosis in prostate cancer. *Semin Urol Oncol* 2002;**20**:89–95.
161. Elston EW, Ellis IO. Method for grading breast cancer. *J Clin Pathol* 1993;**46**:189–90.
162. Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. *Stat Med* 2004;**23**:723–48.
163. American Society of Clinical Oncology. Clinical practice guidelines for the use of tumour markers in breast and colorectal cancer. *J Clin Oncol* 1996;**14**:2843–77.
164. Glasziou P, Vandenbroucke JP, Chalmers I. Assessing the quality of research. *BMJ* 2004;**328**:39–41.
165. Jüni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. *BMJ* 2001;**323**:42–6.
166. Fisher ER, Anderson S, Redmond C, Fisher B. Pathologic findings from the National Surgical Adjuvant Breast Project Protocol B-06. 10-year pathologic and clinical prognostic discriminants. *Cancer* 1993;**71**:2507–14.
167. Fisher ER, Anderson S, Tan-Chiu E, Fisher B, Eaton L, Wolmark N. Fifteen-year prognostic discriminants for invasive breast carcinoma: National Surgical Adjuvant Breast and Bowel Project Protocol-06. *Cancer* 2001;**91**:1679–87.
168. Chapman JA, Trudeau ME, Pritchard KI, Sawka CA, Mobbs BG, Hanna WM, *et al.* A comparison of all-subset Cox and accelerated failure time models with Cox step-wise regression for node-positive breast cancer. *Breast Cancer Res Treat* 1992;**22**:263–72.
169. Blamey RW. The design and clinical use of the Nottingham prognostic index in breast cancer. *Breast* 1996;**5**:156–7.
170. Burke HB, Rosen DB, Goodman PH. Comparing the prediction accuracy of artificial neural networks and other statistical models for breast cancer survival. In Tesauro G, *et al.*, editors. *Advances in neural information processing systems 7*. Cambridge, MA: MIT Press; 1995. pp. 1063–7.

171. Clark G, Hilsenbeck SG, Ravdin PM, Osborne CK. Validation of a model that identifies high risk breast cancer patients for clinical trials of intensive therapy (meeting abstract). *Proc Annu Meet Am Soc Clin Oncol* 1995;**14**:92.
172. De Laurentiis M, Ravdin PM, Clark GM. Prognostic models for producing numerical estimates of disease free survival of breast cancer patients (meeting abstract). *Breast Cancer Res Treat* 1993;**27**:181.
173. Denic S. Three-dimensional survival model of breast cancer patients. *Anticancer Res* 1996;**16**:3217–19.
174. Frkovic-Grazio S, Bracko M. Long term prognostic value of Nottingham histological grade and its components in early (pT1N0M0) breast carcinoma. *J Clin Pathol* 2002;**55**:88–92.
175. Grumett SA, Snow PB. Artificial neural networks: a new model for assessing prognostic factors. *Ann Oncol* 2000;**11**:383–4.
176. Hasebe T, Mukai K, Tsuda H, Ochiai A. New prognostic histological parameter of invasive ductal carcinoma of the breast: clinicopathological significance of fibrotic focus. *Pathol Int* 2000;**50**:263–72.
177. Knox W, Walls J, Donne A, Wallis N, Redford J, Swindell R, *et al.* The prognostic value of full axillary node clearance (meeting abstract). *J Pathol* 1993;**170**:384A.
178. Marchevsky AM, Shah S, Patel S. Reasoning with uncertainty in pathology: artificial neural networks and logistic regression as tools for prediction of lymph node status in breast cancer patients. *Mod Pathol* 1999;**12**:505–13.
179. Pinder SE, Murray S, Ellis IO, Trihia H, Elston CW, Gelber RD, *et al.* The importance of the histologic grade of invasive breast carcinoma and response to chemotherapy. *Cancer* 1998;**83**:1529–39.
180. Quantin C, Moreau T, Asselain B, Maccario J, Lellouch J. A regression survival model for testing the proportional hazards hypothesis. *Biometrics* 1996;**52**:874–85.
181. Rostgaard K, Mouridsen HT, Vaeth M, Holst H, Olesen KP, Lynge E. A modified Nottingham prognostic index for breast cancer patients diagnosed in Denmark 1978–1994. *Acta Oncol* 2001;**40**:838–43.
182. Rudolph P, Alm P, Heidebrecht HJ, Bolte H, Ratjen V, Baldetorp B, *et al.* Immunologic proliferation marker Ki-S2 as prognostic indicator for lymph node-negative breast cancer. *J Nat Cancer Inst* 1999;**91**:2713–18.
183. Yamamoto N, Watanabe T, Katsumata N, Omuro Y, Ando M, Fukuda H, *et al.* Construction and validation of a practical prognostic index for patients with metastatic breast cancer. *J Clin Oncol* 1998;**16**:2401–8.
184. Guerra I, Algorta J, Diaz dO, Pelayo A, Farina J. Immunohistochemical prognostic index for breast cancer in young women. *Mol Pathol* 2003;**56**:323–7.
185. Lundin M, Lundin J, Burke HB, Toikkanen S, Pylkkanen L, Joensuu H. Artificial neural networks applied to survival prediction in breast cancer A. *Oncology* 1999;**57**:281–6.
186. Kaufmann M, Minckwitz VG, Finn HP, Schmid H, Goerttler K, Bastert G. Combination of grading and new biological factors (S-phase fraction and epidermal growth factor receptor) can predict relapse and survival in patients with node-negative primary breast cancer. *Onkologie* 1994;**17**:166–72.
187. Collan Y, Kumpusalo L, Pesonen E, Eskelinen M, Pajarinen P, Kettunen K. Prediction of survival in breast cancer: evaluation of different multivariate models. *Anticancer Res* 1998;**18**:647–50.
188. Burke HB, Goodman PH, Rosen DB, Henson DE, Weinstein JN, Harrell FE Jr, *et al.* Artificial neural networks improve the accuracy of cancer survival prediction. *Cancer* 1997;**79**:857–62.
189. Lockwood CA, Ricciardelli C, Raymond WA, Seshadri R, McCaul K, Horsfall DJ. A simple index using video image analysis to predict disease outcome in primary breast cancer. *Int J Cancer* 1999;**84**:203–8.
190. Baak JPA, Van Dop H, Kurver PHJ, Hermans J. The value of morphometry to classic prognosticators in breast cancer. *Cancer* 1985;**56**:374–82.
191. Knorr KL, Hilsenbeck SG, Wenger CR, Pounds G, Oldaker T, Vendely P, *et al.* Making the most of your prognostic factors: presenting a more accurate survival model for breast cancer patients. *Breast Cancer Res Treat* 1992;**22**:251–62.
192. Henson DE, Ries L, Freedman LS, Carriaga M. Relationship among outcome, stage of disease, and histologic grade for 22,616 cases of breast cancer. The basis for a prognostic index. *Cancer* 1991;**68**:2142–9.
193. Haybittle JL, Blamey RW, Elston CW, Johnson J, Doyle PJ, Campbell FC. A prognostic index in primary breast cancer. *Br J Cancer* 1982;**45**:361–6.
194. Blamey RW, Davies CJ, Elston CW, Johnson J, Haybittle JL, Maynard PV. Prognostic factors in breast cancer – the formation of a prognostic index. *Clin Oncol* 1979;**5**:227–36.
195. Bloom HJG, Richardson WW. Histological grading and prognosis in breast cancer. *Br J Cancer* 1957;**11**:359–77.
196. Todd JH, Dowle C, Williams MR, Elston CW, Ellis IO, Hinton CP, *et al.* Confirmation of a prognostic index in primary breast cancer. *Br J Cancer* 1987;**56**:489–92.

197. Galea MH, Blamey RW, Elston CW, Ellis IO. The Nottingham Prognostic Index in primary breast cancer. *Breast Cancer Res Treat* 1992;**22**:207–19.
198. De Laurentiis M, De Placido S, Bianco AR, Clark GM, Ravdin PM. A prognostic model that makes quantitative estimates of probability of relapse for breast cancer patients. *Clin Cancer Res* 1999;**5**:4133–9.
199. Albain KS, Green S, LeBlanc M, Rivkin S, O'Sullivan J, Osborne CK. Proportional hazards and recursive partitioning and amalgamation analyses of the Southwest Oncology Group node-positive adjuvant CMFVP breast cancer data base: a pilot study. *Breast Cancer Res Treat* 1992;**22**:273–84.
200. Anderson TJ, Alexander FE, Lamb J, Smith A, Forrest AP. Pathology characteristics that optimize outcome prediction of a breast screening trial. *Br J Cancer* 2000;**83**:487–92.
201. Aubele M, Auer G, Voss A, Falkmer U, Rutquist L, Hoefler H. Disease-free survival of node-positive breast cancer patients. Improved prognostication by cytometrical parameters. *Pathol Res Pract* 1995;**191**:982–90.
202. Bryan RM, Mercer RJ, Bennett RC, Rennie GC. Prognostic factors in breast cancer and the development of a prognostic index. *Br J Surg* 1986;**73**:267–71.
203. Caleffi M, Fentiman IS, Birkhead BG. Factors at presentation influencing the prognosis in breast cancer. *Eur J Cancer Clin Oncol* 1989;**25**:51–6.
204. Ceriani RL, Chan CM, Baratta FS, Ozzello L, DeRosa CM, Habif DV. Levels of expression of breast epithelial mucin detected by monoclonal antibody BrE-3 in breast-cancer prognosis. *Int J Cancer* 1992;**51**:343–54.
205. Chan CM, Baratta FS, Ozzello L, Ceriani RL. Monoclonal antibody BrE-3 participation in a multivariate prognostic model for infiltrating ductal carcinoma of the breast. *Breast Cancer Res Treat* 1994;**30**:243–61.
206. Chapman JW, Hanna W, Kahn HJ, Lickley HL, Wall J, Fish EB, *et al.* Alternative multivariate modelling for time to local recurrence for breast cancer patients receiving a lumpectomy alone. *Surg Oncol* 1996;**5**:265–71.
207. Chevallier B, Mosseri V, Dauce JP, Bastit P, Julien JP, Asselain B. A prognostic score in histological node negative breast cancer. *Br J Cancer* 1990;**61**:436–40.
208. Clark GM, Wenger CR, Beardslee S, Owens MA, Pounds G, Oldaker T, *et al.* How to integrate steroid hormone receptor, flow cytometric, and other prognostic information in regard to primary breast cancer. *Cancer* 1993;**71**:2157–62.
209. Coradini D, Boracchi P, Daidone MG, Pellizzaro C, Miodini P, Ammatuna M, *et al.* Contribution of vascular endothelial growth factor to the Nottingham prognostic index in node-negative breast cancer. *Br J Cancer* 2001;**85**:795–7.
210. D'Eredita GA, Giardina C, Martellotta M, Natale T, Ferrarese F. Prognostic factors in breast cancer: the predictive value of the Nottingham Prognostic Index in patients with a long-term follow-up that were treated in a single institution. *Eur J Cancer* 2001;**37**:591–6.
211. Eskelinen M, Collan Y, Pajarinen P, Pesonen E, Kettunen K, Nordling S. An improved prognostic index of axillary node involvement in breast cancer incorporating DNA ploidy and tumour size. *Acta Chir Scand* 1990;**156**:521–7.
212. Fisher ER, Costantino J, Fisher B, Redmond C. Pathologic findings from the National Surgical Adjuvant Breast Project (Protocol 4). Discriminants for 15-year survival. National Surgical Adjuvant Breast and Bowel Project Investigators. *Cancer* 1993;**71**:2141–50.
213. Fuster E, Garcia-Vilanova A, Narbona B, Romero R, Llombart-Bosch A. A statistical approach to an individualized prognostic index (IPI) for breast cancer survivability. *Cancer* 1983;**52**:728–36.
214. Gago FE, Tello OM, Diblasi AM, Ciocca DR. Integration of estrogen and progesterone receptors with pathological and molecular prognostic factors in breast cancer patients. *J Steroid Biochem Mol Biol* 1998;**67**:431–7.
215. Hansen S, Grabau DA, Sorensen FB, Bak M, Vach W, Rose C. Vascular grading of angiogenesis: prognostic significance in breast cancer. *Br J Cancer* 2000;**82**:339–47.
216. Hawkins RA, Tesdale AL, Killen ME, Jack WJ, Chetty U, Dixon JM, *et al.* Prospective evaluation of prognostic factors in operable breast cancer. *Br J Cancer* 1996;**74**:1469–78.
217. Hawkins RA, Tesdale AL, Prescott RJ, Forster T, McIntyre MA, Baker P *et al.* Outcome after extended follow-up in a prospective study of operable breast cancer: key factors and a prognostic index. *Br J Cancer* 2002;**87**:8–14.
218. Haybittle JL. Prognostic indices in breast cancer. *Colloq Semin* 1986;**137**:779–87.
219. Huang XF, Gao F, Xue ZQ, Wang Y, Xu GW. [Multifactorial analysis of survival in breast cancer]. *Chung-hua Chung Liu Tsa Chih* 1988;**10**:117–19.
220. Kollias J, Murphy CA, Elston CW, Ellis IO, Robertson JFR, Blamey RW. The prognosis of small primary breast cancers. *Eur J Cancer* 1999;**35**:908–12.

221. Korzeniowski S, Dyba T, Skolyszewski J. Classical prognostic factors for survival and loco-regional control in breast cancer patients treated with radical mastectomy alone. *Acta Oncol* 1994; **33**:759–65.
222. McCready DR, Chapman JA, Hanna WM, Kahn HJ, Murray D, Fish EB, *et al.* Factors affecting distant disease-free survival for primary invasive breast cancer: use of a log-normal survival model. *Ann of Surg Oncol* 2000; **7**:416–26.
223. Menard S, Bufalino R, Rilke F, Cascinelli N, Veronesi U, Colnaghi MI. Prognosis based on primary breast carcinoma instead of pathological nodal status. *Br J Cancer* 1994; **70**:709–12.
224. Mitra I, Redkar AA, Badwe RA. Prognosis of breast cancer: evidence for interaction between c-erbB-2 overexpression and number of involved axillary lymph nodes. *J Surg Oncol* 1995; **60**:106–11.
225. Parham DM, Hagen N, Brown RA. Simplified method of grading primary carcinomas of the breast. *J Clin Pathol* 1992; **45**:517–20.
226. Perrone F, Carlomagno C, Lauria R, De Laurentiis M, Morabito A, Panico L, *et al.* Selecting high-risk early breast cancer patients: what to add to the number of metastatic nodes? *Eur J Cancer* 1996; **32A**:41–6.
227. Ravdin PM, Clark GM. A practical application of neural network analysis for predicting outcome of individual breast cancer patients. *Breast Cancer Res Treat* 1992; **22**:285–93.
228. Sauerbrei W, Hubner K, Schmoor C, Schumacher M. Validation of existing and development of new prognostic classification schemes in node negative breast cancer. German Breast Cancer Study Group. *Breast Cancer Res Treat* 1997; **42**:149–63.
229. Scholl S, Bieche I, Pallud C, Champeme MH, Beuvon F, Hacene K, *et al.* Relevance of multiple biological parameters in breast cancer prognosis. *Breast* 1996; **5**:21–30.
230. Seshadri R, Horsfall DJ, McCaul K, Leong AS. A simple index to predict prognosis independent of axillary node information in breast cancer. *Aust N Z J Surg* 1997; **67**:765–70.
231. Shek LLM, Godolphin W. Model for breast cancer survival: relative prognostic roles of axillary nodal status, TNM stage, estrogen receptor concentration, and tumor necrosis. *Cancer Res* 1988; **48**:5565–9.
232. Tubiana Hulin M, Le Doussal V, Hacene K, Rouesse J, Brunet M. Sequential identification of factors predicting distant relapse in breast cancer patients treated by conservative surgery. *Cancer* 1993; **72**:1261–71.
233. van der Linden JC, Lindeman J, Baak JP, Meijer CJ, Herman CJ. The Multivariate Prognostic Index and nuclear DNA content are independent prognostic factors in primary breast cancer patients. *Cytometry* 1989; **10**:56–61.
234. Younes M, Lane M, Miller CC, Laucirica R. Stratified multivariate analysis of prognostic markers in breast cancer: a preliminary report. *Anticancer Res* 1997; **17**:1383–90.
235. Harbeck N, Kates R, Ulm K, Graeff H, Schmitt M. Neural network analysis of follow-up data in primary breast cancer. *Int J Biol Markers* 2000; **15**:116–22.
236. Balslev I, Axelsson CK, Zedeler K, Rasmussen BB, Carstensen B, Mouridsen HT. The Nottingham Prognostic Index applied to 9,149 patients from the studies of the Danish Breast Cancer Cooperative Group (DBCG). *Breast Cancer Res Treat* 1994; **32**:281–90.
237. Collett K, Skjaerven R, Maehle BO. The prognostic contribution of estrogen and progesterone receptor status to a modified version of the Nottingham Prognostic Index. *Breast Cancer Res Treat* 1998; **48**:1–9.
238. Sundquist M, Thorstenson S, Brudin L, Nordenskjold B. Applying the Nottingham Prognostic Index to a Swedish breast cancer population. South East Swedish Breast Cancer Study Group. *Breast Cancer Res Treat* 1999; **53**:1–8.
239. Alexander AI, Mercer RJ, Muir IM, Bennett RC, Rennie GC. Validation of a prognostic index in breast cancer. *Aust N Z J Surg* 1987; **57**:399–403.
240. Stoll BA. Prognostic indices in breast cancer. In Stoll BA, editor. *Pointers to cancer prognosis*. Dordrecht: Martinus Nijhoff; 1992. pp. 195–229.
241. Vollmer RT. Malignant melanoma. A multivariate analysis of prognostic factors. *Pathol Annu* 1989; **24** (Pt 1):383–407.
242. Vollmer RT, Seigler HF. Using a continuous transformation of the Breslow thickness for prognosis in cutaneous melanoma. *Am J Clin. Pathol* 2001; **115**:205–12.
243. Ross PL, Scardino PT, Kattan MW. A catalog of prostate cancer nomograms. *J Urol* 2001; **165**:1562–8.
244. Vollmer RT, Keetch DW, Humphrey PA. Predicting the pathology results of radical prostatectomy from preoperative information: a validation study. *Cancer* 1998; **83**:1567–80.
245. Gobbi PG, Zinzani PL, Brogna C, Comelli M, Magagnoli M, Federico M, *et al.* Comparison of prognostic models in patients with advanced Hodgkin disease. Promising results from integration of the best three systems. *Cancer* 2001; **91**:1467–78.
246. Counsell C, Dennis M. Systematic reviews of prognostic models in patients with acute stroke. *Cerebrovasc Dis* 2001; **12**:159–70.

247. Riley RD, Burchill SA, Abrams KR, Heney D, Sutton AJ, Jones DR, *et al.* A systematic review of molecular and biological markers in tumours of the Ewing's sarcoma family. *Eur J Cancer* 2003; **39**:19–30.
248. Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. *Stat Med* 1999; **18**:2529–45.
249. Slamon DJ, Leyland-Jones B, Shak B, Fuchs H, Paton V, Bajamonde A. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* 2001; **344**:783–92.
250. Tetu B, Brisson J, Cote C, Brisson S, Potvin D, Roberge N. Prognostic significance of cathepsin-D expression in node-positive breast carcinoma: an immunohistochemical study. *Int J Cancer* 1993; **55**:429–35.
251. Billgren A-M, Rutqvist LE, Johansson H, Hagerstrom T, Skoog L. The role of cathepsin D and PAI-1 in primary invasive breast cancer as prognosticators and predictors of treatment benefit with adjuvant tamoxifen. *Eur J Cancer* 2000; **36**:1374–80.
252. Tetu B, Brisson J, Lapointe H, Wang CS, Bernard P, Blanchette C. Cathepsin D expression by cancer and stromal cells in breast cancer: an immunohistochemical study of 1348 cases. *Breast Cancer Res Treat* 1999; **55**:137–47.
253. Tetu B, Brisson J, Wang CS, Lapointe H, Beaudry G, Blanchette C. Expression of cathepsin D, stromelysin-2, and urokinase by reactive stromal cells on breast carcinoma prognosis. *Cancer* 2001; **92**:2957–64.
254. Lohrisch C, Piccart M. HER2/neu as a predictive factor in breast cancer. *Clin Breast Cancer* 2001; **2**:129–35.
255. Piccart MJ, Di Leo A, Hamilton A. HER2. a 'predictive factor' ready to use in the daily management of breast cancer patients? *Eur J Cancer* 2000; **36**:1755–61.
256. Hayes DF, Thor AD. c-erbB-2 in breast cancer: development of a clinically useful marker. *Semin Oncol* 2002; **29**:231–45.
257. Yamauchi H, Stearns V, Hayes DF. When is a tumor marker ready for prime time? A case study of c-erbB-2 as a predictive factor in breast cancer. *J Clin Oncol* 2001; **19**:2334–56.
258. Elledge RM, Allred DC. Prognostic and predictive value of p53 and p21 in breast cancer. *Breast Cancer Res Treat* 1998; **52**:79–98.
259. Hamilton A, Piccart M. The contribution of molecular markers to the prediction of response in the treatment of breast cancer: a review of the literature on HER-2, p53 and BCL-2. *Ann Oncol* 2000; **11**:647–63.
260. Calvert NW, Morgan AB, Catto JWF, Hamdy FC, Akehurst RL, Mouncey P, *et al.* Effectiveness and cost-effectiveness of prognostic markers in prostate cancer. *Br J Cancer* 2003; **88**:31–5.
261. Glance LG, Osier T, Shinozaki T. Intensive care unit prognostic scoring systems to predict death: a cost-effectiveness analysis. *Crit Care Med* 1998; **26**:1842–9.
262. Hamel MB, Phillips RS, Davis RB, Desbiens N, Connors AFJ, Teno JM, *et al.* Outcomes and cost-effectiveness of initiating dialysis and continuing aggressive care in seriously ill hospitalized adults. *Ann Intern Med* 1997; **127**:195–202.
263. Littrup PJ, Kane RA, Mettlin CJ, Murphy GP, Lee F, Toi A, *et al.* Cost-effective prostate cancer detection: reduction of low-yield biopsies. *Cancer* 1994; **74**:3146–58.
264. Mason JM. The DiSC assay: a cost-effective guide to treatment for chronic lymphocytic leukemia? *Int J Technol Assess Health Care* 1999; **15**:173–84.
265. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the *BMJ*. *BMJ* 1996; **313**:275–83.
266. Mikhail MG, Smith FA, Gray M, Britton C, Frederiksen SM. Cost-effectiveness of mandatory stress testing in chest pain center patients. *Ann Emerg Med* 1997; **29**:88–98.
267. Orr JWJ, Orr P, Kern DH. Cost-effective treatment of women with advanced ovarian cancer by cytoreductive surgery and chemotherapy directed by an *in vitro* assay for drug resistance. *Cancer J Sci Am* 1999; **5**:174–88.
268. Tonnaire G. Cytogenetic and molecular biology for acute leukemias at diagnosis: a cost/effectiveness comparison. *Leuk Lymphoma* 1998; **28**:363–70.
269. Kent DM, Langa KM, Selker HP. The potential use of ECG-based prognostic instruments in clinical trials and cost-effectiveness analyses of new therapies in acute cardiac ischemia. *J Electrocardiol* 2000; **33**:263–8.
270. Bast RCJ, Desch CE, Ravdin P, Smith TJ. Clinical practice guidelines for the use of tumour markers in breast and colorectal cancer. *J Clin Oncol* 1996; **14**:2843–77.
271. Kievit J, van de Velde CJH. Utility and cost of carcinoembryonic antigen monitoring in colon cancer follow-up evaluation: a Markov analysis. *Cancer* 1990; **65**:2580–7.
272. Lundin J, Lundin M, Isola J, Joensuu H. A web-based system for individualised survival estimation in breast cancer. *BMJ* 2003; **326**:29.
273. Cox DR. Note on grouping. *J Am Stat Assoc* 1957; **52**:543–7.

274. Government Actuary Department. *Interim life tables for England and Wales, women based on the years 1999–2001*. London: Government Actuary Department; 2002.
275. Karnon J, Brown J. Tamoxifen plus chemotherapy versus tamoxifen alone as adjuvant therapies for node-positive postmenopausal women with early breast cancer: a stochastic economic evaluation. *Pharmacoeconomics* 2002;**20**:119–37.
276. Cocquyt V, Moeremans K, Clarys P, Van Belle S, Annemans L. Postmenopausal breast cancer: incidence-based cost of illness. *J Med Econ* 2003;**6**:15–30.
277. Rawlins MD, Culyer AJ. National Institute for Clinical Excellence and its value judgments. *BMJ* 2004;**329**:224–7.
278. Cochrane manual. URL: <http://www.cochrane.org/admin/manual.htm>. Accessed 2004.
279. Haagensen CD, Cooley E. Radical mastectomy for mammary carcinoma. *Ann Surg* 1969;**170**:884–8.
280. Henderson IC, Canellos GP. Cancer of the breast: the past decade (first of two parts). *N Engl J Med* 1980;**302**:17–30.
281. Halsted WS. The results of operations for the cure of cancer of the breast at the Johns Hopkins Hospital from June, 1889 to January, 1894. *Johns Hopkins Hosp Bull* 1907;**4**:1984–5.
282. Early Breast Cancer Trialists' Collaborative Group. Polychemotherapy for early breast cancer: and overview of the randomised trials. *Lancet* 1998;**352**:930–42.
283. Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer: an overview of the randomised trials. *Lancet* 1998;**351**:1451–67.
284. van der Linden JC, Baak JP, Lindeman J, Hermans J, Meyer CJ. Prospective evaluation of prognostic value of morphometry in patients with primary breast cancer. *J Clin Pathol* 1987;**40**:302–6.
285. Richards M, Statham H, Green J, Murton F. Management of breast cancer. Understanding women's preconceptions. *BMJ* 1994;**308**:716.
286. Richards MA, Smith IE, Dixon JM. ABC of breast diseases. Role of systemic treatment for primary operable breast cancer. *BMJ* 1994;**309**:1363–6.

Appendix I

Search strategies

Six searches were carried out to find research relevant to the six areas of the review:

Topic A: Prognostic models in breast cancer

Topic B: Predictive factors in breast cancer

Topic C: Reviews of prognostic variables/factors in breast cancer

Topic D: Quality assessment of prognostic studies (not restricted to cancer)

Topic E: Clinical use of prognostic information in breast and other cancers

Topic F: Quality of life, cost and cost-effectiveness relevant to the modelling.

The search strategies were developed using an iterative process.

The following databases were searched:

Database	Topic A	Topic B	Topic C	Topic D	Topic E	Topic F
MEDLINE (Ovid)	1980–Aug. 2001	1980–Aug. 2001	1980–Aug. 2001	1980–Aug. 2001	1990–Aug. 2001	1966–Dec. 2000
CCTR	Not searched	Not searched	Not searched	Not searched	Not searched	Issue 2001/1
Biosis	Suggested strategy to be run locally in Oxford	Suggested strategy to be run locally in Oxford	Suggested strategy to be run locally in Oxford	Suggested strategy to be run locally in Oxford	Suggested strategy to be run locally in Oxford	Suggested strategy to be run locally in Oxford
EMBASE (winSPIRS)	Developed strategy to be run locally in Oxford	Developed strategy to be run locally in Oxford	Developed strategy to be run locally in Oxford	Developed strategy to be run locally in Oxford	Developed strategy to be run locally in Oxford	Developed strategy to be run locally in Oxford
CancerLit (OVID)	1980–2001/03	1980–2001/03	1980–2001/03	1980–2001/03	1990–2001/03	1963–2000/9
NHSEED (CRD admin. database)	Not searched	Not searched	Not searched	Not searched	Not searched	1995 to date
HEED	Not searched	Not searched	Not searched	Not searched	Not searched	To April 2001

Full details of all strategies are given below.

Topic A. Prognostic models in breast cancer

The CancerLit search was undertaken on two different versions. The period 1995–2001/03 was searched on the OVID CD-Rom version of CancerLit. The period before 1995 was searched on the version of CancerLit offered by the Dialog database service.

SilverPlatterASCII 3.0WINNCANCERLIT 1998–2001/03

(prognos* near (index or indices or indexes) near (relapse or recurrence or surviv* or death* or mortality)) in ti,ab
 (model* near prognos* near (relapse or recurrence or surviv* or death* or mortality)) in ti,ab
 (predictive model* near (recurrence or relapse or surviv* or death* or mortality)) in ti,ab
 neural network* with (recurrence or relapse or surviv* or death* or mortality) in ti,ab
 explode "Prognosis"/ all subheadings
 "Survival-Rate"/ all subheadings
 "Disease-Free-Survival"/all subheadings
 "Mortality"/all subheadings
 "Recurrence"/all subheadings
 "Neural-Networks-Computer"
 explode "Survival-Analysis"/ all subheadings
 explode "Models-Statistical"/all subheadings
 "Breast-Neoplasms"/ all subheadings
 "Mammary-Neoplasms"/all subheadings
 breast cancer in ti,ab
 "Breast-Neoplasms"/secondary
 Metasta* in ti
 explode "Neoplasms"/secondary
 explode "neoplasm-metastasis"/all subheadings
 exact{LETTER} in pt
 exact¹²⁹ in pt
 (animal or cell line* or vitro or invitro or cell or rat or rats or mouse or mice) in ti
 #1 or #2 or #3 or #4
 #5 and (#6 or #7 or #8 or #9) and (#10 or #11 or #12)
 #23 or #24
 #13 or #14 or #15
 #25 and #26
 #27 not (#16 or #17 or #18 or #19 or #20 or #21 or #22)
 #27 not (#20 or #21 or #22)
 #29 not #28

MEDLINE

OVID MEDLINE on CD-ROM was searched for the period 1980–August 2001.

#prognostic methods

(prognos\$ adj4 (index or indices or indexes) adj4 (relapse or recurrence or surviv\$ or death\$ or mortality)).ti,ab.
 (model\$ adj4 prognos\$ adj4 (relapse or recurrence or survival\$ or death\$ or mortality)).ti,ab.
 (predictive model\$ adj4 (recurrence or relapse or surviv\$ or death\$ or mortality)).ti,ab.
 (neural network\$ adj4 (recurrence or relapse or surviv\$ or death\$ or mortality)).ti,ab.
 exp prognosis/
 survival rate/
 disease free survival/
 mortality/
 recurrence/
 neural networks computer/
 exp survival analysis/
 exp models statistical/
 breast neoplasms/
 mammary neoplasms/
 breast cancer.ti,ab.
 breast neoplasms/sc
 metasta\$.ti.
 exp neoplasms/sc
 exp neoplasm metastasis/
 letter.pt.
 comment.pt.
 (animal or cell line\$ or vitro or invitro or cell or rat or rats or mouse or mice).ti.
 or/1-4
 5 and (6 or 7 or 8 or 9) and (10 or 11 or 12)
 23 or 24
 13 or 14 or 15
 25 and 26
 27 not (16 or 17 or 18 or 19 or 20 or 21 or 22)
 27 not (20 or 21 or 22)
 29 not 28

BIOSIS

A BIOSIS strategy was produced for use with the WinSPIRS interface. These searches were carried out in Oxford.

SilverPlatterASCII 3.0WINNEMBASE (R) 2001/07–2001/10

(prognos* near (index or indices or indexes) near (relapse or recurrence or surviv* or death* or mortality)) in ti,ab
 (model* near prognos* near (relapse or recurrence or surviv* or death* or mortality)) in ti,ab
 (predictive model* near (recurrence or relapse or surviv* or death* or mortality)) in ti,ab
 neural network* with (recurrence or relapse or surviv* or death* or mortality) in ti,ab
 Prognos* in ds,cb,mi
 Survival in ds
 Relapse in ds

Mortality in ds
 Recurrence* in ds
 Recurrent in ds
 Model in ds
 Breast adenocarcinoma in ds or neoplastic disease in ds
 breast cancer in ti,ab,mi,ds
 Metastasis in ti,ds
 Metastases in ti,ds
 Animals in tn
 Humans in tn
 Exact {letter} in dt or exact¹²⁹ in dt
 #1 or #2 or #3 or #4
 #5 and (#6 or #7 or #8 or #9 or #10) and #11
 #19 or #20
 #12 or #13
 #21 and #22
 #16 not (#16 and #17)
 #23 not (#14 or #15 or #24)

EMBASE

The EMBASE strategy was developed for the WinSPIRS interface and was run at Oxford.

SilverPlatterASCII 3.0WINNEMBASE (R) 2001/07-2001/10

(prognos* near (index or indices or indexes) near (relapse or recurrence or surviv* or death* or mortality)) in ti,ab
 (model* near prognos* near (relapse or recurrence or surviv* or death* or mortality)) in ti,ab
 (predictive model* near (recurrence or relapse or surviv* or death* or mortality)) in ti,ab
 neural network* with (recurrence or relapse or surviv* or death* or mortality) in ti,ab
 "Prognosis"/ all subheadings
 explode "Survival"/ all subheadings
 "Relapse"/all subheadings
 "Mortality"/all subheadings or "cancer-mortality"/all subheadings
 "Recurrent-Disease"/all subheadings
 "Artificial-Neural-Network"
 "Statistical-Model"/all subheadings
 explode "Breast-Cancer"/ all subheadings
 breast cancer in ti,ab
 Metasta* in ti
 explode "Metastasis"/all subheadings
 exact {LETTER} in dt or "letter"/all subheadings
 (animal or cell line* or vitro or invitro or cell or rat or rats or mouse or mice) in ti
 #1 or #2 or #3 or #4
 #5 and (#6 or #7 or #8 or #9) and (#10 or #11)
 #18 or #19
 #12 or #13
 #20 and #21
 #22 not (#14 or #15 or #16 or #17)

((oestrogen receptor*) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab
 ((estrogen receptor*) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab
 ((progesterone receptor*) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab
 ((erbb2 or c erbb 2 or cerbb2) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab
 ((her 2 or her2 or neu or p53 or bcl2) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab
 #24 or #25 or #26 or #27 or #28
 explode "Estrogen-Receptor"/ all subheadings
 "Progesterone-Receptor"/ all subheadings
 "Oncogene-neu"/ all subheadings
 "Oncogene-c-erb"/ all subheadings or "protein-p53"/all subheadings
 #30 or #31 or #32 or #33
 "Mortality"/ all subheadings
 "Cancer-mortality"/ all subheadings
 explode "Survival"/ all subheadings
 "Recurrent-Disease"/ all subheadings
 "Cancer-growth"/ all subheadings or "Cancer-recurrence"/ all subheadings or "Cancer-inhibition"/ all subheadings or "Cancer-regression"/ all subheadings
 "Tumor-growth"/ all subheadings or "Tumor-recurrence"/ all subheadings or "Tumor-Regression"/ all subheadings
 #35 or #36 or #37 or #38 or #39 or #40
 #34 and #41
 #21 and (#29 or #42)
 "case-report"/all subheadings
 exact{LETTER} in dt or "letter"/all subheadings
 "Case-study"/all subheadings
 #44 or #45 or #46
 #43 not #47
 #48 not #23

Topic B. Predictive factors in breast cancer

The CancerLit search was undertaken on two different versions. The period 1995-2001/03 was searched on the OVID CD-Rom version of CancerLit. The period before 1995 was searched on the version of CancerLit offered by the Dialog database service.

**SilverPlatterASCII 3.0WINNCANCERLIT
1996-1997**

(prognos* near (index or indices or indexes) near (relapse or recurrence or surviv* or death* or mortality)) in ti,ab
(model* near prognos* near (relapse or recurrence or surviv* or death* or mortality)) in ti,ab
(predictive model* near (recurrence or relapse or surviv* or death* or mortality)) in ti,ab
neural network* with (recurrence or relapse or surviv* or death* or mortality) in ti,ab
explode "Prognosis"/ all subheadings
"Survival-Rate"/ all subheadings
"Disease-Free-Survival"/all subheadings
"Mortality"/all subheadings
"Recurrence"/all subheadings
"Neural-Networks-Computer"
explode "Survival-Analysis"/ all subheadings
explode "Models-Statistical"/all subheadings
"Breast-Neoplasms"/ all subheadings
"Mammary-Neoplasms"/all subheadings
breast cancer in ti,ab
"Breast-Neoplasms"/secondary
Metasta* in ti
explode "Neoplasms"/secondary
explode "neoplasm-metastasis"/all subheadings
exact{LETTER} in pt
exact¹²⁹ in pt
(animal or cell line* or vitro or invitro or cell or rat or rats or mouse or mice) in ti
#1 or #2 or #3 or #4
#5 and (#6 or #7 or #8 or #9) and (#10 or #11 or #12)
#23 or #24
#13 or #14 or #15
#25 and #26
#27 not (#16 or #17 or #18 or #19 or #20 or #21 or #22)
((oestrogen receptor*) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab
((estrogen receptor*) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab
((progesterone receptor*) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab
((erbb2 or c erbb 2 or cerbb2) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab
((her 2 or her2 or neu or p53 or bcl2) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab
#29 or #30 or #31 or #32 or #33
explode "Receptors-Estrogen"/ all subheadings
"Receptors-Progesterone"/ all subheadings

"Genes-erbB-2"/ all subheadings
"Genes-p53"/ all subheadings
#35 or #36 or #37 or #38
"Mortality"/ all subheadings
"Fatal-Outcome"/ all subheadings
"Survival-Rate"/ all subheadings
"Recurrence"/ all subheadings
"Disease-Progression"/ all subheadings
"Disease-Free-Survival"/ all subheadings
#40 or #41 or #42 or #43 or #44 or #45
#39 and #46
#26 and (#34 or #47)
exact{CASES} in PT
exact{LETTER} in PT
exact{NEWS} in PT
#49 or #50 or #51 or #22
#48 not #52
#53 not #28

MEDLINE

OVID MEDLINE on CD-ROM was searched for the period 1980-August 2001.

#Predictive factors
(prognos\$ adj4 (index or indices or indexes) adj4 (relapse or recurrence or surviv\$ or death\$ or mortality)).ti,ab.
(model\$ adj4 prognos\$ adj4 (relapse or recurrence or survival\$ or death\$ or mortality)).ti,ab.
(predictive model\$ adj4 (recurrence or relapse or surviv\$ or death\$ or mortality)).ti,ab.
(neural network\$ adj4 (recurrence or relapse or surviv\$ or death\$ or mortality)).ti,ab.
exp prognosis/
survival rate/
disease free survival/
mortality/
recurrence/
neural networks computer/
exp survival analysis/
exp models statistical/
breast neoplasms/
mammary neoplasms/
breast cancer.ti,ab.
breast neoplasms/sc
metasta\$.ti.
exp neoplasms/sc
exp neoplasm metastasis/
letter.pt.
comment.pt.
(animal or cell line\$ or vitro or invitro or cell or rat or rats or mouse or mice).ti.
or/1-4
5 and (6 or 7 or 8 or 9) and (10 or 11 or 12)
23 or 24
13 or 14 or 15
25 and 26

27 not (16 or 17 or 18 or 19 or 20 or 21 or 22)
 (oestrogen receptor\$ adj4 (death or recurrence or survival or relapse or prognosis or mortality)).ti,ab.
 (estrogen receptor\$ adj4 (death or recurrence or survival or relapse or prognosis or mortality)).ti,ab.
 (progesterone receptor\$ adj4 (death or recurrence or survival or relapse or prognosis or mortality)).ti,ab.
 ((erbb2 or c erbb 2 or cerbb2) adj4 (death or recurrence or survival or relapse or prognosis or mortality)).ti,ab.
 ((her 2 or her2 or neu or p53 or bc12) adj4 (death or recurrence or survival or relapse or prognosis or mortality)).ti,ab.
 or/29-33
 exp receptors-estrogen/
 receptors, progesterone/
 genes-erbb-2/
 genes-p53/
 or/35-38
 mortality/
 fatal outcome/
 survival rate/
 recurrence/
 disease progression/
 disease free survival/
 or/40-45
 39 and 46
 26 and (34 or 47)
 cases.pt.
 letter.pt.
 news.pt.
 49 or 50 or 51 or 22
 48 not 52
 53 not 28

BIOSIS

A BIOSIS strategy was produced for use with the WinSPIRS interface. These searches were carried out in Oxford.

SilverPlatterASCII 3.0WINNEMBASE (R) 2001/07-2001/10

(prognos* near (index or indices or indexes) near (relapse or recurrence or surviv* or death* or mortality)) in ti,ab
 (model* near prognos* near (relapse or recurrence or surviv* or death* or mortality)) in ti,ab
 (predictive model* near (recurrence or relapse or surviv* or death* or mortality)) in ti,ab
 neural network* with (recurrence or relapse or surviv* or death* or mortality) in ti,ab
 Prognos* in ds,cb,mi
 Survival in ds
 Relapse in ds

Mortality in ds
 Recurrence* in ds
 Recurrent in ds
 Model in ds
 Breast adenocarcinoma in ds or neoplastic disease in ds
 breast cancer in ti,ab,mi,ds
 Metastasis in ti,ds
 Metastases in ti,ds
 Animals in tn
 Humans in tn
 Exact {letter} in dt or exact ¹²⁹ in dt
 #1 or #2 or #3 or #4
 #5 and (#6 or #7 or #8 or #9 or #10) and #11
 #19 or #20
 #12 or #13
 #21 and #22
 #16 not (#16 and #17)
 #23 not (#14 or #15 or #24)
 ((oestrogen receptor*) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab
 ((estrogen receptor*) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab
 ((progesterone receptor*) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab
 ((erbb2 or c erbb 2 or cerbb2) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab
 ((her 2 or her2 or neu or p53 or bcl2) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab
 #26 or #27 or #28 or #29 or #30
 Estrogen* in cb
 Progesterone Receptor* in cb
 Erbb2* in cb
 Cerb* in cb
 #32 or #33 or #34 or #35
 Survival in ds
 Relapse in ds
 Mortality in ds
 Recurrence* in ds
 Recurrent in ds
 #37 or #38 or #39 or #40 or #41
 #36 and #42
 #22 and (#31 or #43)
 #44 not (#14 or #15 or #24 or #25)

EMBASE

The EMBASE strategy was developed for the WinSPIRS interface and was run at Oxford.

SilverPlatterASCII 3.0WINNEMBASE (R) 2001/07-2001/10

(prognos* near (index or indices or indexes) near

(relapse or recurrence or surviv* or death* or mortality)) in ti,ab
(model* near prognos* near (relapse or recurrence or surviv* or death* or mortality)) in ti,ab
(predictive model* near (recurrence or relapse or surviv* or death* or mortality)) in ti,ab
neural network* with (recurrence or relapse or surviv* or death* or mortality) in ti,ab
"Prognosis"/ all subheadings
explode "Survival"/ all subheadings
"Relapse"/all subheadings
"Mortality"/all subheadings or "cancer-mortality"/all subheadings
"Recurrent-Disease"/all subheadings
"Artificial-Neural-Network"
"Statistical-Model"/all subheadings
explode "Breast-Cancer"/ all subheadings
breast cancer in ti,ab
Metasta* in ti
explode "Metastasis"/all subheadings
exact {LETTER} in dt or "letter"/all subheadings
(animal or cell line* or vitro or invitro or cell or rat or rats or mouse or mice) in ti
#1 or #2 or #3 or #4
#5 and (#6 or #7 or #8 or #9) and (#10 or #11)
#18 or #19
#12 or #13
#20 and #21
#22 not (#14 or #15 or #16 or #17)
((oestrogen receptor*) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab
((estrogen receptor*) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab
((progesterone receptor*) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab
((erbb2 or c erbb 2 or cerbb2) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab
((her 2 or her2 or neu or p53 or bcl2) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab
#24 or #25 or #26 or #27 or #28
explode "Estrogen-Receptor"/ all subheadings
"Progesterone-Receptor"/ all subheadings
"Oncogene-neu"/ all subheadings
"Oncogene-c-erb"/ all subheadings or "protein-p53"/all subheadings
#30 or #31 or #32 or #33
"Mortality"/ all subheadings
"Cancer-mortality"/ all subheadings
explode "Survival"/ all subheadings
"Recurrent-Disease"/ all subheadings
"Cancer-growth"/ all subheadings or "Cancer-

recurrence"/ all subheadings or "Cancer-inhibition"/ all subheadings or "Cancer-regression"/ all subheadings
"Tumor-growth"/ all subheadings or "Tumor-recurrence"/ all subheadings or "Tumor-Regression"/ all subheadings
#35 or #36 or #37 or #38 or #39 or #40
#34 and #41
#21 and (#29 or #42)
"case-report"/all subheadings
exact{LETTER} in dt or "letter"/all subheadings
"Case-study"/all subheadings
#44 or #45 or #46
#43 not #47
#48 not #23
((prognostic variable*) near (death or recurrence or survival or relapse or prognosis)) in ti,ab
((prognostic factor*) near (death or recurrence or survival or relapse or prognosis)) in ti,ab
"Cancer-Staging"/ all subheadings
Explode "Mitosis"/ all subheadings
"Epidermal-Growth-factor-receptor"/ all subheadings
"Cell-Surface-Receptor"/ all subheadings
"Ploidy"/ all subheadings
"Transforming-Growth-Factor-alpha"/ all subheadings
"Cathepsin-D"/ all subheadings
explode "Estrogen-Receptor"/ all subheadings
"Progesterone-Receptor"/ all subheadings
"Oncogene-neu"/ all subheadings
"Oncogene-c-erb"/ all subheadings or "protein-p53"/ all subheadings
"Prognosis"/ all subheadings
explode "Survival"/ all subheadings
#39 or #40
"Mortality"/all subheadings
"Recurrent-Disease"/all subheadings
(tnm staging near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
(histologic grade near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
(histologic type near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
(mitotic near count* near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
(hormone receptor status near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
((cerbb2 or erbb2)near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
(her2 near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab

(proliferation marker* near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (invasion near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (p53 near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (dna ploidy analysis near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (microvessel density near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (epidermal growth factor receptor* near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (transforming growth factor near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 ((bc12 or bc1 2 or ps2) near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 ((cathepsin d) near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 #50 or #51
 (#52 or #53 or #54 or #55 or #56 or #57 or #58 or #59 or #60 or #61 or #62) and #63 and (#64 or #65 or #66 or #67)
 #68 or #69 or #70 or #71 or #72 or #73 or #74 or #75 or #76 or #77 or #78 or #79 or #80 or #81 or #82 or #83
 #84 or #85 or #86
 "meta-analysis"/ all subheadings
 metaanalys* in ti,ab
 meta-analys* in ti,ab
 meta analys* in ti,ab
 cochrane in ti,ab
 (review* or overview*) in ti
 review in dt
 (synthes* near3 (literature* or research* or studies or data)) in ti,ab
 (pooled analys*) in ti,ab
 (data near2 pool*) and studies
 (medline or medlars or embase or cinahl or scisearch or psychinfo or psycinfo or psychlit or psyclit) in ti,ab
 ((hand or manual or database* or computer*) near2 search*) in ti,ab
 ((electronic or bibliographic*) near2 (database* or data base*)) in ti,ab
 ((review* or overview*) near10 (systematic* or methodologic* or quantitativ* or research* or literature* or studies or trial* or effective*)) in ab
 #88 or #89 or #90 or #91 or #92 or #93 or #94 or #95 or #96 or #97 or #99 or #100 or #101
 #87 and #102

#103 not (#47 or #23 or #49)
 #104 and #21

Topic C. Reviews of prognostic variables/factors in breast cancer

The CancerLit search was undertaken on two different versions. The period 1995–2001/03 was searched on the OVID CD-Rom version of CancerLit. The period before 1995 was searched on the version of CancerLit offered by the Dialog database service.

SilverPlatterASCII 3.0WINNCANCERLIT 1996–1997

(prognos* near (index or indices or indexes) near (relapse or recurrence or surviv* or death* or mortality)) in ti,ab
 (model* near prognos* near (relapse or recurrence or surviv* or death* or mortality)) in ti,ab
 (predictive model* near (recurrence or relapse or surviv* or death* or mortality)) in ti,ab
 neural network* with (recurrence or relapse or surviv* or death* or mortality) in ti,ab
 explode "Prognosis"/ all subheadings
 "Survival-Rate"/ all subheadings
 "Disease-Free-Survival"/all subheadings
 "Mortality"/all subheadings
 "Recurrence"/all subheadings
 "Neural-Networks-Computer"
 explode "Survival-Analysis"/ all subheadings
 explode "Models-Statistical"/all subheadings
 "Breast-Neoplasms"/ all subheadings
 "Mammary-Neoplasms"/all subheadings
 breast cancer in ti,ab
 "Breast-Neoplasms"/secondary
 Metasta* in ti
 explode "Neoplasms"/secondary
 explode "neoplasm-metastasis"/all subheadings
 exact{LETTER} in pt
 exact¹²⁹ in pt
 (animal or cell line* or vitro or invitro or cell or rat or rats or mouse or mice) in ti
 #1 or #2 or #3 or #4
 #5 and (#6 or #7 or #8 or #9) and (#10 or #11 or #12)
 #23 or #24
 #13 or #14 or #15
 #25 and #26
 #27 not (#16 or #17 or #18 or #19 or #20 or #21 or #22)
 ((oestrogen receptor*) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab

((estrogen receptor*) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab
 ((progesterone receptor*) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab
 ((erbb2 or c erbb 2 or cerbb2) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab
 ((her 2 or her2 or neu or p53 or bcl2) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab
 #29 or #30 or #31 or #32 or #33
 explode "Receptors-Estrogen"/ all subheadings
 "Receptors-Progesterone"/ all subheadings
 "Genes-erbB-2"/ all subheadings
 "Genes-p53"/ all subheadings
 #35 or #36 or #37 or #38
 "Mortality"/ all subheadings
 "Fatal-Outcome"/ all subheadings
 "Survival-Rate"/ all subheadings
 "Recurrence"/ all subheadings
 "Disease-Progression"/ all subheadings
 "Disease-Free-Survival"/ all subheadings
 #40 or #41 or #42 or #43 or #44 or #45
 #39 and #46
 #26 and (#34 or #47)
 exact{CASES} in PT
 exact{LETTER} in PT
 exact{NEWS} in PT
 #49 or #50 or #51 or #22
 #48 not #52
 #53 not #28
 ((prognostic variable*) near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 ((prognostic factor*) near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 "Neoplasm-Staging"/ all subheadings
 "Mitosis"/ all subheadings
 "Mitotic-Index"/ all subheadings
 explode "Receptors-Cell-Surface"/ all subheadings
 explode "Ploidies"/ all subheadings
 "Transforming-Growth-Factor-alpha"/ all subheadings
 "Cathepsin-D"/ all subheadings
 explode "Receptors-Estrogen"/ all subheadings
 "Receptors-Progesterone"/ all subheadings
 "Genes-erbB-2"/ all subheadings
 "Genes-p53"/ all subheadings
 explode "Prognosis"/ all subheadings
 "Survival-Rate"/ all subheadings
 "Disease-Free-Survival"/all subheadings
 "Mortality"/all subheadings
 "Recurrence"/all subheadings
 (tnm staging near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab

(histologic grade near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (histologic type near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (mitotic near count* near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (hormone receptor status near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 ((cerbb2 or erbb2)near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (her2 near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (proliferation marker* near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (invasion near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (p53 near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (dna ploidy analysis near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (microvessel density near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (epidermal growth factor receptor* near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (transforming growth factor near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 ((bc12 or bcl 2 or ps2) near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 ((cathepsin d) near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 exact{META-ANALYSIS} in PT
 exact{REVIEW} in PT
 systematic review* in ti,ab
 overview* in ti,ab
 review in ti
 #55 or #56
 (#67 or #58 or #59 or #60 or #61 or #62 or #63 or #64 or #65 or #66 or #67) and #68 and (#69 or #70 or #71 or #72)
 #73 or #74 or #75 or #76 or #77 or #78 or #79 or #80 or #81 or #82 or #83 or #84 or #85 or #86 or #87 or #88
 #94 or #95 or #96
 #89 or #90 or #91 or #92 or #93
 #97 and #98

#99 not (#52 or #28 or #54)
#100 and #26

MEDLINE

OVID MEDLINE on CD-ROM was searched for the period 1980–August 2001.

#reviews of prognosis
(prognos\$ adj4 (index or indices or indexes) adj4 (relapse or recurrence or survival\$ or death\$ or mortality)).ti,ab.
(model\$ adj4 prognos\$ adj4 (relapse or recurrence or survival\$ or death\$ or mortality)).ti,ab.
(predictive model\$ adj4 (recurrence or relapse or survival\$ or death\$ or mortality)).ti,ab.
(neural network\$ adj4 (recurrence or relapse or survival\$ or death\$ or mortality)).ti,ab.
exp prognosis/
survival rate/
disease free survival/
mortality/
recurrence/
neural networks computer/
exp survival analysis/
exp models statistical/
breast neoplasms/
mammary neoplasms/
breast cancer.ti,ab.
breast neoplasms/sc
metasta\$.ti.
exp neoplasms/sc
exp neoplasm metastasis/
letter.pt.
comment.pt.
(animal or cell line\$ or vitro or invitro or cell or rat or rats or mouse or mice).ti.
or/1-4
5 and (6 or 7 or 8 or 9) and (10 or 11 or 12)
23 or 24
13 or 14 or 15
25 and 26
27 not (16 or 17 or 18 or 19 or 20 or 21 or 22)
(oestrogen receptor\$ adj4 (death or recurrence or survival or relapse or prognosis or mortality)).ti,ab.
(estrogen receptor\$ adj4 (death or recurrence or survival or relapse or prognosis or mortality)).ti,ab.
(progesterone receptor\$ adj4 (death or recurrence or survival or relapse or prognosis or mortality)).ti,ab.
((erbb2 or c erbb 2 or cerbb2) adj4 (death or recurrence or survival or relapse or prognosis or mortality)).ti,ab.
((her 2 or her2 or neu or p53 or bc12) adj4 (death or recurrence or survival or relapse or prognosis or mortality)).ti,ab.

or/29-33
exp receptors-estrogen/
receptors, progesterone/
genes-erbb-2/
genes-p53/
or/35-38
mortality/
fatal outcome/
survival rate/
recurrence/
disease progression/
disease free survival/
or/40-45
39 and 46
26 and (34 or 47)
cases.pt.
letter.pt.
news.pt.
49 or 50 or 51 or 22
48 not 52
53 not 28
(prognostic variable\$ adj4 (death or recurrence or survival or relapse or prognosis)).ti,ab.
(prognostic factor\$ adj4 (death or recurrence or survival or relapse or prognosis)).ti,ab.
neoplasm staging/
mitosis/
mitotic index/
exp receptors cell surface/
exp ploidy/
transforming growth factor alpha/
cathepsin d/
exp receptors estrogen/
receptors progesterone/
genes erbb 2/
genes p53/
exp prognosis/
survival rate/
disease free survival/
mortality/
recurrence/
(tnm staging adj4 prognos\$ adj4 (death or recurrence or survival or relapse)).ti,ab.
(histologic grade adj4 prognos\$ adj4 (death or recurrence or survival or relapse)).ti,ab.
(histologic type adj4 prognos\$ adj4 (death or recurrence or survival or relapse)).ti,ab.
(mitotic adj4 count adj4 prognos\$ adj4 (death or recurrence or survival or relapse)).ti,ab.
(hormone receptor status adj4 prognos\$ adj4 (death or recurrence or survival or relapse)).ti,ab.
((cerbb2 or erbb2) adj4 prognos\$ adj4 (death or recurrence or survival or relapse)).ti,ab.
(her2 adj4 prognos\$ adj4 (death or recurrence or survival or relapse)).ti,ab.

(proliferation marker\$ adj4 prognos\$ adj4 (death or recurrence or survival or relapse)).ti,ab.
 (invasion adj4 prognos\$ adj4 (death or recurrence or survival or relapse)).ti,ab.
 (p53 adj4 prognos\$ adj4 (death or recurrence or survival or relapse)).ti,ab.
 (dna ploidy analysis adj4 prognos\$ adj4 (death or recurrence or survival or relapse)).ti,ab.
 (microvessel density adj4 prognos\$ adj4 (death or recurrence or survival or relapse)).ti,ab.
 (epidermal growth factor receptor\$ adj4 prognos\$ adj4 (death or recurrence or survival or relapse)).ti,ab.
 (transforming growth factor adj4 prognos\$ adj4 (death or recurrence or survival or relapse)).ti,ab.
 ((bc12 or bcl 2 or ps2) adj4 prognos\$ adj4 (death or recurrence or survival or relapse)).ti,ab.
 (cathepsin d adj4 prognos\$ adj4 (death or recurrence or survival or relapse)).ti,ab.
 meta-analysis.pt.
 review.pt.
 systematic review\$.ti,ab.
 overview\$.ti,ab.
 review.ti.
 or/55-56
 (57 or 58 or 59 or 60 or 61 or 62 or 63 or 64 or 65 or 66 or 67) and 68 and (69 or 70 or 71 or 72)
 or/73-88
 or/94-96
 or/89-93
 97 and 98
 99 not (52 or 28 or 54)
 100 and 26

BIOSIS

A BIOSIS strategy was produced for use with the WinSPIRS interface. These searches were carried out in Oxford.

SilverPlatterASCII 3.0WINNEMBASE (R) 2001/07-2001/10

(prognos* near (index or indices or indexes) near (relapse or recurrence or surviv* or death* or mortality)) in ti,ab
 (model* near prognos* near (relapse or recurrence or surviv* or death* or mortality)) in ti,ab
 (predictive model* near (recurrence or relapse or surviv* or death* or mortality)) in ti,ab
 neural network* with (recurrence or relapse or surviv* or death* or mortality) in ti,ab
 Prognos* in ds,cb,mi
 Survival in ds
 Relapse in ds
 Mortality in ds
 Recurrence* in ds
 Recurrent in ds

Model in ds
 Breast adenocarcinoma in ds or neoplastic disease in ds
 breast cancer in ti,ab,mi,ds
 Metastasis in ti,ds
 Metastases in ti,ds
 Animals in tn
 Humans in tn
 Exact {letter} in dt or exact¹²⁹ in dt
 #1 or #2 or #3 or #4
 #5 and (#6 or #7 or #8 or #9 or #10) and #11
 #19 or #20
 #12 or #13
 #21 and #22
 #16 not (#16 and #17)
 #23 not (#14 or #15 or #24)
 ((oestrogen receptor*) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab
 ((estrogen receptor*) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab
 ((progesterone receptor*) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab
 ((erbb2 or c erbb 2 or cerbb2) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab
 ((her 2 or her2 or neu or p53 or bcl2) near (death or recurrence or survival or relapse or prognosis or mortality)) in ti,ab
 #26 or #27 or #28 or #29 or #30
 Estrogen* in cb
 Progesterone Receptor* in cb
 Erbb2* in cb
 Cerb* in cb
 #32 or #33 or #34 or #35
 Survival in ds
 Relapse in ds
 Mortality in ds
 Recurrence* in ds
 Recurrent in ds
 #37 or #38 or #39 or #40 or #41
 #36 and #42
 #22 and (#31 or #43)
 #44 not (#14 or #15 or #24 or #25)
 tnm staging in md
 cancer staging in ti,ab
 Mitosis in cb
 Ploidy in cb
 Ploidies in cb
 Cathepsin d in cb
 #46 or #47 or #48 or #49 or #50 or #51 or #36
 #52 and #5 and #42
 ((prognostic variable*) near (death or recurrence or survival or relapse or prognosis)) in ti,ab

((prognostic factor*) near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (tnm staging near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (histologic grade near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (histologic type near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (mitotic near count* near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (hormone receptor status near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 ((cerbb2 or erbb2)near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (her2 near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (proliferation marker* near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (invasion near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (p53 near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (dna ploidy analysis near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (microvessel density near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (epidermal growth factor receptor* near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 (transforming growth factor near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 ((bc12 or bc1 2 or ps2) near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 ((cathepsin d) near prognos* near (death or recurrence or survival or relapse or prognosis)) in ti,ab
 #54 or #55 or #56 or #57 or #58 or #59 or #60 or #61 or #62 or #63 or #64 or #65 or #66 or #67 or #68 or #69 or #70 or #71
 #53 or #72
 "meta-analysis"/ all subheadings
 metaanalys* in ti,ab
 meta-analys* in ti,ab
 meta analys* in ti,ab
 cochrane in ti,ab
 (review* or overview*) in ti

review in dt
 (synthes* near3 (literature* or research* or studies or data)) in ti,ab
 (pooled analys*) in ti,ab
 (data near2 pool*) and studies
 (medline or medlars or embase or cinahl or scisearch or psychinfo or psycinfo or psychlit or psyclit) in ti,ab
 ((hand or manual or database* or computer*) near2 search*) in ti,ab
 ((electronic or bibliographic*) near2 (database* or data base*)) in ti,ab
 ((review* or overview*) near10 (systematic* or methodologic* or quantitativ* or research* or literature* or studies or trial* or effective*)) in ab
 #74 or #75 or #76 or #77 or #78 or #79 or #80 or #81 or #82 or #83 or #84 or #85 or #86 or #87
 #88 and #73
 #89 not (#45 or #25 or #14 or #15 or #24)
 #90 and #22

EMBASE

The EMBASE strategy was developed for the WinSPIRS interface and was run at Oxford.

SilverPlatterASCII 3.0WINNEMBASE (R) 2001/07-2001/10

(quality or scoring or score* or apprais* or usefulness or reliability) with (prognostic stud*)
 (quality or scoring or score* or apprais* or usefulness or reliability) with prognos* with research
 (quality or scoring or score* or apprais* or usefulness or reliability) with prognos* with article*
 (quality or scoring or score* or apprais* or usefulness or reliability) with prognos* with methodolog*
 (quality or scoring or score* or apprais* or usefulness or reliability) with prognos* with conduct*
 #1 or #2 or #3 or #4 or #5
 (prognosis or prognostic) in ti
 explode "Evidence-Based-Medicine"/ all subheadings
 #7 and #8
 checklist* or quality scores or quality assessment or check list* or criteria
 #10 with #7
 #6 or #9 or #11
 "Quality-of-Life"/ all subheadings
 (quality near life) in ti,ab
 #13 or #14
 #12 not #15

Topic D. Quality assessment of prognostic studies (not restricted to cancer)

The CancerLit search was undertaken on two different versions. The period 1995–2001/03 was searched on the OVID CD-Rom version of CancerLit. The period before 1995 was searched on the version of CancerLit offered by the Dialog database service.

The CancerLit search strategy was as follows.

SilverPlatterASCII 3.0WINNCANCERLIT 1998–2001/03

((quality or scoring or score* or apprais* or usefulness or reliability) with (prognostic stud*))
 ((quality or scoring or score* or apprais* or usefulness or reliability) with prognos* with research
 (quality or scoring or score* or apprais* or usefulness or reliability) with prognos* with article*
 (quality or scoring or score* or apprais* or usefulness or reliability) with prognos* with methodolog*
 (quality or scoring or score* or apprais* or usefulness or reliability) with prognos* with conduct*
 #1 or #2 or #3 or #4 or #5
 (prognosis or prognostic) in ti
 "Evidence-Based-Medicine"/ all subheadings
 #7 and #8
 checklist* or quality scores or quality assessment or check list* or criteria
 #10 with #7
 #6 or #9 or #11
 explode "Quality-of-Life"/ all subheadings
 (quality near life) in ti,ab
 #13 or #14
 #12 not #15

MEDLINE

OVID MEDLINE on CD-ROM was searched for the period 1980–August 2001.

#quality assessment

((quality or scoring or score\$ or apprais\$ or usefulness or reliability) adj4 prognostic stud\$).ti,ab.
 ((quality or scoring or score\$ or apprais\$ or usefulness or reliability) adj4 prognos\$ adj4 research).ti,ab.
 ((quality or scoring or score\$ or apprais\$ or usefulness or reliability) adj4 prognos\$ adj4 article).ti,ab.

((quality or scoring or score\$ or apprais\$ or usefulness or reliability) adj4 prognos\$ adj4 methodolog\$).ti,ab.

((quality or scoring or score\$ or apprais\$ or usefulness or reliability) adj4 prognos\$ adj4 conduct\$).ti,ab.

or/1-5

(prognosis or prognostic).ti.

evidence based medicine/

7 and 8

((checklist\$ or quality scores or quality assessment or check list\$ or criteria) adj4 (prognosis or prognostic)).ti.

6 or 9 or 10

exp quality of life/

(quality adj4 life).ti,ab.

or/12-13

11 not 14

BIOSIS

A BIOSIS strategy was produced for use with the WinSPIRS interface. These searches were carried out in Oxford.

SilverPlatterASCII 3.0WINNEMBASE (R) 2001/07–2001/10

((quality or scoring or score* or apprais* or usefulness or reliability) with (prognostic stud*))

((quality or scoring or score* or apprais* or usefulness or reliability) with prognos* with research

(quality or scoring or score* or apprais* or usefulness or reliability) with prognos* with article*

(quality or scoring or score* or apprais* or usefulness or reliability) with prognos* with methodolog*

(quality or scoring or score* or apprais* or usefulness or reliability) with prognos* with conduct*

((prognosis or prognostic) with (checklist* or quality scores or quality assessment or check list* or criteria)) in ti,ab

#1 or #2 or #3 or #4 or #5 or #6

(quality near life) in ti,ab

#7 not #8

Topic E. Clinical use of prognostic information in breast and other cancers

The CancerLit search was undertaken on two different versions. The period 1995–2001/03 was searched on the OVID CD-Rom version of CancerLit. The period before 1995 was searched

on the version of CancerLit offered by the Dialog database service.

**SilverPlatterASCII 3.0WINNEMBASE (R)
2001/07-2001/10**

(prognos* near (index or indices or indexes) near (relapse or recurrence or surviv* or death* or mortality)) in ti,ab

(model* near prognos* near (relapse or recurrence or surviv* or death* or mortality)) in ti,ab

(predictive model* near (recurrence or relapse or surviv* or death* or mortality)) in ti,ab

neural network* with (recurrence or relapse or surviv* or death* or mortality) in ti,ab

"Prognosis"/ all subheadings

explode "Survival"/ all subheadings

"Relapse"/all subheadings

"Mortality"/all subheadings or "cancer-mortality"/all subheadings

"Recurrent-Disease"/all subheadings

"Artificial-Neural-Network"

"Statistical-Model"/all subheadings

explode "Breast-Cancer"/ all subheadings

breast cancer in ti,ab

Metasta* in ti

explode "Metastasis"/all subheadings

exact {LETTER} in dt or "letter"/all subheadings

(animal or cell line* or vitro or invitro or cell or rat or rats or mouse or mice) in ti

#1 or #2 or #3 or #4

#5 and (#6 or #7 or #8 or #9) and (#10 or #11)

#18 or #19

#12 or #13

#20 and #21

#22 not (#14 or #15 or #16 or #17)

prognostic factors in ti,ab

prognostic index in ti,ab

prognostic indices in ti,ab

prognostic indexes in ti,ab

prognostic indicator* in ti,ab

((prognosis or prognostic) with information) in ti,ab

(prognostic marker* or genetic marker* or genetic predisposition) in ti,ab

((prognosis or prognostic) with data) in ti,ab

((prognosis or prognostic) with database*) in ti,ab

((prognosis or prognostic) with datafile*) in ti,ab

((prognosis or prognostic) with databank*) in ti,ab

#24 or #25 or #26 or #27 or #28 or #29 or #30 or #31 or #32 or #33 or #34

(clinical use or monitoring practice or routine practice) in ti,ab

(clinical setting* or audit*) in ti,ab

clinical governance in ti,ab

(treatment with (choice or decision*)) in ti,ab

(decision making or decision analysis) in ti,ab

(clinically useful or clinical usefulness) in ti,ab

(clinical utility or clinical value) in ti,ab

(clinical impact or clinical decision*) in ti,ab

#35 and (#36 or #37 or #38 or #39 or #40 or #41 or #42 or #43)

explode "neoplasm"/ all subheadings

#44 and #45

#46 not (#23 or #16 or #17)

MEDLINE

OVID MEDLINE on CD-ROM was searched for the period 1990–August 2001.

#clinical setting

(prognos\$ adj4 (index or indices or indexes) adj4

(relapse or recurrence or surviv\$ or death\$ or mortality)).ti,ab.

(model\$ adj4 prognos\$ adj4 (relapse or recurrence or survival\$ or death\$ or mortality)).ti,ab.

(predictive model\$ adj4 (recurrence or relapse or surviv\$ or death\$ or mortality)).ti,ab.

(neural network\$ adj4 (recurrence or relapse or surviv\$ or death\$ or mortality)).ti,ab.

exp prognosis/

survival rate/

disease free survival/

mortality/

recurrence/

neural networks computer/

exp survival analysis/

exp models statistical/

breast neoplasms/

mammary neoplasms/

breast cancer.ti,ab.

breast neoplasms/sc

metasta\$.ti.

exp neoplasms/sc

exp neoplasm metastasis/

letter.pt.

comment.pt.

(animal or cell line\$ or vitro or invitro or cell or rat or rats or mouse or mice).ti.

or/1-4

5 and (6 or 7 or 8 or 9) and (10 or 11 or 12)

23 or 24

13 or 14 or 15

25 and 26

27 not (16 or 17 or 18 or 19 or 20 or 21 or 22)

prognostic factors.ti,ab.

prognostic index.ti,ab.

prognostic indices.ti,ab.

prognostic indexes.ti,ab.

prognostic indicator\$.ti,ab.

((prognostic or prognosis) adj4 information).ti,ab.

(prognostic marker\$ or genetic marker\$ or genetic predisposition).ti,ab.

((prognostic or prognosis) adj4 data).ti,ab.

((prognostic or prognosis) adj4 database\$).ti,ab.
 ((prognostic or prognosis) adj4 datafile\$).ti,ab.
 ((prognostic or prognosis) adj4 databank\$).ti,ab.
 or/29-39
 (monitoring practice or routine practice).ti,ab.
 (clinical setting\$ or audit\$).ti,ab.
 clinical governance.ti,ab.
 (treatment adj4 (choice or decision\$)).ti,ab.
 (decision making or decision analysis).ti,ab.
 (clinically useful or clinical usefulness).ti,ab.
 (clinical utility or clinical value).ti,ab.
 (clinical impact or clinical decision\$).ti,ab.
 40 and (41 or 42 or 43 or 44 or 45 or 46 or 47 or 48)
 exp neoplasms/
 49 and 50
 51 not (28 or 20 or 21 or 22)

BIOSIS

A BIOSIS strategy was produced for use with the WinSPIRS interface. These searches were carried out in Oxford.

SilverPlatterASCII 3.0WINNEMBASE (R) 2001/07-2001/10

(prognos* near (index or indices or indexes) near (relapse or recurrence or surviv* or death* or mortality)) in ti,ab
 (model* near prognos* near (relapse or recurrence or surviv* or death* or mortality)) in ti,ab
 (predictive model* near (recurrence or relapse or surviv* or death* or mortality)) in ti,ab
 neural network* with (recurrence or relapse or surviv* or death* or mortality) in ti,ab
 Prognos* in ds,cb,mi
 Survival in ds
 Relapse in ds
 Mortality in ds
 Recurrence* in ds
 Recurrent in ds
 Model in ds
 Breast adenocarcinoma in ds or neoplastic disease in ds
 breast cancer in ti,ab,mi,ds
 Metastasis in ti,ds
 Metastases in ti,ds
 Animals in tn
 Humans in tn
 Exact {letter} in dt or exact ¹²⁹ in dt
 #1 or #2 or #3 or #4
 #5 and (#6 or #7 or #8 or #9 or #10) and #11
 #19 or #20
 #12 or #13
 #21 and #22
 #16 not (#16 and #17)
 #23 not (#14 or #15 or #24)

prognostic factors in ti,ab
 prognostic index in ti,ab
 prognostic indices in ti,ab
 prognostic indexes in ti,ab
 prognostic indicator* in ti,ab
 ((prognosis or prognostic) with information) in ti,ab
 (prognostic marker* or genetic marker* or genetic predisposition) in ti,ab
 ((prognosis or prognostic) with data) in ti,ab
 ((prognosis or prognostic) with database*) in ti,ab
 ((prognosis or prognostic) with datafile*) in ti,ab
 ((prognosis or prognostic) with databank*) in ti,ab
 #26 or #27 or #28 or #29 or #30 or #31 or #32 or #33 or #34 or #35 or #36
 (clinical use or monitoring practice or routine practice) in ti,ab
 (clinical setting* or audit*) in ti,ab
 clinical governance in ti,ab
 (treatment with (choice or decision*)) in ti,ab
 (decision making or decision analysis) in ti,ab
 (clinically useful or clinical usefulness) in ti,ab
 (clinical utility or clinical value) in ti,ab
 (clinical impact or clinical decision*) in ti,ab
 #37 and (#38 or #39 or #40 or #41 or #42 or #43 or #44 or #45)
 (neoplasm or neoplasms or cancer) in ti,ab,ds,cb,mi
 #46 and #47
 #48 not (#25 or #14 or #15 or #24)

EMBASE

The EMBASE strategy was developed for the WinSPIRS interface and was run at Oxford.

SilverPlatterASCII 3.0WINNEMBASE (R) 2001/07-2001/10

(prognos* near (index or indices or indexes) near (relapse or recurrence or surviv* or death* or mortality)) in ti,ab
 (model* near prognos* near (relapse or recurrence or surviv* or death* or mortality)) in ti,ab
 (predictive model* near (recurrence or relapse or surviv* or death* or mortality)) in ti,ab
 neural network* with (recurrence or relapse or surviv* or death* or mortality) in ti,ab
 "Prognosis"/ all subheadings
 explode "Survival"/ all subheadings
 "Relapse"/all subheadings
 "Mortality"/all subheadings or "cancer-mortality"/all subheadings
 "Recurrent-Disease"/all subheadings
 "Artificial-Neural-Network"
 "Statistical-Model"/all subheadings
 explode "Breast-Cancer"/ all subheadings
 breast cancer in ti,ab

Metasta* in ti
 explode "Metastasis"/all subheadings
 exact {LETTER} in dt or "letter"/all subheadings
 (animal or cell line* or vitro or invitro or cell or
 rat or rats or mouse or mice) in ti
 #1 or #2 or #3 or #4
 #5 and (#6 or #7 or #8 or #9) and (#10 or #11)
 #18 or #19
 #12 or #13
 #20 and #21
 #22 not (#14 or #15 or #16 or #17)
 prognostic factors in ti,ab
 prognostic index in ti,ab
 prognostic indices in ti,ab
 prognostic indexes in ti,ab
 prognostic indicator* in ti,ab
 ((prognosis or prognostic) with information) in ti,ab
 (prognostic marker* or genetic marker* or genetic
 predisposition) in ti,ab
 ((prognosis or prognostic) with data) in ti,ab
 ((prognosis or prognostic) with database*) in ti,ab
 ((prognosis or prognostic) with datafile*) in ti,ab
 ((prognosis or prognostic) with databank*) in ti,ab
 #24 or #25 or #26 or #27 or #28 or #29 or #30
 or #31 or #32 or #33 or #34
 (clinical use or monitoring practice or routine
 practice) in ti,ab
 (clinical setting* or audit*) in ti,ab
 clinical governance in ti,ab
 (treatment with (choice or decision*)) in ti,ab
 (decision making or decision analysis) in ti,ab
 (clinically useful or clinical usefulness) in ti,ab
 (clinical utility or clinical value) in ti,ab
 (clinical impact or clinical decision*) in ti,ab
 #35 and (#36 or #37 or #38 or #39 or #40 or
 #41 or #42 or #43)
 explode "neoplasm"/ all subheadings
 #44 and #45
 #46 not (#23 or #16 or #17)

Topic F. Quality of life, cost and cost-effectiveness relevant to the modelling

A two-part search approach was used:

- Part 1: cost estimates of breast cancer treatments (adjuvant and non-adjuvant), cost-effectiveness estimates and decision analysis models of breast cancer treatments
- Part 2: costs and cost-effectiveness models incorporating prognostic models (not specific to breast cancer).

I. Databases

Ia. MEDLINE

The OVID CD-ROM version was searched from 1966 to 2000/12.

Two searches were undertaken to capture the two aspects of the question.

Part 1

(breast adj cancer\$.ti,ab.
 breast neoplasms/
 or/1-2
 cost-benefit analysis/
 exp models, economic/
 exp decision trees/
 ((cost-effect\$ or cost) adj effective\$.ti,ab.
 (resource adj use).tw.
 (resource adj util\$).tw.
 eq5d.tw.
 (eq adj 5d).tw.
 qwb.tw.
 (quality adj3 well\$).tw.
 hui.tw.
 (health adj utilit\$).tw.
 (life adj year\$).tw.
 (quality adj adjusted).tw.
 (decision adj analysis).tw.
 (decision adj analytic).tw.
 (monte adj carlo).ti,ab.
 markov.tw.
 (simulation adj model\$).tw.
 (cost adj utilit\$).tw.
 (utility adj value\$).tw.
 (weight\$ adj3 preference\$).tw.
 euroqol.tw.
 (((visual adj analog) or visual) adj analogue).tw.
 (standard adj gamble).tw.
 (time adj trade).tw.
 ((qtwtist or q) adj twist).tw.
 (economic adj evaluation\$).tw.
 (quality adj3 life adj value\$).tw.
 (quality adj3 life adj measure\$).tw.
 (valu\$ adj quality).tw.
 (measur\$ adj quality adj3 life).tw.
 (quality adj3 life adj scale\$).tw.
 quality-adjusted life years/
 qaly\$.tw.
 exp *mass screening/
 exp *mammography/
 mammography.ti.
 or/4-38
 or/39-41
 3 and 42
 44 not 43
 limit 45 to english language
 editorial.pt.
 letter.pt.

comment.pt.
or/47-49
46 not 50

Part 2

This search excludes items already found in the Part 1 search.

#oxford hta review: second strategy
(breast adj cancer\$).ti,ab.
breast neoplasms/
or/1-2
cost-benefit analysis/
exp models, economic/
exp decision trees/
((cost-effect\$ or cost) adj effective\$).ti,ab.
(resource adj use).tw.
(resource adj util\$).tw.
eq5d.tw.
(eq adj 5d).tw.
qwb.tw.
(quality adj3 well\$).tw.
hui.tw.
(health adj utilit\$).tw.
(life adj year\$).tw.
(quality adj adjusted).tw.
(decision adj analysis).tw.
(decision adj analytic).tw.
(monte adj carlo).ti,ab.
markov.tw.
(simulation adj model\$).tw.
(cost adj utilit\$).tw.
(utility adj value\$).tw.
(weight\$ adj3 preference\$).tw.
euroqol.tw.
(((visual adj analog) or visual) adj analogue).tw.
(standard adj gamble).tw.
(time adj trade).tw.
((qtwtist or q) adj twist).tw.
(economic adj evaluation\$).tw.
(quality adj3 life adj value\$).tw.
(quality adj3 life adj measure\$).tw.
(valu\$ adj quality).tw.
(measur\$ adj quality adj3 life).tw.
(quality adj3 life adj scale\$).tw.
quality-adjusted life years/
qaly\$.tw.
exp *mass screening/
exp *mammography/
mammography.ti.
or/4-38
or/39-41
3 and 42
44 not 43
limit 45 to english language
editorial.pt.
letter.pt.

comment.pt.
or/47-49
46 not 50
or/4-9,18-22
(prognostic adj (model\$ or tool\$ or marker\$ or factor\$ or indicator\$)).ti,ab.
(model\$ adj prognosis).ti,ab.
(probability adj3 (recurrence or survival or mortality or recovery or death)).ti,ab.
(disease adj free adj rate\$).ti,ab.
(predicti\$ adj3 (outcome\$ or death or survival or mortality or recovery or recurrence or complication\$)).ti,ab.
(disease adj free adj survival).ti,ab.
(predictive adj ability).ti,ab.
(survival adj function\$).ti,ab.
(predictive adj model\$).ti,ab.
or/53-61
life tables/
proportional hazards models/
disease-free survival/
prognosis/
risk/
or/63-65
or/66-67
68 and 69
or/62,70
52 and 71
limit 72 to english language
73 not (50 or 51)

1b. EMBASE

The following strategies were prepared for Oxford (WinSPIRS EMBASE) to run locally:

Part 1

SilverPlatterASCII 3.0WINNEMBASE (R)
2001/01-2001/02
breast cancer* in ti,ab
explode "breast-cancer"/ all subheadings
explode "breast-tumor"/ all subheadings
"cost-effectiveness-analysis"/ all subheadings
"economic-evaluation"/ all subheadings
"cost-utility-analysis"/ all subheadings
"model"/ all subheadings
"decision-theory"/ all subheadings
"quality-adjusted-life-year"/ all subheadings
"system-analysis"/ all subheadings
"probability"/ all subheadings
"simulation"/ all subheadings
(cost-effect* or cost effective*) in ti,ab
resource use in ti,ab
resource util* in ti,ab
eq5d in ti,ab
eq 5d in ti,ab
qwb in ti,ab

(quality near well*) in ti,ab
 hui in ti,ab
 health utilit* in ti,ab
 life year* in ti,ab
 quality adjusted in ti,ab
 decision analysis in ti,ab
 decision analytic in ti,ab
 monte carlo in ti,ab
 markov in ti,ab
 simulation model* in ti,ab
 cost utilit* in ti,ab
 utility value* in ti,ab
 weight* near preference* in ti,ab
 euroqol in ti,ab
 (visual analog or visual analogue) in ti,ab
 standard gamble in ti,ab
 time trade in ti,ab
 qtwist or q twist in ti,ab
 economic evaluation* in ti,ab
 (quality near life value*) in ti,ab
 (quality near life measure*) in ti,ab
 valu* quality in ti,ab
 (measur* quality near life) in ti,ab
 (quality near life scale*) in ti,ab
 qaly* in ti,ab
 mammography in ti
 "cancer-screening"/ all subheadings
 explode "mammography"/ all subheadings
 "editorial"/ all subheadings
 "letter"/ all subheadings
 #1 or #2 or #3
 #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11
 #12 or #13 or #14 or #15 or #16 or #17 or #18
 or #19 or #20
 #21 or #22 or #23 or #24 or #25 or #26 or #27
 or #28 or #29
 #30 or #31 or #32 or #33 or #34 or #35 or #36
 or #37 or #38 or #39 or #40 or #41 or #42 or
 #43
 #50 or #51 or #52 or #53
 #44 or #45 or #46
 #49 and #54
 #56 not #55
 #47 or #48
 #57 not #58
 LA = "ENGLISH"
 #59 and (LA = "ENGLISH")

Part 2

SilverPlatterASCII 3.0WINNEMBASE (R)
 2001/01-2001/02
 breast cancer* in ti,ab
 explode "breast-cancer"/ all subheadings
 explode "breast-tumor"/ all subheadings
 "cost-effectiveness-analysis"/ all subheadings
 "economic-evaluation"/ all subheadings
 "cost-utility-analysis"/ all subheadings

"model"/ all subheadings
 "decision-theory"/ all subheadings
 "quality-adjusted-life-year"/ all subheadings
 "system-analysis"/ all subheadings
 "probability"/ all subheadings
 "simulation"/ all subheadings
 (cost-effect* or cost effective*) in ti,ab
 resource use in ti,ab
 resource util* in ti,ab
 eq5d in ti,ab
 eq 5d in ti,ab
 qwb in ti,ab
 (quality near well*) in ti,ab
 hui in ti,ab
 health utilit* in ti,ab
 life year* in ti,ab
 quality adjusted in ti,ab
 decision analysis in ti,ab
 decision analytic in ti,ab
 monte carlo in ti,ab
 markov in ti,ab
 simulation model* in ti,ab
 cost utilit* in ti,ab
 utility value* in ti,ab
 weight* near preference* in ti,ab
 euroqol in ti,ab
 (visual analog or visual analogue) in ti,ab
 standard gamble in ti,ab
 time trade in ti,ab
 qtwist or q twist in ti,ab
 economic evaluation* in ti,ab
 (quality near life value*) in ti,ab
 (quality near life measure*) in ti,ab
 valu* quality in ti,ab
 (measur* quality near life) in ti,ab
 (quality near life scale*) in ti,ab
 qaly* in ti,ab
 mammography in ti
 "cancer-screening"/ all subheadings
 explode "mammography"/ all subheadings
 "editorial"/ all subheadings
 "letter"/ all subheadings
 #1 or #2 or #3
 #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11
 #12 or #13 or #14 or #15 or #16 or #17 or #18
 or #19 or #20
 #21 or #22 or #23 or #24 or #25 or #26 or #27
 or #28 or #29
 #30 or #31 or #32 or #33 or #34 or #35 or #36
 or #37 or #38 or #39 or #40 or #41 or #42 or
 #43
 #50 or #51 or #52 or #53
 #44 or #45 or #46
 #49 and #54
 #56 not #55
 #47 or #48
 #57 not #58

LA = "ENGLISH"
 #59 and (LA = "ENGLISH")
 prognostic model* or prognostic tool* or
 prognostic marker* or prognostic factor* or
 prognostic indicator* in ti,ab
 model* prognosis in ti,ab
 probability near (recurrence or survival or
 mortality or recovery or death) in ti,ab
 disease free rate* in ti,ab
 predicti* near (outcome* or death or survival or
 mortality or recovery or recurrence or
 complication*) in ti,ab
 disease free survival in ti,ab
 predictive ability in ti,ab
 survival function* in ti,ab
 predictive model* in ti,ab
 "prognosis"/ all subheadings
 "risk-factor"/ all subheadings
 "recurrence-risk"/ all subheadings
 explode "survival"/ all subheadings
 "cancer-mortality"/ all subheadings
 "prediction"/ all subheadings
 "life-table"/ all subheadings
 "nonbiological-model"/ all subheadings
 #4 or #5 or #7 or #8 or #11 or #12 or #13 or
 #14 or #24 or #25 or #26 or #27 or #28
 #71 or #72 or #73
 #74 or #75 or #76 or #77 or #78
 #62 or #63 or #64 or #65 or #66 or #67 or #68
 or #69 or #70
 #80 and #81
 #82 or #83
 #79 and #84
 #85 not #61
 LA = "ENGLISH"
 #86 and (LA = "ENGLISH")

1c. CCTR

CCTR was searched on Issue 2001/1 of the
 Cochrane Library.

Two searches were carried out.

Part 1

prog1

(BREAST next CANCER)
 BREAST-NEOPLASMS*:ME
 (#1 or #2)
 COST-BENEFIT-ANALYSIS*:ME
 MODELS-ECONOMIC*:ME
 DECISION-TREES:ME
 (COST-EFFECT* or (COST next EFFECTIVE*))
 "RESOURCE UTIL*"
 (EQ5D or QWB)
 (QUALITY near WELLBEING:TI) or (QUALITY
 near WELLBEING:AB)
 (HUI or (HEALTH next UTILIT*))

((LIFE next YEAR*:TI) or (LIFE next YEAR*:AB))
 ((QUALITY next ADJUSTED:TI) or (QUALITY
 next ADJUSTED:AB))
 ((DECISION next ANALY*) or (MONTE next
 CARLO))
 markov:TI or markov:AB
 ((SIMULATION next MODEL*) or (COST next
 UTILIT*))
 (UTILITY next VALUE*)
 ((WEIGHT* near PREFERENCE*:TI) or
 (WEIGHT* near PREFERENCE*:TI))
 (EUROQOL or (VISUAL next ANALOG*))
 ((STANDARD next GAMBLE) or (TIME next
 TRADE))
 (QTWIST or (ECONOMIC next EVALUATION*))
 (QUALITY near VALUE*:TI) or (QUALITY near
 VALUE*:ab)
 (QUALITY near MEASUR*:TI) OR (QUALITY
 near MEASUR*:AB)
 (VALU* next QUALITY)
 (QUALITY near SCALE*:TI)
 QALY*
 (QUALITY near SCALE*:AB)
 mammography:TI
 mass-screening*:ME
 mammography*:ME

Part 2

(BREAST next CANCER)
 BREAST-NEOPLASMS*:ME
 (#1 or #2)
 COST-BENEFIT-ANALYSIS*:ME
 MODELS-ECONOMIC*:ME
 DECISION-TREES:ME
 (COST-EFFECT* or (COST next EFFECTIVE*))
 (RESOURCE and UTIL*)
 (EQ5D or QWB)
 ((QUALITY near WELLBEING:TI) or (QUALITY
 near WELLBEING:AB))
 (HUI or (HEALTH next UTILIT*))
 ((LIFE next YEAR*:TI) or (LIFE next YEAR*:AB))
 ((QUALITY next ADJUSTED:TI) or (QUALITY
 next ADJUSTED:AB))
 ((DECISION next ANALY*) or (MONTE next
 CARLO))
 (MARKOV:TI or MARKOV:AB)
 ((SIMULATION next MODEL*) or (COST next
 UTILIT*))
 (UTILITY next VALUE*)
 ((WEIGHT* near PREFERENCE*:TI) or
 (WEIGHT* near PREFERENCE*:TI))
 (EUROQOL or (VISUAL next ANALOG*))
 ((STANDARD next GAMBLE) or (TIME next
 TRADE))
 (QTWIST or (ECONOMIC next EVALUATION*))
 ((QUALITY near VALUE*:TI) or (QUALITY near
 VALUE*:AB))

((QUALITY near MEASUR*:TI) or (QUALITY near MEASUR*:AB))
 (VALU* next QUALITY)
 (QUALITY near SCALE*:TI)
 QALY*
 (QUALITY near SCALE*:AB)
 MAMMOGRAPHY:TI
 MASS-SCREENING*:ME
 MAMMOGRAPHY*:ME
 ((PROGNOSTIC next MODEL*:TI) or (PROGNOSTIC next MODEL*:AB))
 ((PROGNOSTIC next TOOL*:TI) or (PROGNOSTIC next TOOL*:AB))
 ((PROGNOSTIC next MARKER*:TI) or (PROGNOSTIC next MARKER*:AB))
 ((PROGNOSTIC next FACTOR*:TI) or (PROGNOSTIC next FACTOR*:AB))
 ((PROGNOSTIC next INDICATOR*:TI) or (PROGNOSTIC next INDICATOR*:AB))
 ((MODEL* next PROGNOSIS:TI) or (MODEL* next PROGNOSIS:AB))
 ((PROBABILITY near RECURRENCE:TI) or (PROBABILITY near RECURRENCE:AB))
 ((PROBABILITY near SURVIVAL:TI) or (PROBABILITY near SURVIVAL:AB))
 ((PROBABILITY near MORTALITY:TI) or (PROBABILITY near MORTALITY:AB))
 ((PROBABILITY near RECOVERY:TI) or (PROBABILITY near RECOVERY:AB))
 ((PROBABILITY near DEATH:TI) or (PROBABILITY near DEATH:AB))
 (((DISEASE next FREE) next RATE*:TI) or ((DISEASE next FREE) next RATE*:AB))
 ((PREDICT* near OUTCOME*:TI) or (PREDICT* near OUTCOME*:AB))
 ((PREDICT* near DEATH:TI) or (PREDICT* near DEATH:AB))
 ((PREDICT* near SURVIVAL:TI) or (PREDICT* near SURVIVAL:AB))
 ((PREDICT* near MORTALITY:TI) or (PREDICT* near MORTALITY:AB))
 ((PREDICT* near RECOVERY:TI) or (PREDICT* near RECOVERY:AB))
 ((PREDICT* near RECURRENCE:TI) or (PREDICT* near RECURRENCE:AB))
 ((PREDICT* near COMPLICATIONS:TI) or (PREDICT* near COMPLICATIONS:AB))
 (((DISEASE next FREE) next SURVIVAL:TI) or ((DISEASE next FREE) next SURVIVAL:AB))
 ((PREDICTIVE next ABILITY:TI) or (PREDICTIVE next ABILITY:AB))
 ((SURVIVAL next FUNCTION*:TI) or (SURVIVAL next FUNCTION*:AB))
 ((PREDICTIVE next MODEL*:TI) or (PREDICTIVE next MODEL*:AB))
 LIFE-TABLES*:ME
 PROPORTIONAL-HAZARDS-MODELS:ME

DISEASE-FREE-SURVIVAL:ME
 PROGNOSIS:ME
 RISK*:ME
 ((((((#4 or #5) or #6) or #7) or #8) or #14) or #15) or #16)
 ((((((((((((((((((((((#31 or #32) or #33) or #34) or #35) or #36) or #37) or #38) or #39) or #40) or #41) or #42) or #43) or #44) or #45) or #46) or #47) or #48) or #49) or #50) or #51) or #52) or #53)
 ((#54 or #55) or #56)
 (#57 or #58)
 (#62 and #63)
 (#61 or #64)
 (#60 and #65)
 ((((((((((((((((((((((#4 or #5) or #6) or #7) or #8) or #9) or #10) or #11) or #12) or #13) or #14) or #15) or #16) or #17) or #18) or #19) or #20) or #21) or #22) or #23) or #24) or #25) or #26) or #27)
 ((#28 or #29) or #30)
 (#3 and #67)
 (#69 not #68)
 (#66 not #70)

Notes on the CCTR search

1. The Cochrane Library does not allow searches on certain stopwords (e.g. use) or on words of less than three letters. This means the following search phrases were not possible: 'resource use', 'eq 5d', 'q twist'.

Id. Cochrane Cancer Network Controlled Trials Register

The CCTR strategy was suitable for use within this database, which was searched in Oxford.

Ie. BIOSIS

The Oxford team use the terms from the EconLit strategy on their local BIOSIS databases. These strategies were solely text word based and have no subject index terms. The strategies were prepared for WinSPIRS.

If. CancerLit

Searched CancerLit on Silverplatter 1995–2000/9.

Search strategy

SilverPlatterASCII 3.0WINNCANCERLIT
1995–1996

breast cancer in ti,ab
 explode "Breast-Neoplasms"/ all subheadings
 #1 or #2
 "Cost-Benefit-Analysis"/ all subheadings
 explode "Models-Economic"/ all subheadings
 "Decision-Trees"
 #4 or #5 or #6

(cost-effect* or cost effective*) in ti,ab
 (resource use or resource util*) in ti,ab
 eq5d in ti,ab
 eq 5d in ti,ab
 qwb in ti,ab
 quality near (well* in ti,ab)
 hui in ti,ab
 health utilit* in ti,ab
 life year* in ti,ab
 quality adjusted in ti,ab
 decision analysis in ti,ab
 decision analytic in ti,ab
 monte carlo in ti,ab
 markov in ti,ab
 simulation model* in ti,ab
 cost utilit* in ti,ab
 utility value* in ti,ab
 weight* near (preference* in ti,ab)
 euroqol in ti,ab
 (visual analog in ti,ab) or (visual analogue in ti,ab)
 standard gamble in ti,ab
 time trade in ti,ab
 (qtwist in ti,ab) or (q twist in ti,ab)
 economic evaluation* in ti,ab
 quality near life near (value* in ti,ab)
 quality near life near (measur* in ti,ab)
 valu* quality in ti,ab
 quality near life near (scale* in ti,ab)
 "Quality-Adjusted-Life-Years"
 qaly*
 explode "Mass-Screening"/ all subheadings
 explode "Mammography"/ all subheadings
 #38 or #39
 mammography in ti
 #7 or #8 or #9 or #10 or #11 or #12 or #13 or
 #14 or #15 or #16 or #17 or #18 or #19 or #20
 #22 or #23 or #24 or #25 or #26 or #27 or #28
 or #29 or #30 or #31 or #32 or #33
 #21 or #34 or #35 or #36 or #37
 #42 or #43 or #44
 #40 or #41
 #3 and #45
 #47 not #46
 #48 and (LA = "ENGLISH")
 exact{EDITORIAL} in PT
 exact{LETTER} in PT
 exact¹²⁹ in PT
 #50 or #51 or #52
 #49 not #53 (set 54)
 exact{MEDLINE} in SB
 #54 not #55 (set 56)
 #7 or #8 or #9 or #18 or #19 or #20 or #21 or
 #22
 prognostic model* in ti,ab
 prognostic tool* in ti,ab
 prognostic marker* in ti,ab
 prognostic factor* in ti,ab

prognostic indicator* in ti,ab
 model* prognosis in ti,ab
 probability near recurrence in ti,ab
 probability near survival in ti,ab
 probability near mortality in ti,ab
 probability near recovery in ti,ab
 probability near death in ti,ab
 disease free rate* in ti,ab
 predict* near outcome* in ti,ab
 predict* near death in ti,ab
 predict* near survival in ti,ab
 predict* near mortality in ti,ab
 predict* near recovery in ti,ab
 predict* near recurrence in ti,ab
 predict* near complications in ti,ab
 disease free survival in ti,ab
 predictive ability in ti,ab
 survival function* in ti,ab
 predictive model* in ti,ab
 explode "Life-Tables"/ all subheadings
 "Proportional-Hazards-Models"
 "Disease-Free-Survival"
 "Prognosis-"
 explode "Risk"/ all subheadings
 #81 or #82 or #83 or #84 or #85
 #58 or #59 or #60 or #61 or #62 or #63 or #64
 or #65 or #66 or #67 or #68
 #69 or #70 or #71 or #72 or #73 or #74 or #75
 or #76 or #77 or #78 or #79 or #80
 (#81 or #82 or #83) and (#84 or #85)
 #87 or #88 or #89
 #90 and #57
 #91 and (LA = "ENGLISH")
 #92 not #54 (set 93)
 #93 not #55 (set 94)

This search combines both aspects of the economic searches: sets 54/56 are the breast cancer-specific parts and sets 93/94 are the broader 'all cancer' modelling results (excluding results found in sets 54/56).

Notes on the CancerLit searches

1. CRD only has a 5-year subscription to CancerLit.
2. The use of the 'near' operator in several of the search lines finds words in any order.
3. The Silverplatter WinSPIRS version of CancerLit does not retain the major and minor subject headings option which is found on some implementations of MeSH (e.g. OVID). This means that the mammography exclusions are not focused only to studies where they form the focus of the publication (as with the MEDLINE strategy we have prepared). This may have the effect of removing some studies which are relevant but which have the

screening subject headings as minor descriptors.

4. CancerLit is a combination of records which appear on MEDLINE and records which are unique to CancerLit. The facility to remove records which are also in MEDLINE and therefore probably already picked up by the MEDLINE searches is available: select exact{MEDLINE} in SB and then NOT this set from the result set). In one of the CancerLit disks (1995–6) the exclusion produced 0 hits, which is hard to understand. It is assumed that a bug has affected that disk and the whole result set has been sent, i.e. sets 54 and 93 not sets 56 and 94. In the other years the limit to unique CancerLit records seems to have worked correctly and results for set 56 have been sent.

Ig. NHS Economic Evaluation Database

The administration database for the NHS EED database was searched on 2 May 2001. This database contains all candidate material for the public NHS EED database. Because NHS EED focuses on economic evaluations, a high-level search without reference to economics was undertaken. (s) searches for words within the same sentence; (w) searches for words adjacent to each other; + represents OR; & represents AND.

Part 1

S breast(w)neoplasm
S breast(w)cancer
S mammography or screening
S s1+s2
S s4&s3

Part 2

S prognostic(s)model*
S Prognostic(s)tool*
S prognostic(s)(marker* or factor* or indicator*)
S model*(s)prognosis
S probability(s)(recurrence or survival or mortality or recovery or death)
S disease(w)free(w)rate
S predict*(s)(outcome* or death or survival or mortality or recovery or recurrence or complication*)
S disease(w)free(w)survival
S predictive(w)ability
S survival(w)function*
S predictive(w)model*
S proportional(w)hazards(w)model*
S s1+s2+s3+s4+s5+s6+s7+s8+s9+s10+s11+s12
S cost(w)effective*
S economic(w)model*

S decision(w)tree*
S resource(w)(use or util*)
S decision(w)(analysis or analytic)
S monte(w)carlo or markov or simulation(w)model*
S s14+s15+s16+s17+s18+s19
S s13&s20

Ih. HEED

The April 2001 issue of HEED was searched.

1. AX = 'BREAST CANCER' OR 'BREAST NEOPLASMS'
2. EE = 'EFFECTIVENESS' OR EE = 'UTILITY'
3. AX = 'ECONOMIC MODEL' OR AX = 'DECISION TREE'
4. AX = 'RESOURCE USE' OR AX = 'RESOURCE UTILISATION'
5. AX = 'RESOURCE UTILIZATION' OR AX = EQ5D
6. AX = 'EQ 5D' OR AX = QWB OR AX = WELLBEING
7. AX = HUI OR AX = 'HEALTH UTILITIES' OR AX = 'LIFE YEAR'
8. AX = 'LIFE YEARS' OR AX = 'QUALITY ADJUSTED'
9. AX = 'DECISION ANALYSIS' OR AX = 'DECISION ANALYTIC'
10. AX = 'MONTE CARLO' OR AX = MARKOV OR AX = SIMULATION
11. AX = 'UTILITY VALUE' OR AX = 'UTILITY VALUES'
12. AX = 'WEIGHTED PREFERENCE' OR AX = 'WEIGHTED PREFERENCES'
13. AX = EUROQOL OR AX = ANALOG OR AX = ANALOGUE
14. AX = 'STANDARD GAMBLE'
15. AX = 'TIME TRADE' OR AX = QTWIST OR AX = 'Q TWIST'
16. AX = 'QUALITY OF LIFE' OR AX = 'VALUING QUALITY'
17. AX = QALY*
18. CS = 2 OR CS = 3 OR CS = 4 OR CS = 5 OR CS = 6 OR CS = 7 OR CS = 8 OR CS = 9 OR CS = 10
19. CS = 11 OR CS = 12 OR CS = 13 OR CS = 14 OR CS = 15 OR CS = 16
20. CS = 1 AND (CS = 18 OR CS = 19)
21. AX = 'BREAST CANCER SCREENING' OR AX = MAMMOGRAPHY
22. CS = 20 ANDNOT CS = 21
23. CS = 2 OR CS = 4 OR CS = 5 OR CS = 9 OR CS = 10
24. AX = PROGNOSTIC OR AX = PROGNOSIS OR AX = PROBABILITY
25. AX = 'DISEASE FREE RATES' OR AX = PREDICT*
26. AX = 'SURVIVAL FUNCTION'

27. AX='LIFE TABLES'
 28. AX='PROPORTIONAL HAZARDS'
 29. CS=24 OR CS=25 OR CS=26 OR CS=27
 OR CS=28
 30. CS=23 AND CS=29
 31. CS=30 ANDNOT CS=22

Set 22 is the results of the breast cancer search and set 31 is the result of the broader prognostic search.

li EconLit

The following strategies were prepared for WinSPIRS and forwarded to Oxford and were run locally.

Part 1

SilverPlatterASCII 3.0WINNEconLit 1969-2001/03

breast cancer*
 (cost-effect* or cost effective*)
 resource use
 resource util*
 eq5d
 eq 5d
 qwb
 (quality near well*)
 hui
 health utilit*
 life year*
 quality adjusted
 decision analysis
 decision analytic
 monte carlo
 markov
 simulation model*
 cost utilit*
 utility value*
 weight* near preference*
 euroqol
 (visual analog or visual analogue)
 standard gamble
 time trade
 qtwist or q twist
 economic evaluation*
 (quality near life value*)
 (quality near life measure*)
 valu* quality
 (measur* quality near life)
 (quality near life scale*)
 qaly*
 mammography in ti
 #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9
 or #10 or #11
 #12 or #13 or #14 or #15 or #16 or #17 or #18
 or #19 or #20
 #21 or #22 or #23 or #24 or #25 or #26 or #27
 or #28 or #29

- #30 or #31 or #32
 #34 or #35 or #36 or #37
 #1 and #38
 #39 not #33

Part 2

SilverPlatterASCII 3.0WINNEconLit 1969-2001/03

breast cancer*
 (cost-effect* or cost effective*)
 resource use
 resource util*
 eq5d
 eq 5d
 qwb
 (quality near well*)
 hui
 health utilit*
 life year*
 quality adjusted
 decision analysis
 decision analytic
 monte carlo
 markov
 simulation model*
 cost utilit*
 utility value*
 weight* near preference*
 euroqol
 (visual analog or visual analogue)
 standard gamble
 time trade
 qtwist or q twist
 economic evaluation*
 (quality near life value*)
 (quality near life measure*)
 valu* quality
 (measur* quality near life)
 (quality near life scale*)
 qaly*
 mammography in ti
 #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9
 or #10 or #11
 #12 or #13 or #14 or #15 or #16 or #17 or #18
 or #19 or #20
 #21 or #22 or #23 or #24 or #25 or #26 or #27
 or #28 or #29
 #30 or #31 or #32
 #34 or #35 or #36 or #37
 #1 and #38
 #39 not #33
 prognostic model* or prognostic tool* or
 prognostic marker* or prognostic factor* or
 prognostic indicator*
 model* prognosis
 probability near (recurrence or survival or
 mortality or recovery or death)
 disease free rate*

predicti* near (outcome* or death or survival or mortality or recovery or recurrence or complication*)
 disease free survival
 predictive ability
 survival function*
 predictive model*
 #2 or #3 or #4 or #13 or #14 or #15 or #16 or #17
 #41 or #42 or #43 or #44 or #45 or #46 or #47 or #48 or #49
 #50 and #51
 #52 not #40

The EMBASE strategy was developed for the WinSPIRS interface and was run at Oxford.

SilverPlatterASCII 3.0WINNEMBASE (R)
2001/07-2001/10

(prognos* near (index or indices or indexes) near (relapse or recurrence or surviv* or death* or mortality)) in ti,ab
 (model* near prognos* near (relapse or recurrence or surviv* or death* or mortality)) in ti,ab

(predictive model* near (recurrence or relapse or surviv* or death* or mortality)) in ti,ab
 neural network* with (recurrence or relapse or surviv* or death* or mortality) in ti,ab
 "Prognosis"/ all subheadings
 explode "Survival"/ all subheadings
 "Relapse"/all subheadings
 "Mortality"/all subheadings or "cancer-mortality"/all subheadings
 "Recurrent-Disease"/all subheadings
 "Artificial-Neural-Network"
 "Statistical-Model"/all subheadings
 explode "Breast-Cancer"/ all subheadings
 breast cancer in ti,ab
 Metasta* in ti
 explode "Metastasis"/all subheadings
 exact {LETTER} in dt or "letter"/all subheadings
 (animal or cell line* or vitro or invitro or cell or rat or rats or mouse or mice) in ti
 #1 or #2 or #3 or #4
 #5 and (#6 or #7 or #8 or #9) and (#10 or #11)
 #18 or #19
 #12 or #13
 #20 and #21
 #22 not (#14 or #15 or #16 or #17)

Appendix 2

Forms used to assess the eligibility of studies identified using searches for Topics A–E

Topic A. Prognostic models in breast cancer

	Yes	No
Abstract indicates that:		
• article is a primary study or an evaluation of a proposed model	<input type="checkbox"/>	<input type="checkbox"/>
• the model can be utilised to predict time to death or disease recurrence	<input type="checkbox"/>	<input type="checkbox"/>
• that at least 2 prognostic factors have been combined to formulate the model	<input type="checkbox"/>	<input type="checkbox"/>
• article is concerned with patients with first diagnosis, operable breast cancer	<input type="checkbox"/>	<input type="checkbox"/>

The definition of model for this project:

The incorporation of at least 2 disparate factors from either one across the following areas: clinical, pathological, and demographic. The model should have a broad perspective, and not narrowly focus on 2 very similar factors.

Article relevant to PRINCE project, search question A?

Article relevant to other PRINCE project search question?

If yes, please indicate which search question _____

Reviewer initials _____

Topic B. Predictive factors in breast cancer

[A study that investigates patient/tumour factors as possible predictors of response to treatment].

	Yes	No
Abstract indicates that:		
• the article is a review article or a single study [epidemiological or RCT]	<input type="checkbox"/>	<input type="checkbox"/>
• incorporates treatment [drug/chemosensitivity, drug/chemoresponsiveness, drug, interaction]; or 1+ prognostic factors/variables/markers as predictive of response [OS/DFS and/or mortality] to treatment	<input type="checkbox"/>	<input type="checkbox"/>
• different predictive features on a tumour behave/respond differently with different treatments/interventions	<input type="checkbox"/>	<input type="checkbox"/>
• article is concerned with patients with first diagnosis, operable breast cancer	<input type="checkbox"/>	<input type="checkbox"/>

Article relevant to PRINCE project, search question B? Yes No

Article relevant to other PRINCE project search question? Yes No

If yes, please indicate which search question _____

Reviewer initials _____

Topic C. Reviews of prognostic information in breast cancer

	Yes	No
Abstract indicates that:		
• Study is a review [overview, systematic or meta-analysis]	<input type="checkbox"/>	<input type="checkbox"/>
• Concerned with breast cancer [preferably early disease]	<input type="checkbox"/>	<input type="checkbox"/>
• Refers to survival and/or prognosis	<input type="checkbox"/>	<input type="checkbox"/>
• Review is concerned with patients with first diagnosis, operable breast cancer	<input type="checkbox"/>	<input type="checkbox"/>

Article relevant to PRINCE project, search question C? Yes No

Article relevant to other PRINCE project search question? Yes No

If yes, please indicate which search question _____

Reviewer initials _____

Topic D. Quality assessment of prognostic studies

Papers should be general and not specific to cancer.

		Yes	No
Abstract indicates that:			
• Paper presents an assessment scoring scheme checklist for the quality of studies of prognosis		<input type="checkbox"/>	<input type="checkbox"/>
		<input type="checkbox"/>	<input type="checkbox"/>
		<input type="checkbox"/>	<input type="checkbox"/>

More?

Article relevant to PRINCE project, search question D?

Article relevant to other PRINCE project search question?

If yes, please indicate which search question _____

Reviewer initials _____

Topic E. Clinical use of prognostic information in breast and other cancers

		Yes	No
Abstract indicates that:			
• Paper is concerned with the use of prognostic information in a large group [routine] clinical setting;		<input type="checkbox"/>	<input type="checkbox"/>
		<input type="checkbox"/>	<input type="checkbox"/>

More?

Article relevant to PRINCE project, search question E?

Article relevant to other PRINCE project search question?

If yes, please indicate which search question _____

Reviewer initials _____

Appendix 3

Forms used to extract data from studies assessed as eligible for Topics A, B and C

Topic A. Prognostic models in breast cancer

Only include papers that **fulfil** the following inclusion criteria:

	Yes	No
• development and/or validation of a model	<input type="checkbox"/>	<input type="checkbox"/>
• primary cancer	<input type="checkbox"/>	<input type="checkbox"/>
• non-metastatic and non-advanced disease [If includes patients with metastatic disease, this must not be included in analysis for prognostic model]	<input type="checkbox"/>	<input type="checkbox"/>
• operable b.c. where patients received surgery as initial treatment for their disease	<input type="checkbox"/>	<input type="checkbox"/>
• patient's receiving adjuvant therapy only	<input type="checkbox"/>	<input type="checkbox"/>

Paper Number: _____ Reviewer Initials: _____ Year of Publication: _____

Author of Paper: _____ Paper number of other papers that this may link: _____

Number of factors included in the model: _____ Names of these factors: _____

Category of Studies to be included: **1: Prognostic Model Development:** complete sections 1–12
2: Validation of prognostic model: complete sections 1–7 and 13–14
3: Combined development and validation of model: complete all sections

Pg. No	1 Study Characteristics		
	Are all 3 study dates [see below*] specified? [0 = no, 1 = yes, 2 = partially]		<input type="checkbox"/>
	Start of recruitment*	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>
	End of recruitment*	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>
	End of follow-up*	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>
	Focus of the paper [1 = development of a prognostic model, 2 = evaluation of an existing model, 3 = both]		<input type="checkbox"/>

	<p>2 Design</p> <p>Study design [1 = randomised clinical trial, 2 = non-randomised comparative study, 3 = cohort study (all patients treated same way) 4 = other] <input type="checkbox"/></p> <p>Design [1 = prospective, 2 = retrospective] <input type="checkbox"/></p> <p>Reason for sample size [0 = none given, 1 = power, 2 = justified time interval, 3 = (group) sequential design, 4 = other] <input type="checkbox"/></p> <p>Sample Size: * [Indicate sample size dependent on the category of study: model development, validation or both].</p> <table style="margin-left: 40px;"> <thead> <tr> <th></th> <th style="text-align: center;">Initial</th> <th style="text-align: center;">In analysis</th> </tr> </thead> <tbody> <tr> <td>*Developing Model</td> <td style="text-align: center;"><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/></td> <td style="text-align: center;"><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/></td> </tr> <tr> <td>*Validating Model</td> <td style="text-align: center;"><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/></td> <td style="text-align: center;"><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/></td> </tr> </tbody> </table> <p>Details provided of subjects rejected from the analysis? [0 = no, 1 = yes] <input type="checkbox"/></p>		Initial	In analysis	*Developing Model	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	*Validating Model	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
	Initial	In analysis								
*Developing Model	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>								
*Validating Model	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>								
	<p>3 Sample Characteristics</p> <p>Inclusion criteria [0 = not stated, 1 = no exclusions, 2 = stated explicitly, 3 = partly stated] <input type="checkbox"/></p> <p>Stage of breast cancer included in the study: lowest: <input type="checkbox"/> highest: <input type="checkbox"/></p> <p>If the highest stage included is greater than Stage 2, please comment on the operability of disease within this patient grouping. _____</p> <p>Menopausal status? [1 = pre, 2 = post, 3 = both] <input type="checkbox"/></p>									
	<p>4 Treatment</p> <p>What proportion of patients received surgery as treatment for their disease: _____</p> <p>Type of treatment received by patients included in the paper? Did some patients receive [<i>tick all that apply</i>]:</p> <p>Chemotherapy <input type="checkbox"/> Hormone therapy <input type="checkbox"/> Radiotherapy <input type="checkbox"/> Immunotherapy <input type="checkbox"/></p> <p>Not stated <input type="checkbox"/></p> <p>Please indicate timing of adjuvant therapy in relation to any surgery received _____</p> <p>Was treatment explicitly determined by the prognostic model? <input type="checkbox"/> [0 = no, 1 = yes, 2 = not stated]</p> <p>Treatment received, was it: 1 = standardised, 2 = randomised, 3 = unit protocol, 4 = not stated, 5 = other: specify <input type="checkbox"/> <input type="checkbox"/></p>									
	<p>5 End-points for outcome analysis</p> <p>Number of outcome measures used in univariate analysis <input type="checkbox"/><input type="checkbox"/></p> <p>Number of outcome measures used in multivariate analyses [0 = none] <input type="checkbox"/><input type="checkbox"/></p> <p>Death [0 = no, 1 = explicitly any death, 2 = yes – unclear, 3 = ‘overall survival’] <input type="checkbox"/></p> <p>Cancer death [0 = no, 1 = yes] <input type="checkbox"/></p>									

<p>Recurrence [breast cancer] [0 = no, 1 = yes, 2 = yes but unclear how deaths treated] <input type="checkbox"/></p> <p>Any event [death or relapse [0 = no, 1 = yes] <input type="checkbox"/></p> <p>Is time origin stated? [0 = no, 1 = yes] specify: _____ <input type="checkbox"/></p> <p>Number of events: Number of deaths [entire sample] <input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/> Number of recurrences <input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/></p>
<p>6 Follow-up</p> <p>Summary of follow-up given? [0 = no, 1 = median only, 2 = median and range, 3 = other] <input type="checkbox"/></p> <p>Actual length of follow-up [median months] <input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/></p> <p>Clear statement on subject loss to follow-up? [0 = no, 1 = yes, 2 = yes: none lost] <input type="checkbox"/></p> <p>Clear statement on how loss to follow-up treated in analysis? [0 = no, 1 = yes] <input type="checkbox"/></p>
<p>7 Data Quality</p> <p>Discussion in text of missing data? [0 = no, 1 = yes, 2 = no: none missing, 3 = yes: none missing] <input type="checkbox"/></p> <p>Number of patients excluded due to missing data. <input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/></p>
<p>8 Prognostic factors used in developing the model</p> <p>Number of factors considered in the study <input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/></p> <p>Was information provided as to why these factors were included? [0 = no, 1 = yes] <input type="checkbox"/></p> <p>If yes, please specify: _____</p>
<p>9 Univariate Analysis</p> <p>Were study results presented in a tabulated form? [0 = no, 1 = yes] <input type="checkbox"/></p> <p>Actuarial survival rates presented per factor? [0 = no, 1 = yes: all factors, 2 = yes: some factors, 3 = yes: only if test is significant, 4 = occasional subgroups] <input type="checkbox"/></p> <p>Confidence intervals presented? [0 = no, 1 = for median, 2 = for proportion, 3 = other] <input type="checkbox"/></p> <p>Hazard ratio [0 = no, 1 = O/E, 2 = ratio of median survival times, 3 = e from univ. Cox, 4 = other] <input type="checkbox"/></p>
<p>10 Multivariate Analysis</p> <p>Multivariate analyses done? [0 = no, 1 = yes] <input type="checkbox"/></p> <p>Model used [0 = none, 1 = Cox, 2 = Weibull, 3 = Artificial Neural Network, 4 = other] <input type="checkbox"/></p> <p>Type of Cox model [1 = time fixed, 2 = time dependent, 3 = both] <input type="checkbox"/></p> <p>Stratified Cox model used? [0 = no, 1 = yes] <input type="checkbox"/></p>

Strategy for building the multivariate model [1 = stepwise, 2 = forward stepwise, 3 = backward stepwise, 4 = all subsets, 5 = all significant in univariate, 6 = other] <i>If other, please specify:</i> _____	<input type="checkbox"/>
Choice of variables to include [1 = all used in univariate analyses, 2 = all with $p < 0.05$ in univariate, 3 = all with $p <$ some higher level in univariate, 4 = unjustified list, 5 = all available, 6 = other] <i>If other, please specify:</i> _____	<input type="checkbox"/>
Forced inclusion of variables in full model known a priori to affect survival? [0 = no, 1 = yes]	<input type="checkbox"/>
Any interaction(s) examined? [0 = no, 1 = yes] Specify: _____	<input type="checkbox"/>
Model assumptions discussed? [0 = no, 1 = yes]	<input type="checkbox"/>
Model assumptions assessed? [0 = no, 1 = yes]	<input type="checkbox"/>
How goodness of fit was assessed [0 = not assessed, 1 = residual plots (not necessarily shown), 2 = DBETAS (deltaBeta), 3 = other]	<input type="checkbox"/>
Number of factors in final model	<input type="checkbox"/> <input type="checkbox"/>
11 Presentation of multivariate model	
Regression coefficients [0 = no, 1 = some, 2 = all]	<input type="checkbox"/>
Relative risks given [0 = no, 1 = some, 2 = all]	<input type="checkbox"/>
SE or CI given (for beta or RR)? [0 = no, 1 = some, 2 = all]	<input type="checkbox"/>
p values from final model? [0 = no, 1 = some, 2 = all, 3 = significance but no p values]	<input type="checkbox"/>
Calculation of prognostic index [0 = no, 1 = yes but not explained, 2 = yes, explained]	<input type="checkbox"/>
Have risk groups been created? [0 = no, 1 = yes]	<input type="checkbox"/>
If yes, how many risk groups were created?	<input type="checkbox"/>
What methods were used to create these risk groups? _____	
Expected survival as a function of the PI [0 = no, 1 = yes: plot, 2 = yes: some values given]	<input type="checkbox"/>
Other graphs based on survival model created? [0 = no, 1 = yes] If yes, please give details: _____	<input type="checkbox"/>
12 Treatment of continuous prognostic variables	
How treated? [0 = no continuous variables, 1 = all continuous, 2 = all categorised, 3 = some of each]	<input type="checkbox"/>
<i>For those variables categorised: Logrank:</i> How categorised? [1 = all dichotomised, 2 = all 3+ groups, 3 = mixture]	<input type="checkbox"/>

Choice of cutpoints [1 = quantiles, 2 = reason given, 3 = no reason given, 4 = mixture, 5 = data dependent]	<input type="checkbox"/>
<i>For those variables categorised: (semi) parametric</i> (Including univariate or multivariate Cox): How categorised? [1 = all dichotomised, 2 = all 3+ groups, 3 = mixture]	<input type="checkbox"/>
Choice of cutpoints [1 = quantiles, 2 = reason given, 3 = no reason given, 4 = mixture, 5 = data dependent]	<input type="checkbox"/>
Comments	
Has this model been compared to other published models? [0 = no, 1 = yes, 2 = not clear]	<input type="checkbox"/>
Is this model useable by others [researchers/clinicians]? [0 = no, 1 = yes, 2 = not clear]	<input type="checkbox"/>
<hr/> Remember to write summary paragraph <hr/>	
<i>When writing study summary paragraph, please comment on:</i>	
Study design	
Choice of variables	
Data quality	
Methods of analysis	
Presentation of paper	
Usefulness of model [can it be used in clinical practice/by other researchers]	
Other: please specify	
<hr/> Validation	
Type of validation	
Internal [e.g. bootstrap] <input type="checkbox"/>	Random split <input type="checkbox"/>
Temporal <input type="checkbox"/>	External <input type="checkbox"/>
Larger series including original sample <input type="checkbox"/>	
By whom is the validation done [1 = same investigator, 2 = different investigators]	
<input type="checkbox"/>	
Were all statistical methods clearly identified? [0 = no, 1 = yes]	
<input type="checkbox"/>	

Topic B. Studies of predictive factors and response to treatment

Only include papers that **fulfil** the following inclusion criteria:

	Yes	No
• primary cancer;	<input type="checkbox"/>	<input type="checkbox"/>
• non-metastatic and non-advanced disease [If includes patients with metastatic disease, this must Not be included in analysis for prognostic model]	<input type="checkbox"/>	<input type="checkbox"/>
• operable b.c. where patients received surgery as initial treatment for their disease	<input type="checkbox"/>	<input type="checkbox"/>
• patient's receiving adjuvant therapy only	<input type="checkbox"/>	<input type="checkbox"/>
• comparison of subgroups of patients, divided by predictive factors	<input type="checkbox"/>	<input type="checkbox"/>

Paper Number: _____ Reviewer Initials: _____ Year of Publication: _____

Author of Paper: _____ Paper number of other papers that this may link: _____

Name of predictive factor [patient/tumour] studied within this paper _____

Pg. No **1 Study Dates**

	Are all 3 study dates specified [see below*]? [0 = no, 1 = yes, 2 = partially]	<input type="checkbox"/>
	Start of recruitment*	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
	End of recruitment*	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
	End of follow-up*	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
2 Design		
	Study design [1 = randomised clinical trial, 2 = non-randomised comparative study, 3 = cohort study (all patients treated same way), 4 = other]	<input type="checkbox"/>
	Design [1 = prospective, 2 = retrospective]	<input type="checkbox"/>
	From where was the sample recruited? [0 = not known, 1 = one centre, 2 = >1centre, 3 =other] [<i>please specify other</i>] _____	<input type="checkbox"/>
3 Sample Characteristics		
	Inclusion criteria [0 = not stated, 1 = no exclusions, 2 = stated explicitly, 3 =partly stated]	<input type="checkbox"/>
	Age range of study group [enter 000 if not known] _____	
	Stage of breast cancer included in the study: lowest: <input type="text"/> highest: <input type="text"/>	

	<p>If the highest stage is greater than Stage 2, comment on the operability of disease within this patient grouping. _____</p> <p>Menopausal status? [1 = pre, 2 = post, 3 = both] <input type="checkbox"/></p> <p>Any other sample characteristics that were of particular relevance to this study: please comment _____</p>
	<p>4 Treatment</p> <p>Please specify the target treatment received by patients in the study. _____</p> <p>Was the target treatment influenced by the presence of the target predictive factor? [0 = no, 1 = yes] <input type="checkbox"/></p> <p>Treatment received, was it: 1 = standardised [all patients received the same], <input type="checkbox"/> <input type="checkbox"/> 2 = randomised, 3 = unit protocol, 4 = not stated, 5 = other: specify _____</p> <p>Did patients receive any other treatment? [0 = no, 1 = yes], <input type="checkbox"/> If yes, please specify: [tick all that apply]</p> <p>Chemotherapy <input type="checkbox"/> specify: _____ Hormone therapy <input type="checkbox"/> specify: _____ Immunotherapy <input type="checkbox"/> specify: _____ Radiotherapy <input type="checkbox"/></p> <p>For each factor and each treatment included in the study: please complete a separate DEF</p>
	<p>5 Predictive factors</p> <p>Total number of factors investigated within the study <input type="text"/> <input type="text"/></p> <p>Name of predictive factor: _____</p> <p>Number of subjects used in analysis of predictive factor and response to treatment <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/></p> <p>Were details provided of the assay techniques/laboratory methodology employed for the predictive factor? [0 = no, 1 = yes, 2 = not relevant to the predictive factor, 3 = not stated] <input type="checkbox"/></p> <p>Were details provided of the methodology used for storage of this tissue sample? [0 = no, 1 = yes, 2 = not relevant to this predictive factor, 3 = not stated] <input type="checkbox"/></p>
	<p>6 End-points for outcome analysis</p> <p>Number of outcome measures used in univariate analysis <input type="text"/> <input type="text"/></p> <p>Number of outcome measures used in multivariate analyses [0 = none] <input type="text"/> <input type="text"/></p> <p>Death [0 = no, 1 = explicitly any death, 2 = yes – unclear, 3 = ‘overall survival’] <input type="checkbox"/></p> <p>Cancer death [0 = no, 1 = yes] <input type="checkbox"/></p> <p>Recurrence/relapse [breast cancer] [0 = no, 1 = yes, 2 = yes but unclear how deaths treated] <input type="checkbox"/></p> <p>Any event [death or relapse] [0 = no, 1 = yes] <input type="checkbox"/></p>

	Is time origin stated? [0 = no, 1 = yes] specify: _____ <input type="checkbox"/>
	Number of events: Number of deaths [entire sample] <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Number of recurrences <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
7 Follow-up	
	Summary of follow-up given? [0 = no, 1 = median only, 2 = median and range, 3 = other] <input type="checkbox"/>
	Actual length of follow-up [median months] <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
	Clear statement on loss to follow-up? [0 = no, 1 = yes, 2 = yes: none lost] <input type="checkbox"/>
	Clear statement on how loss to follow-up treated in analysis? [0 = no, 1 = yes] <input type="checkbox"/>
8 Data Quality	
	Discussion in text of missing data? [0 = no, 1 = yes, 2 = no: none missing, 3 = yes: none missing] <input type="checkbox"/>
	Number of patients excluded from analysis, due to missing data <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
9 Analysis	
	For uncontrolled studies: do they give survival curves for groups defined by the predictive factor? [0 = no, 1 = yes] <input type="checkbox"/>
	For controlled studies: do they compare treatment effects between the groups by the predictive factor? [0 = no, 1 = interaction tests, 2 = comparison of <i>p</i> values, 3 = other] <input type="checkbox"/>
	Confidence intervals presented? [0 = no, 1 = yes] <input type="checkbox"/>
	Hazard ratio [0 = no, 1 = yes] <input type="checkbox"/>
	Did the authors perform a multivariate analysis, adjusting for other predictive variables? [0 = no, 1 = yes] <input type="checkbox"/>
10 Results of Primary Study	
	Do the authors conclude that the factor is predictive of outcome? [0 = no, 1 = yes] <input type="checkbox"/>
	Provide summary details of the findings of this study.
	Do you think these conclusions are reasonable given the results of the study? Please comment.

	<p>11 Comments</p> <p><i>Please note below any key points from this study, using the suggested headings where applicable.</i></p> <p>Study design</p> <p>Data quality</p> <p>Methods of analysis</p> <p>Presentation of paper</p> <p>Other: please specify</p>
	<p>Include paper in HTA review? [0 = no, 1 = yes, 2 = need to discuss with group] <input type="checkbox"/></p>
	<p>Time taken to complete this form [minutes] <input type="text"/><input type="text"/><input type="text"/></p>

Date form completed: _____

	<p>Stage[s] of disease [0 = not stated, 1 = partly stated, 2 = stated for all included studies] <input type="checkbox"/></p> <p>Stages of breast cancer considered within the studies lowest: <input type="text"/> highest: <input type="text"/></p> <p>Did the study participants receive adjuvant treatment? [0 = no, 1 = yes, 2 = not stated] <input type="checkbox"/></p> <p>Please comment on the extent of/lack of clinical heterogeneity within the studies included in the review _____ _____</p>
	<p>4 Prognostic Factors</p> <p>Name of prognostic factor: _____</p> <p>Were details provided of the assay techniques/laboratory methodology that was used to measure this prognostic factor in the included studies? [0 = no, 1 = yes, 2 = not relevant to the prognostic factor, 3 = not stated] <input type="checkbox"/></p> <p>Was the prognostic factor measured/analysed in the same way by each of the included studies? [0 = no, 1 = yes, 2 = not relevant to the prognostic factor, 3 = not stated] <input type="checkbox"/></p> <p>Were the same conditions for storage of the prognostic factor utilised by each of the included studies? [0 = no, 1 = yes, 2 = not relevant to this prognostic factor, 3 = not stated] <input type="checkbox"/></p>
	<p>5 End-points for outcome analysis</p> <p>Which of the following end-points were considered in the review: <input type="checkbox"/></p> <p>Death [0 = no, 1 = explicitly any death, 2 = yes – unclear, 3 = ‘overall survival’]</p> <p>Cancer death [0 = no, 1 = yes] <input type="checkbox"/></p> <p>Recurrence/relapse [breast cancer] [0 = no, 1 = yes, 2 = yes but unclear how deaths treated] <input type="checkbox"/></p> <p>Any event [death or relapse [0 = no, 1 = yes] <input type="checkbox"/></p> <p>Other [0 = no, 1 = yes] If yes, please specify: <input type="checkbox"/></p> <p>_____</p> <p>Did each of the included studies utilise the same end-points for outcome analysis? [0 = no, 1 = yes, 2 = not possible to ascertain] <input type="checkbox"/></p>
	<p>6 Data Quality</p> <p>Did the reviewers assess the methodological quality of the included study papers? [0 = no, 1 = yes, 2 = not stated] <input type="checkbox"/></p> <p>If yes, how was this undertaken? _____ _____</p> <p>Discussion in text of loss to follow-up in the included studies? [0 = no, 1 = yes, 2 = no: none missing, 3 = yes: none missing] <input type="checkbox"/></p>

7 Results of Individual Studies	
Number of studies included within the review	<input type="text"/>
Method by which the results from the included studies are presented. [0 = not presented, 1 = narrative summary of results, 2 = tabulated summary of some of part of the results, 3 = other]	<input type="checkbox"/>
Hazard ratio given [0 = no, 1 = yes]	<input type="checkbox"/>
<i>p</i> value given? [0 = no, 1 = yes]	<input type="checkbox"/>
Are cut-points specified for each study? [0 = no, 1 = yes]	<input type="checkbox"/>
Is the prevalence of high/elevated or low/reduced levels specified for each study? [0 = no, 1 = yes]	<input type="checkbox"/>
8 Meta-Analysis	
Was statistical heterogeneity assessed? [0 = no, 1 = yes]	<input type="checkbox"/>
If yes , please provide details of how statistical heterogeneity was assessed.	 <hr/> <hr/>
Was a meta-analysis undertaken? [0 = no, 1 = yes]	<input type="checkbox"/>
If no , what method of data synthesis was used for the review? [provide details]	<hr/>
If yes , what method was used to synthesise data from included studies? [please specify]	<hr/>
Which outcomes were considered in meta-analysis?	
Death [0 = no, 1 = yes]	<input type="checkbox"/>
Recurrence/relapse [0 = no, 1 = yes]	<input type="checkbox"/>
Disease free survival [0 = no, 1 = yes]	<input type="checkbox"/>
Were summary results presented as <i>p</i> values? [0 = no, 1 = yes]	<input type="checkbox"/>
Hazard ratio given? [0 = no, 1 = yes]	<input type="checkbox"/>
Relative risk given? [0 = no, 1 = yes]	<input type="checkbox"/>
Confidence intervals presented? [0 = no, 1 = yes]	<input type="checkbox"/>
Assessment of publication biases? [0 = no, 1 = yes, 2 = not stated]	<input type="checkbox"/>
Any sensitivity analysis [eg omitting poorer studies]? [0 = no, 1 = yes]	<input type="checkbox"/>

	<p>9 Results of Review</p> <p>Provide summary details of the findings of this review.</p> <p>Do you think that these conclusions are valid given the results of the review? Please comment.</p>
	<p>10 Comments</p> <p>Does the review add any new data/information to the existing knowledge about prognostic factors in the treatment of breast cancer? [0 = no, 1 = yes, 2 = not stated] <input type="checkbox"/></p>
	<p><i>Please note below any key points from this study, using where applicable the suggested headings.</i></p> <p>Study design</p> <p>Data quality</p> <p>Methods of analysis</p> <p>Presentation of paper</p> <p>Other [please specify]</p>
	<p>Include paper in HTA review [0 = no, 1 = yes, 2 = needs discussion with group] <input type="checkbox"/></p>
	<p>Time taken to complete this form [minutes] <input type="text"/> <input type="text"/> <input type="text"/></p>

Date form completed: _____

Appendix 4

Questionnaires used to survey clinicians and histopathologists for Topic E

Clinicians' survey letter

Dr

Address

.....

.....

.....

Dear Dr

Survey of current use of prognostic and predictive factors in the United Kingdom to select women for adjuvant therapy of breast cancer.

There has been rapid development in adjuvant therapy in the management of early breast cancer in the last decade. However there is little information available on how prognostic and predictive factors affect outcome and how they are used within clinical practice.

A group of clinicians and methodologists working alongside the Cochrane Cancer Network have been commissioned by the NHS Health Technology Assessment programme to address this issue.

One of the missing pieces of information is how prognostic and predictive factors are used in clinical practice within the UK. In order to attain this important information, we would be grateful if you would pass this survey on to the lead breast cancer clinician in your hospital for them to complete this survey. I know how much pressure everyone is under, but your input is crucial to addressing this issue.

If you have any enquiries about this survey or the overall project, please contact Susan Brunskill, Research Fellow, by telephone on 01865 226645, or email at sbrunskill@canet.org

Thank you for your response,

Yours sincerely

Dr C Williams
Lead for HTA Project.
Cochrane Cancer Network.
Institute of Health Sciences
Headington
PO Box 777
Oxford OX3 7LF

Clinicians' questionnaire

Survey of Current Use of Prognostic and Predictive Factors in the United Kingdom, to Select Women for Adjuvant Systemic Therapy of Breast Cancer

To be completed by the Lead Breast Cancer Clinician managing adjuvant therapy.

This survey forms part of a NHS Health Technology Assessment funded project, concerning the use of prognostic and predictive factors, to select women for adjuvant chemotherapy of breast cancer. The project is being co-ordinated and run by a group of clinicians and methodologists working alongside the Cochrane Cancer Network.

One of the missing pieces of information is how prognostic and predictive factors are used in clinical practice within the UK. In order to attain this important information, we would be grateful if you would complete this survey. We know how much pressure everyone is under, but your input is crucial to addressing this issue.

For enquiries about this survey, please email Susan Brunskill, [Project Research Fellow] at sbrunskill@canet.org

Please respond to as many questions as you are able to.

Please return by

Form completed by

Date

Please provide contact details:

.....

Please answer these survey questions in accordance with current practice in your unit.

Please respond to questions 1 and 2 using the table below.

Please respond by placing a tick in the box[es] in the table below, next to those factor[s] that you consider to be important. *Please respond separately for patients aged 50 years and under, and for patients aged over 50 years.*

1. Which of the noted factors do you consider to be clinically important when selecting newly diagnosed, regionally localised breast cancer patients for adjuvant chemotherapy?
2. Which of the noted factors do you consider to be clinically important when selecting newly diagnosed, regionally localised breast cancer patients for adjuvant hormone therapy?

Factor	Question			
	1 [Chemotherapy]		2 [Hormone Ther.]	
	≤ 50 yrs.	> 50 yrs.	≤ 50 yrs.	> 50 yrs.
Age				
Menopausal status				
Physiological Age				
Grade				
Histological subtype				
Margins positive with invasive cancer				
Nodal Status [pathological]				
Nodal Status [clinical]				
Prognostic index [e.g. Nottingham]				
Size				
Vascular invasion				
BCL 2				
Estrogen receptor				
HER2 erbB2				
Progesterone receptor				
Proliferation index				
<i>Other factors available</i>				
<i>.....</i>				

3. In your unit, when deciding on adjuvant therapy for newly diagnosed, regionally localised breast cancer patients, which factors are generally available? [Please place a **tick** in the box[es] next to those factors that are available].

Factor	Factor
Age	Size
Menopausal status	Vascular invasion
Physiological Age	BCL 2
Grade	Estrogen receptor
Histological subtype	HER2 erbB2
Margins positive with invasive cancer	Progesterone receptor
Nodal Status [pathological]	Proliferation index
Nodal Status [clinical]	<i>Other factors available</i>
Prognostic index [e.g. Nottingham]

4. In your unit, are there any other factors that you use when deciding on therapy for selected patients?

Yes No

If **yes**, what are they?:

.....

5. Are there factors that are not currently available in your unit, which you would like to have available when making treatment decisions for newly diagnosed, regionally localised breast cancer patients?

Yes No

If **yes**, what are they?

.....

6. In your unit, are there a minimal number of nodes sampled that is considered acceptable for staging the axilla?

Yes No

If **yes**, what number

7. a) In your unit, is there a written protocol for making decisions on adjuvant treatment for newly diagnosed, regionally localised breast cancer patients?

Yes No

b) If **yes**, is it based on a prognostic index?

Yes No

c) If **yes**, which prognostic index is it based on?

.....

d) If you have a written unit protocol, would you be prepared to send us a copy of the protocol when you return this survey?

Yes No

Protocol enclosed: Yes No

8. a) How many patients are seen in your unit each year with newly diagnosed, regionally localised breast cancer? [Please place a tick in the relevant box].

0-20 20-50 50-100
 100-150 150+

b) Approximately what proportion [%] of these patients received:

- i] adjuvant chemotherapy?.....
- ii] adjuvant hormonal therapy?.....
- iii] both adjuvant chemotherapy and adjuvant hormonal therapy?.....

9. In your unit, do you use a conventional chemotherapy regime [e.g. CMF, CAF] for treatment of newly diagnosed, regionally localised breast cancer?

Yes No

If **yes**, what is the regime?.....

10. a) In your unit, do you use a non-conventional adjuvant chemotherapy regime[s], which includes new cytotoxic reagents for treatment of all or some patients with newly diagnosed, regionally localised breast cancer?

Yes No

If **yes**, what are the regime[s]?

b) Approximately what proportion [%] of patients who are eligible for adjuvant chemotherapy received the non-conventional regime[s] in the last year?

.....

11. Would you be willing to take part in a short survey about quality of life and breast cancer?

Yes No

Thank you for taking the time to complete this survey.

Please return this survey in the enclosed S.A.E.

Dr C Williams
 Lead for HTA Project.
 Cochrane Cancer Network.
 Institute of Health Sciences
 PO Box 777
 Headington
 Oxford OX3 7LF

Histopathologists' survey letter

«Job_Title_»
«Department»
«Address1»
«Address2»
«Address3»
«Address4»
«Address5»

Dear «Salutation»

Survey of current use of prognostic and predictive factors in the United Kingdom to select women for adjuvant therapy of breast cancer.

There has been rapid development in adjuvant therapy in the management of early breast cancer in the last decade. However, there is little information available on how prognostic and predictive factors affect outcome and how they are used within clinical practice.

A group of clinicians and methodologists working alongside the Cochrane Cancer Network have been commissioned by the NHS Health Technology Assessment [HTA] programme to address this issue.

One of the missing pieces of information is the availability and the use of prognostic and predictive factors in clinical practice within the UK. In order to obtain this important information, we would be grateful if you would take the time to complete this survey. All we are interested in, is whether assessments for particular predictive and prognostic factors are available to a patient's clinician.

I know how much pressure everyone is under, but your input is crucial to addressing this issue. The survey will take a few minutes to complete. If you have any enquiries about this survey or the overall project, please contact Susan Brunskill, Research Fellow, by telephone on 01865 226645, or email at sbrunskill@canet.org

Thank you in advance for your response,

Yours sincerely

Dr C Williams
Lead for HTA Project.
Cochrane Cancer Network.
Institute of Health Sciences
PO Box 777
Headington
Oxford OX3 7LF

Histopathologists' questionnaire

Survey of Current Use of Prognostic and Predictive Factors in the United Kingdom, to Select Women for Adjuvant Systemic Therapy of Breast Cancer

To be completed by the Lead Histopathologist managing breast cancer diagnosis.

Form completed by

Please provide contact details:

.....

*For further details about the survey, please email
Susan Brunskill [Research Fellow] at sbrunskill@canet.org*

Please answer these survey questions in accordance with current practice in your unit.

1. Which factors are available on a pathology report to oncology breast clinicians making decisions about **newly diagnosed, early stage breast cancer** patients? [Please place a **tick** in the box next to those factors that are available].

Please indicate your responses to the question below in the following table.

	Question
Factor	1
Histological subtype	
Grade	
Size	
Margins positive with invasive cancer	
Vascular invasion	
Nodal Status [pathological]	
BCL 2	
HER2 erbB2	
Estrogen receptor	
Progesterone receptor	
Proliferation index	
<i>Other factors available</i>	
<i>.....</i>	
<i>.....</i>	

Thank you for taking the time to complete this survey.

Please return this survey in the enclosed S.A.E.

Dr C Williams
 Director
 Cochrane Cancer Network.
 Institute of Health Sciences
 PO Box 777
 Headington
 Oxford OX3 7LF



Health Technology Assessment Programme

Director,
Professor Tom Walley,
Director, NHS HTA Programme,
Department of Pharmacology &
Therapeutics,
University of Liverpool

Deputy Director,
Professor Jon Nicholl,
Director, Medical Care Research
Unit, University of Sheffield,
School of Health and Related
Research

Prioritisation Strategy Group

Members

Chair,
Professor Tom Walley,
Director, NHS HTA Programme,
Department of Pharmacology &
Therapeutics,
University of Liverpool

Professor Bruce Campbell,
Consultant Vascular & General
Surgeon, Royal Devon & Exeter
Hospital

Dr Edmund Jessop, Medical
Advisor, National Specialist,
Commissioning Advisory Group
(NSCAG), Department of
Health, London

Professor Jon Nicholl, Director,
Medical Care Research Unit,
University of Sheffield, School
of Health and Related Research

Dr John Reynolds, Clinical
Director, Acute General
Medicine SDU, Radcliffe
Hospital, Oxford

Dr Ron Zimmern, Director,
Public Health Genetics Unit,
Strangeways Research
Laboratories, Cambridge

HTA Commissioning Board

Members

Programme Director,
Professor Tom Walley,
Director, NHS HTA Programme,
Department of Pharmacology &
Therapeutics,
University of Liverpool

Chair,
Professor Jon Nicholl,
Director, Medical Care Research
Unit, University of Sheffield,
School of Health and Related
Research

Deputy Chair,
Professor Jenny Hewison,
Professor of Health Care
Psychology, Academic Unit of
Psychiatry and Behavioural
Sciences, University of Leeds
School of Medicine

Dr Jeffrey Aronson
Reader in Clinical
Pharmacology, Department of
Clinical Pharmacology,
Radcliffe Infirmary, Oxford

Professor Deborah Ashby,
Professor of Medical Statistics,
Department of Environmental
and Preventative Medicine,
Queen Mary University of
London

Professor Ann Bowling,
Professor of Health Services
Research, Primary Care and
Population Studies,
University College London

Dr Andrew Briggs, Public
Health Career Scientist, Health
Economics Research Centre,
University of Oxford

Professor John Cairns, Professor
of Health Economics, Public
Health Policy, London School of
Hygiene and Tropical Medicine,
London

Professor Nicky Cullum,
Director of Centre for Evidence
Based Nursing, Department of
Health Sciences, University of
York

Mr Jonathan Deeks,
Senior Medical Statistician,
Centre for Statistics in
Medicine, University of Oxford

Dr Andrew Farmer, Senior
Lecturer in General Practice,
Department of Primary
Health Care,
University of Oxford

Professor Fiona J Gilbert,
Professor of Radiology,
Department of Radiology,
University of Aberdeen

Professor Adrian Grant,
Director, Health Services
Research Unit, University of
Aberdeen

Professor F D Richard Hobbs,
Professor of Primary Care &
General Practice, Department of
Primary Care & General
Practice, University of
Birmingham

Professor Peter Jones, Head of
Department, University
Department of Psychiatry,
University of Cambridge

Professor Sallie Lamb,
Professor of Rehabilitation,
Centre for Primary Health Care,
University of Warwick

Professor Stuart Logan,
Director of Health & Social
Care Research, The
Peninsula Medical School,
Universities of Exeter &
Plymouth

Dr Linda Patterson,
Consultant Physician,
Department of Medicine,
Burnley General Hospital

Professor Ian Roberts, Professor
of Epidemiology & Public
Health, Intervention Research
Unit, London School of
Hygiene and Tropical Medicine

Professor Mark Sculpher,
Professor of Health Economics,
Centre for Health Economics,
Institute for Research in the
Social Services, University of York

Dr Jonathan Shapiro, Senior
Fellow, Health Services
Management Centre,
Birmingham

Ms Kate Thomas,
Deputy Director,
Medical Care Research Unit,
University of Sheffield

Ms Sue Ziebland,
Research Director, DIPEX,
Department of Primary Health
Care, University of Oxford,
Institute of Health Sciences

Current and past membership details of all HTA 'committees' are available from the HTA website (www.hta.ac.uk)

Diagnostic Technologies & Screening Panel

Members

<p>Chair, Dr Ron Zimmern, Director of the Public Health Genetics Unit, Strangeways Research Laboratories, Cambridge</p>	<p>Professor Adrian K Dixon, Professor of Radiology, University Department of Radiology, University of Cambridge Clinical School</p>	<p>Dr Susanne M Ludgate, Medical Director, Medicines & Healthcare Products Regulatory Agency, London</p>	<p>Professor Lindsay Wilson Turnbull, Scientific Director, Centre for MR Investigations & YCR Professor of Radiology, University of Hull</p>
<p>Ms Norma Armston, Lay Member, Bolton</p>	<p>Dr David Elliman, Consultant Paediatrician/Hon. Senior Lecturer, Population Health Unit, Great Ormond St. Hospital, London</p>	<p>Professor William Rosenberg, Professor of Hepatology, Liver Research Group, University of Southampton</p>	<p>Professor Martin J Whittle, Associate Dean for Education, Head of Department of Obstetrics and Gynaecology, University of Birmingham</p>
<p>Professor Max Bachmann Professor of Health Care Interfaces, Department of Health Policy and Practice, University of East Anglia</p>	<p>Professor Glyn Elwyn, Primary Medical Care Research Group, Swansea Clinical School, University of Wales Swansea</p>	<p>Dr Susan Schonfield, Consultant in Public Health, Specialised Services Commissioning North West London, Hillingdon Primary Care Trust</p>	<p>Dr Dennis Wright, Consultant Biochemist & Clinical Director, Pathology & The Kennedy Galton Centre, Northwick Park & St Mark's Hospitals, Harrow</p>
<p>Professor Rudy Bilous Professor of Clinical Medicine & Consultant Physician, The Academic Centre, South Tees Hospitals NHS Trust</p>	<p>Mr Tam Fry, Honorary Chairman, Child Growth Foundation, London</p>	<p>Dr Phil Shackley, Senior Lecturer in Health Economics, School of Population and Health Sciences, University of Newcastle upon Tyne</p>	
<p>Dr Paul Cockcroft, Consultant Medical Microbiologist and Clinical Director of Pathology, Department of Clinical Microbiology, St Mary's Hospital, Portsmouth</p>	<p>Dr Jennifer J Kurinczuk, Consultant Clinical Epidemiologist, National Perinatal Epidemiology Unit, Oxford</p>	<p>Dr Margaret Somerville, PMS Public Health Lead, Peninsula Medical School, University of Plymouth</p>	
		<p>Dr Graham Taylor, Scientific Director & Senior Lecturer, Regional DNA Laboratory, The Leeds Teaching Hospitals</p>	

Pharmaceuticals Panel

Members

<p>Chair, Dr John Reynolds, Chair Division A, The John Radcliffe Hospital, Oxford Radcliffe Hospitals NHS Trust</p>	<p>Mr Peter Cardy, Chief Executive, Macmillan Cancer Relief, London</p>	<p>Dr Christine Hine, Consultant in Public Health Medicine, South Gloucestershire Primary Care Trust</p>	<p>Professor Jan Scott, Professor of Psychological Treatments, Institute of Psychiatry, University of London</p>
<p>Professor Tony Avery, Head of Division of Primary Care, School of Community Health Services, Division of General Practice, University of Nottingham</p>	<p>Professor Imti Choonara, Professor in Child Health, Academic Division of Child Health, University of Nottingham</p>	<p>Professor Stan Kaye, Cancer Research UK Professor of Medical Oncology, Section of Medicine, The Royal Marsden Hospital, Sutton</p>	<p>Mrs Katrina Simister, Assistant Director New Medicines, National Prescribing Centre, Liverpool</p>
<p>Ms Anne Baileiff, Consultant Nurse in First Contact Care, Southampton City Primary Care Trust, University of Southampton</p>	<p>Dr Robin Ferner, Consultant Physician and Director, West Midlands Centre for Adverse Drug Reactions, City Hospital NHS Trust, Birmingham</p>	<p>Ms Barbara Meredith, Lay Member, Epsom</p>	<p>Dr Richard Tiner, Medical Director, Medical Department, Association of the British Pharmaceutical Industry, London</p>
<p>Professor Stirling Bryan, Professor of Health Economics, Health Services Management Centre, University of Birmingham</p>	<p>Dr Karen A Fitzgerald, Consultant in Pharmaceutical Public Health, National Public Health Service for Wales, Cardiff</p>	<p>Dr Andrew Prentice, Senior Lecturer and Consultant Obstetrician & Gynaecologist, Department of Obstetrics & Gynaecology, University of Cambridge</p>	<p>Dr Helen Williams, Consultant Microbiologist, Norfolk & Norwich University Hospital NHS Trust</p>
	<p>Mrs Sharon Hart, Head of DTB Publications, <i>Drug & Therapeutics Bulletin</i>, London</p>	<p>Dr Frances Rotblat, CPMP Delegate, Medicines & Healthcare Products Regulatory Agency, London</p>	

Therapeutic Procedures Panel

Members

Chair,

Professor Bruce Campbell,
Consultant Vascular and
General Surgeon, Department
of Surgery, Royal Devon &
Exeter Hospital

Dr Carl E Counsell, Clinical
Senior Lecturer in Neurology,
Department of Medicine and
Therapeutics, University of
Aberdeen

Ms Maryann L Hardy,
Lecturer, Division of
Radiography, University of
Bradford

Professor James Neilson,
Professor of Obstetrics and
Gynaecology, Department of
Obstetrics and Gynaecology,
University of Liverpool

Ms Amelia Curwen, Executive
Director of Policy, Services and
Research, Asthma UK, London

Professor Alan Horwich,
Director of Clinical R&D,
Academic Department of
Radiology, The Institute of
Cancer Research,
London

Dr John C Pounsford,
Consultant Physician,
Directorate of Medical Services,
North Bristol NHS Trust

Professor Gene Feder, Professor
of Primary Care R&D,
Department of General Practice
and Primary Care, Barts & the
London, Queen Mary's School
of Medicine and Dentistry,
London

Dr Simon de Lusignan,
Senior Lecturer,
Primary Care Informatics,
Department of Community
Health Sciences,
St George's Hospital Medical
School, London

Karen Roberts, Nurse
Consultant, Queen Elizabeth
Hospital, Gateshead

Dr Aileen Clarke,
Reader in Health Services
Research, Public Health &
Policy Research Unit, Barts &
the London School of Medicine
& Dentistry, London

Professor Paul Gregg,
Professor of Orthopaedic
Surgical Science, Department of
General Practice and Primary
Care, South Tees Hospital NHS
Trust, Middlesbrough

Professor Neil McIntosh,
Edward Clark Professor of
Child Life & Health,
Department of Child Life &
Health, University of
Edinburgh

Dr Vimal Sharma, Consultant
Psychiatrist/Hon. Senior Lecturer,
Mental Health Resource Centre,
Cheshire and Wirral Partnership
NHS Trust, Wallasey

Dr L David Smith, Consultant
Cardiologist, Royal Devon &
Exeter Hospital

Dr Matthew Cooke, Reader in
A&E/Department of Health
Advisor in A&E, Warwick
Emergency Care and
Rehabilitation, University of
Warwick

Ms Bec Hanley, Co-Director,
TwoCan Associates,
Hurstpierpoint

Professor Norman Waugh,
Professor of Public Health,
Department of Public Health,
University of Aberdeen

Expert Advisory Network

Members

Professor Douglas Altman,
Director of CSM & Cancer
Research UK Med Stat Gp,
Centre for Statistics in
Medicine, University of Oxford,
Institute of Health Sciences,
Headington, Oxford

Professor John Bond,
Director, Centre for Health
Services Research, University of
Newcastle upon Tyne, School of
Population & Health Sciences,
Newcastle upon Tyne

Mr Shaun Brogan,
Chief Executive, Ridgeway
Primary Care Group, Aylesbury

Mrs Stella Burnside OBE,
Chief Executive, Office of the
Chief Executive, Trust
Headquarters, Altnagelvin
Hospitals Health & Social
Services Trust, Altnagelvin Area
Hospital, Londonderry

Ms Tracy Bury,
Project Manager, World
Confederation for Physical
Therapy, London

Professor Iain T Cameron,
Professor of Obstetrics and
Gynaecology and Head of the
School of Medicine,
University of Southampton

Dr Christine Clark,
Medical Writer & Consultant
Pharmacist, Rossendale

Professor Collette Clifford,
Professor of Nursing & Head of
Research, School of Health
Sciences, University of
Birmingham, Edgbaston,
Birmingham

Professor Barry Cookson,
Director, Laboratory of
Healthcare Associated Infection,
Health Protection Agency,
London

Professor Howard Cuckle,
Professor of Reproductive
Epidemiology, Department of
Paediatrics, Obstetrics &
Gynaecology, University of
Leeds

Dr Katherine Darton,
Information Unit, MIND –
The Mental Health Charity,
London

Professor Carol Dezateux,
Professor of Paediatric
Epidemiology, London

Mr John Dunning,
Consultant Cardiothoracic
Surgeon, Cardiothoracic
Surgical Unit, Papworth
Hospital NHS Trust, Cambridge

Mr Jonothan Earnshaw,
Consultant Vascular Surgeon,
Gloucestershire Royal Hospital,
Gloucester

Professor Martin Eccles,
Professor of Clinical
Effectiveness, Centre for Health
Services Research, University of
Newcastle upon Tyne

Professor Pam Enderby,
Professor of Community
Rehabilitation, Institute of
General Practice and Primary
Care, University of Sheffield

Mr Leonard R Fenwick,
Chief Executive, Newcastle
upon Tyne Hospitals NHS Trust

Professor David Field,
Professor of Neonatal Medicine,
Child Health, The Leicester
Royal Infirmary NHS Trust

Mrs Gillian Fletcher,
Antenatal Teacher & Tutor and
President, National Childbirth
Trust, Henfield

Professor Jayne Franklyn,
Professor of Medicine,
Department of Medicine,
University of Birmingham,
Queen Elizabeth Hospital,
Edgbaston, Birmingham

Ms Grace Gibbs,
Deputy Chief Executive,
Director for Nursing, Midwifery
& Clinical Support Services,
West Middlesex University
Hospital, Isleworth

Dr Neville Goodman,
Consultant Anaesthetist,
Southmead Hospital, Bristol

Professor Alastair Gray,
Professor of Health Economics,
Department of Public Health,
University of Oxford

Professor Robert E Hawkins,
CRC Professor and Director of
Medical Oncology, Christie CRC
Research Centre, Christie
Hospital NHS Trust, Manchester

Professor Allen Hutchinson,
Director of Public Health &
Deputy Dean of SchARR,
Department of Public Health,
University of Sheffield

Dr Duncan Keeley,
General Practitioner (Dr Burch
& Ptms), The Health Centre,
Thame

Dr Donna Lamping,
Research Degrees Programme
Director & Reader in Psychology,
Health Services Research Unit,
London School of Hygiene and
Tropical Medicine, London

Mr George Levvy,
Chief Executive, Motor
Neurone Disease Association,
Northampton

Professor James Lindesay,
Professor of Psychiatry for the
Elderly, University of Leicester,
Leicester General Hospital

Professor Julian Little,
Professor of Human Genome
Epidemiology, Department of
Epidemiology & Community
Medicine, University of Ottawa

Professor Rajan Madhok,
Medical Director & Director of
Public Health, Directorate of
Clinical Strategy & Public
Health, North & East Yorkshire
& Northern Lincolnshire Health
Authority, York

Professor David Mant,
Professor of General Practice,
Department of Primary Care,
University of Oxford

Professor Alexander Markham,
Director, Molecular Medicine
Unit, St James's University
Hospital, Leeds

Dr Chris McCall,
General Practitioner, The
Hadleigh Practice, Castle Mullen

Professor Alistair McGuire,
Professor of Health Economics,
London School of Economics

Dr Peter Moore,
Freelance Science Writer, Ashtead

Dr Sue Moss, Associate Director,
Cancer Screening Evaluation
Unit, Institute of Cancer
Research, Sutton

Mrs Julietta Patnick,
Director, NHS Cancer Screening
Programmes, Sheffield

Professor Tim Peters,
Professor of Primary Care
Health Services Research,
Academic Unit of Primary
Health Care, University of
Bristol

Professor Chris Price,
Visiting Chair – Oxford, Clinical
Research, Bayer Diagnostics
Europe, Cirencester

Professor Peter Sandercock,
Professor of Medical Neurology,
Department of Clinical
Neurosciences, University of
Edinburgh

Dr Eamonn Sheridan,
Consultant in Clinical Genetics,
Genetics Department,
St James's University Hospital,
Leeds

Dr Ken Stein,
Senior Clinical Lecturer in
Public Health, Director,
Peninsula Technology
Assessment Group,
University of Exeter

Professor Sarah Stewart-Brown,
Professor of Public Health,
University of Warwick,
Division of Health in the
Community Warwick Medical
School, LWMS, Coventry

Professor Ala Szczepura,
Professor of Health Service
Research, Centre for Health
Services Studies, University of
Warwick

Dr Ross Taylor,
Senior Lecturer, Department of
General Practice and Primary
Care, University of Aberdeen

Mrs Joan Webster,
Consumer member, HTA –
Expert Advisory Network

Feedback

The HTA Programme and the authors would like to know your views about this report.

The Correspondence Page on the HTA website (<http://www.hta.ac.uk>) is a convenient way to publish your comments. If you prefer, you can send your comments to the address below, telling us whether you would like us to transfer them to the website.

We look forward to hearing from you.