Health Technology Assessment 2005; Vol. 9: No. 35

Conceptual framework and systematic review of the effects of participants' and professionals' preferences in randomised controlled trials

M King, I Nazareth, F Lampe, P Bower,
M Chandler, M Morou, B Sibbald and R Lai

September 2005

Health Technology Assessment
NHS R&D HTA Programme

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.

Conceptual framework and systematic review of the effects of participants' and professionals' preferences in randomised controlled trials

M King,[1*] I Nazareth,[2] F Lampe,[2] P Bower,[3] M Chandler,[1] M Morou,[1] B Sibbald[3] and R Lai[4]

[1] Department of Mental Health Sciences, Royal Free and University College Medical School, London, UK
[2] Department of Primary Care and Population Sciences, Royal Free and University College Medical School, London, UK
[3] National Primary Care Research and Development Centre, University of Manchester, UK
[4] Medical Library, Royal Free and University College Medical School, London, UK

* Corresponding author

Declared competing interests of authors: none

Published September 2005

This report should be referenced as follows:

King M, Nazareth I, Lampe F, Bower P, Chandler M, Morou M, *et al*. Conceptual framework and systematic review of the effects of participants' and professionals' preferences in randomised controlled trials. *Health Technol Assess* 2005;**9**(35).

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, consumer groups and professional bodies such as Royal Colleges and NHS Trusts.

Research suggestions are carefully considered by panels of independent experts (including consumers) 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 designing 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 limited 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 98/26/03. The contractual start date was in November 2001. The draft report began editorial review in August 2003 and was accepted for publication in June 2004. 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.

ISSN 1366-5278

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

Conceptual framework and systematic review of the effects of participants' and professionals' preferences in randomised controlled trials

M King,[1*] I Nazareth,[2] F Lampe,[2] P Bower,[3] M Chandler,[1] M Morou,[1] B Sibbald[3] and R Lai[4]

[1] Department of Mental Health Sciences, Royal Free and University College Medical School, London, UK
[2] Department of Primary Care and Population Sciences, Royal Free and University College Medical School, London, UK
[3] National Primary Care Research and Development Centre, University of Manchester, UK
[4] Medical Library, Royal Free and University College Medical School, London, UK
* Corresponding author

Objectives: To develop a conceptual framework of preferences for interventions in the context of randomised controlled trials (RCTs), as well as to examine the extent to which preferences affect recruitment to RCTs and modify the measured outcome in RCTs through a systematic review of RCTs that incorporated participants' and professionals' preferences. Also to make recommendations on the role of participants' and professionals' preferences in the evaluation of health technologies.
Data sources: Electronic databases.
Review methods: The conceptual review was carried out on published papers in the psychology and economics literature concerning concepts of relevance to patient decision-making and preferences, and their measurement. For the systematic review, studies across all medical specialities meeting strict criteria were selected. Data were then extracted, synthesised and analysed.
Results: Key elements for a conceptual framework were found to be that preferences are evaluations of an intervention in terms of its desirability and these preferences relate to expectancies and perceived value of the process and outcome of interventions. RCTs differed in the information provided to patients, the complexity of techniques used to provide that information and the degree to which preference elicitation may simply produce pre-existing preferences or actively construct them. Most current RCTs used written information alone. Preference can be measured in many different ways and most RCTs did not provide quantitative measures of preferences, and those that did tended to use very simple measures. The second part of the study, the systematic review included 34 RCTs. The findings gave support to the hypothesis that preferences affect trial recruitment. However, there was less evidence that external validity was seriously compromised. There was some evidence that preferences influenced outcome in a proportion of trials. However, evidence for preference effects was weaker in large trials and after accounting for baseline differences. Preference effects were also inconsistent in direction. There was no evidence that preferences influenced attrition. Therefore, the available evidence does not support the operation of a consistent and important 'preference effect'. Interventions cannot be categorised consistently on degree of participation. Examining differential preference effects based on unreliable categories ran the risk of drawing incorrect conclusions, so this was not carried out.
Conclusions: Although patients and physicians often have intervention preferences, our review gives less support to the hypothesis that preferences significantly compromise the internal and external validity of trials. This review adds to the growing evidence that when preferences based on informed expectations or strong ethical objections to an RCT exist, observational methods are a valuable alternative. All RCTs in which participants and/or professionals cannot be masked to treatment arms should attempt to estimate participants' preferences. In this way, the amount of evidence available to answer questions about the effect of treatment preferences within and outwith RCTs could be increased. Furthermore, RCTs should routinely attempt to report the proportion of eligible

patients who refused to take part because of their preferences for treatment. The findings also indicate a number of approaches to the design, conduct and analysis of RCTs that take account of participants' and/or professionals' preferences. This is referred to as a methodological tool kit for undertaking RCTs that incorporate some consideration of patients' or professionals' preferences. Future research into the amount and source of information available to patients about interventions in RCTs could be considered, with special emphasis on the relationship between sources inside and outside the RCT context. Qualitative research undertaken as part of ongoing RCTs might be especially useful. The processes by which this information leads to preferences in order to develop or extend the proposed expectancy–value framework could also be examined. Other areas for consideration include: how information about interventions changes participants' preferences; a comparison of the feasibility and effectiveness of different informed consent procedures; how strength of preference varies for different interventions within the same RCT and how these differences can be taken account of in the analysis; the differential effects of patients' and professionals' preferences on evidence arising from RCTs; and whether the standardised measurement of preferences within all RCTs (and analysis of the effect on outcome) would allow the rapid development of a significant evidence base concerning patient preferences, albeit in relation to a single preference design.

Contents

List of abbreviations

AIDS	acquired immunodeficiency syndrome
CABG	coronary artery bypass grafting
CBT	cognitive behaviour therapy
CI	confidence interval
EBM	evidenced-based medicine
IP	interpersonal psychotherapy
ITT	intention-to-treat
NP	non-preferred
PTCA	percutaneous transluminal coronary angioplasty
QoL	quality of life
RCT	randomised controlled trial
URTI	upper respiratory tract infection

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

Participants in randomised controlled trials (RCTs) may have preferences for particular interventions that threaten external and internal validity. We tested three hypotheses: preferences affect recruitment to RCTs; preferences are important effect modifiers in RCTs; and the size of the effect modifier is larger in RCTs that require greater effort and participation by participants.

Objectives

The objective of this study was to develop a conceptual framework of preferences for interventions in the context of RCTs, as well as to examine the extent to which preferences affect recruitment to RCTs and modify the measured outcome in RCTs through a systematic review of RCTs that incorporated participants' and professionals' preferences. A further objective was to make recommendations on the role of participants' and professionals' preferences in the evaluation of health technologies.

Methods

The conceptual framework and review of measurement methods was based on a review of published papers in the psychology and economics literature concerning concepts of relevance to patient decision-making and preferences, and their measurement.

For the systematic review we included RCTs in the world literature that measured or recorded preferences, allocated participants based on preference and had follow-ups of non-randomised cohorts (registry studies) where patients received preferred treatment. We excluded reviews where there was no measurement or recording of preferences, RCTs of decision aids, reviews with *post hoc* measurement of preferences, registry studies with follow-up without regard to preferences and experiments testing normal volunteers.

Data extraction

The following data were extracted:

- general/study information
- setting and population
- experimental/control interventions
- RCT design
- elicitation/measurement of preference
- quality of randomisation
- baseline data
- participation
- management of attrition; type of analysis
- nature of primary outcome and whether defined by trialists or reviewers
- methods and results of analysis
- summary data for primary outcome(s).

Data were synthesised and analysed as follows:

- RCT quality
- elicitation/measurement of preference
- analysis of recruitment
- restriction of participants' preferences in the study design
- baseline differences between randomised and preference cohorts
- treatment participation
- attrition
- analyses in each report
- impact of preferences on outcomes.

Results

Conceptual framework

The following were found to be key elements for a conceptual framework of preferences in the context of RCTs:

- Preferences are evaluations of an intervention in terms of its desirability. Concepts in the wider literature of greatest relevance are utility in economics and attitude in psychology.
- Preferences relate to (a) expectancies concerning the process and outcome of interventions and (b) the perceived value of those processes and outcomes.
- Development of preferences and their influence on decision-making can be conceived of in terms of a four-stage model. The stages relate to information received about an intervention, the assimilation of that information, the development of a global preference and decision-making about randomisation

- RCTs differ in the information provided to patients, the complexity of techniques used to provide that information and the degree to which preference elicitation may simply elicit pre-existing preferences or actively construct them. Most current RCTs use written information alone.
- Preferences can be measured in a number of ways. Willingness-to-pay methods and attitude measurement within psychometrics may be most applicable.
- Most RCTs did not provide quantitative measures of preferences, and those that did tended to use very simple measures.

Systematic review

The search identified 10,023 citations, of which 44 were eventually included in the systematic review. This covered 34 RCTs.

- Most (25) were comprehensive cohort designs.
- Many failed to define a primary outcome(s), make a pre-RCT estimation of treatment effect, conceal randomisation or mask treatment groups to the outcome assessor.
- Quality of statistical analysis varied. Participants with missing data were often excluded from analysis, introducing potential bias.
- There was no consistent approach to examining preference effects.

Our findings give support to our first hypothesis, namely that preferences affect trial recruitment. However there was less evidence of bias in the characteristics of individuals agreeing to be randomised and therefore limited evidence that external validity was seriously compromised. With regard to our second hypothesis, there was some evidence that participant or physician preferences influenced outcome in a proportion of trials. However, evidence for moderate or large preference effects was weaker in large trials and after accounting for baseline differences. Preference effects were also inconsistent in direction. There was no evidence that preferences influenced attrition. Therefore, the available evidence does not support the operation of a consistent and important 'preference effect'. Interventions cannot be categorised consistently on degree of participation. Examining differential preference effects based on unreliable categories ran the risk of drawing incorrect conclusions, so we refrained from testing our third hypothesis.

Conclusions

Preferences are hypothesised to be based on expectancies concerning the process and outcomes associated with the intervention and the perceived value placed on those outcomes and processes. However, participants' preferences may be based on insufficient or incorrect information. In addition, decisions about treatment choice may not always accord with preferences and may be influenced by clinicians, relatives or friends. When preferences are likely to affect the external validity of an RCT, it is important to present potential participants with appropriate evidence, without straying into coercion. We have suggested how preferences might best be measured. Once participants have been recruited, preferences may affect perceptions of the intervention and satisfaction but appear to exert few major effects on further participation or clinical outcome. Comprehensive cohort designs may still be worthwhile; however, when a significant proportion of patients refuse to be randomised and (1) follow-up data are economical to collect, for example, from routinely collected sources, or (2) when costs of follow-up are higher, a random sub-sample of participants are allocated to their preferred treatment and followed up.

Our review also adds to the growing evidence that when preferences based on informed expectations or strong ethical objections to an RCT exist, observational methods are a valuable alternative. Data from observational studies may be valuable in situations where:

- there are strong preferences based on informed expectations on the part of eligible participants or physicians and when only a small proportion of them will accept randomisation;
- known confounders of treatment outcome (including strength of preference) are measured and taken account of in the analysis;
- there are strong ethical or legal objections to undertaking an RCT.

All RCTs in which participants and/or professionals cannot be masked to treatment arms should attempt to estimate participants' preferences. This would increase the amount of evidence available to answer questions about the effect of treatment preferences within and outwith RCTs. Furthermore, RCTs should routinely attempt to report the proportion of eligible patients who refused to take part because of their preferences for treatment. Beyond these two general recommendations, our findings also indicate a number of approaches to the design, conduct and analysis of RCTs that take account of participants' and/or professionals' preferences. We refer to these as a methodological tool kit for undertaking RCTs that incorporate some consideration of patients' or professionals' preferences.

Relevance to the NHS

Besides understanding more about how participants' and professionals' preferences affect the internal validity of RCTs and informing professionals and patients about the need for good evidence of efficacy, we need greater application of information systems within the NHS to make use of routine data collection as one source of evidence on effectiveness.

Recommendations for research

The following areas are suggested for future research:

- An assessment of the amount and source of information available to patients about interventions in RCTs, with special emphasis on the relationship between sources inside and outside the RCT context. Qualitative research undertaken as part of ongoing RCTs might be especially useful.
- An examination of the processes by which this information leads to preferences in order to develop or extend the proposed expectancy–value framework. Key questions relate to the type of expectancies that enter into decision-making, and the way in which different expectancies are valued by patients. Conjoint analysis may be especially useful in this regard.
- An investigation into how information about interventions changes participants' preferences and a comparison of the feasibility and effectiveness of different informed consent procedures.
- A study of how strength of preference varies for different interventions within the same RCT and how these differences can be taken account of in the analysis.
- An exploration of the differential effects of patients' and professionals' preferences on evidence arising from RCTs. Our findings suggest that patients' preferences act mainly at recruitment. Professionals' preferences may affect external and internal validity but the number of RCTs in which professionals' preferences were reported was very small.
- An assessment of whether the standardised measurement of preferences within all RCTs (and analysis of the effect on outcome) would allow the rapid development of a significant evidence base concerning patient preferences, albeit in relation to a single preference design.

Chapter 1
General introduction

Introduction to the report

This report begins with an outline of the basic tenets of randomised controlled trials (RCTs), how participants' and professionals' preferences have an impact on the intervention and the research process and how RCT designs have been adapted to take account of preferences. We then describe the objectives of the present study, before describing the conceptual framework. The systematic review follows and the results of both are brought together in a discussion of common methods and findings in this field. We end with a number of conclusions and recommendations.

General introduction

A key tenet of the evidence-based medicine (EBM) paradigm is that the objective evaluation of the outcome of health service interventions is a requirement for their efficient provision in clinical practice.[1] The most reliable evidence for treatment efficacy comes from RCTs, because of the high levels of internal validity that are associated with this form of study design.[2] Randomising patients between intervention and control arms means that the average intervention effect, uncontaminated by all confounding, can be estimated by comparing the main outcome measure between the groups. Statistical theory is based on the assumption of random sampling and therefore in an RCT the differences between intervention arms behave like the differences between random samples from a single population.[3] Generally, it is hypothesised that less rigorous study designs than RCTs lead to **overestimates** of the effect of interventions[4,5] and that this is due to selection effects increasing confounding with important prognostic variables, although the evidence base is limited and this may not always be the case.[6,7]

Beyond the requirements of informed consent, patients are conceptualised as relatively passive recipients of care in experimental research such as RCTs. However, there is a significant literature in areas such as the placebo effect,[8] demand characteristics[9] and the Hawthorne effect[10,11] suggesting that patients are far more active participants in research.[12] One important way in which patients may not be passive recipients of experimental interventions is that, in an RCT, there are two or more competing interventions available: patients may have **preferences** for particular interventions, and would prefer to receive one over the other if given the choice. This may relate to particular aspects of the intervention that are perceived favourably (such as the lack of side-effects in psychological therapy as compared with medication or differences in travel requirements between home-based and clinic-based treatments) or may simply relate to the fact that one intervention is 'new', the other is a 'control' and 'new' may have connotations of 'improved'.[13]

Internal validity

In RCTs, the influence of factors such as preferences is minimised through blinding procedures.[13,14] When patients cannot tell which intervention they have received, their preferences cannot interact with their intervention assignment to reduce the internal validity of the RCT design and bias the results. However, the use of blinding procedures can be problematic.

Despite the use of blinding procedures, it may be possible for RCT participants to determine which intervention they have received, even if they are ostensibly identical pill preparations in an RCT of medication. Differences in side-effects from the interventions or changes in intermediate outcomes may provide information about the intervention to which a patient has been randomised,[15] which means that subjective processes such as preferences may still influence eventual outcomes and bias RCT results. Asking participants which intervention they have received provides one method of determining whether a blind has been successful: if participants' guesses are more accurate than that expected by chance, then the blind may have been broken.[13]

Second, RCTs are increasingly being used to evaluate interventions that are far less amenable to blinding than medication, such as surgical or psychosocial interventions. Where treatments cannot be blinded, preferences have the potential to impact on the internal validity of RCTs, **if preferences are capable of interacting with the intervention to influence outcome**.

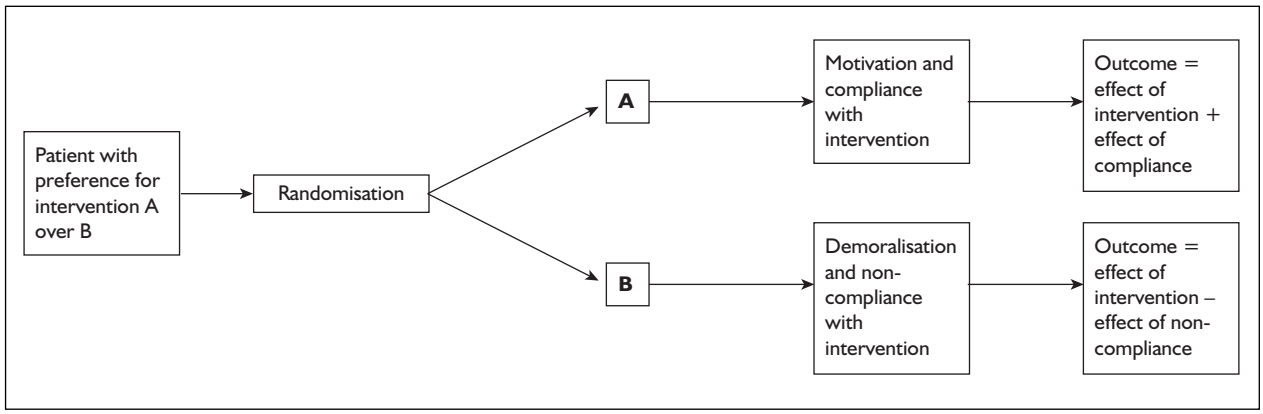

FIGURE 1 *Direct effect of preferences*

Preference–intervention interactions and the process and outcome of the intervention

Preference–intervention interactions are of two types. The first type involves a **direct effect** on the process of care, where patients who do not receive their preferred intervention may suffer **resentful demoralisation**[12] and **refuse to comply with an intervention** (*Figure 1*).

Although direct effects have the potential to impact on almost all interventions, the scope of that effect will depend on the exact nature of the intervention, particularly in terms of the level of **patient involvement** or **participation**. Where patients are passive recipients of interventions such as surgery, they may of course simply refuse to receive the intervention at all. However, once they have agreed to receive it, they can have little influence on the process of a surgical operation. In contrast, an intervention such as a medication regime requires more active cooperation on the part of the patient, although the scope for active participation and autonomy is relatively limited. Finally, there are interventions which require informed, active and motivated participants to engage fully in the **process** of a particular intervention.[16] Examples of such interventions are psychological therapy, self-care interventions, and professional education and training, all of which are increasingly the subject of RCTs. The scope and magnitude of preference–intervention interactions may be far greater in these contexts.

In addition to the direct effects of preference on compliance, there is also the possibility of **indirect effects**. Even when a patient cannot readily influence the process of the intervention (e.g. surgery), patient's preferences may relate to **expectancies about the nature or effectiveness** of the different interventions.[17,18] Such expectancies are a key mechanism of action of the **placebo** effect,[8] and a recent HTA review showed that such expectancies can have a substantial impact on outcomes (*Figure 2*).

The impact of both direct and indirect preference–intervention interactions may vary according to whether the RCT uses objective outcomes (i.e. almost always physiological or biological parameters, such as blood pressure or mortality) or subjective outcomes (i.e. self-report health status or satisfaction with the intervention). Although both direct and indirect effects may impact on both subjective and objective outcome measures, indirect effects may have a greater influence on subjective outcomes, as subjective outcomes may be more vulnerable to preference effects mediated through a fundamentally psychological mechanism such as expectancy.

The combination of direct and indirect effects and objective and subjective outcomes provides a number of possible causal pathways. *Figure 3* illustrates these (only positive effects of receiving a preferred treatment are considered in the figure for simplicity).

The possible causal pathways for direct effects are:

1. Increased compliance with an intervention which affects objective outcomes only, such as blood pressure medication.
2. Increased compliance with an intervention which affects subjective outcomes only, such as psychological therapy.
3. Increased compliance with an intervention which affects objective outcomes, which in turn affect subjective outcomes, such as blood pressure medication which reduces

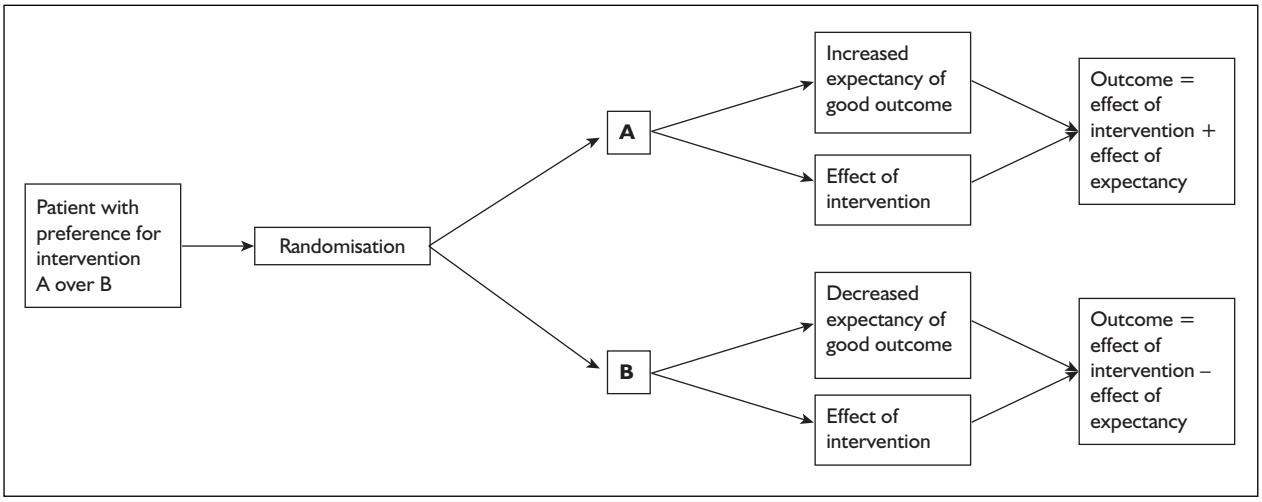

FIGURE 2 *Indirect effect of preferences*

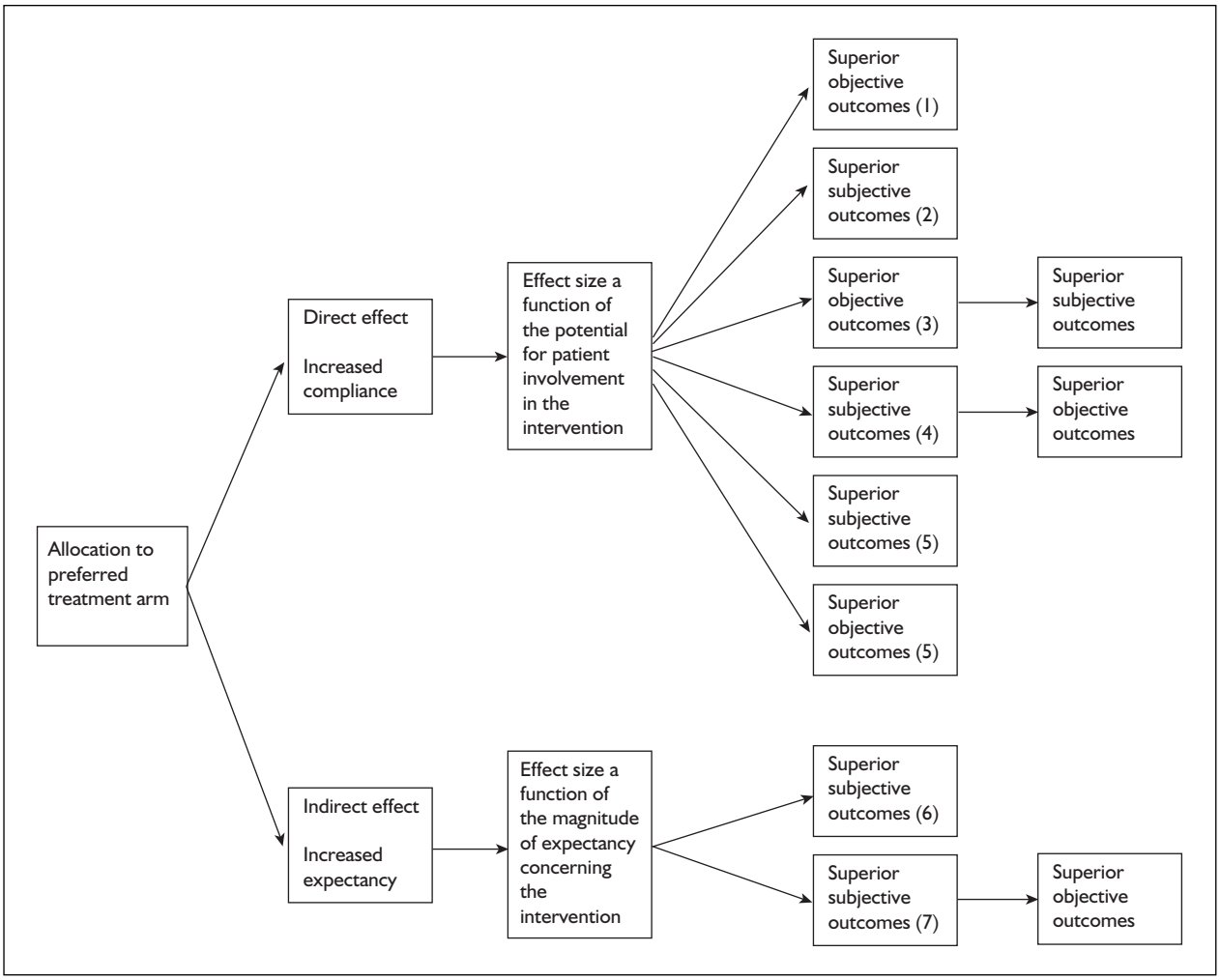

FIGURE 3 *Causal mechanisms for preference effects. Numbers (1) to (7) refer to examples in the text.*

blood pressure and improves quality of life (QoL).
4. Increased compliance with an intervention which affects subjective outcomes, which in turn affect objective outcomes, such as using a self-help programme which improves QoL and thus reduces health service utilisation.
5. Increased compliance with an intervention which affects objective and subjective outcomes simultaneously without one being mediated through the other, such as an exercise intervention that reduces obesity **and** depression through different causal mechanisms.

The possible causal pathways for indirect effects are:

6. Increased expectancy concerning an intervention which affects subjective outcomes only ('psychological expectancy' effect).
7. Increased expectancy concerning an intervention which affects subjective outcomes, which in turn affect objective outcomes ('psychosomatic expectancy' effect), such as a graded exercise intervention for chronic fatigue syndrome that reduces fatigue and thus increases an objective strength exercise or a return to work.

Preference–intervention interactions and the research process

There are two other possible effects of preference on RCTs that may influence internal validity through their influence on aspects of the research process, rather than the process of care within the intervention. First, being denied choice of treatment may **bias** subjective outcome measurement. For example, patients who are suffering the same subjective outcomes (e.g. symptoms of anxiety and depression) as other patients may rate the impact or significance of those symptoms as greater or their satisfaction with that treatment as lower if they were denied their preferred treatment. This might represent the effect of resentment at being denied preference and choice *per se*, rather than an effect mediated through expectancy concerning that treatment. However, given that subjective outcomes are based on self-report, such effects may be almost impossible to disentangle from general indirect effects.

Second, in addition to not complying with the process of the intervention (i.e. the direct effect), patients may refuse to comply with **research** assessments designed to evaluate the intervention. This may reduce sample size and available power if attrition is high in both arms of the RCT, and reduce internal validity if there is differential attrition between arms in an RCT, which is associated with preferences.

Internal validity: summary

Preference–treatment interactions have important implications for the internal validity of RCTs. Generally, it is held that less rigorous study designs than RCTs lead to **overestimates** of the effect of interventions, owing to selection effects. However, if preference–intervention interactions are apparent and of significant magnitude, it is possible that removing choice and randomising patients to their non-preferred interventions might impact on outcome such that the results of an RCT provide a biased estimate of the effects of the intervention, that is, randomisation may be a threat to internal validity.[19]

External validity

External validity refers to the confidence with which a researcher can expect relationships found in the context of one particular study to **generalise** to other contexts. For example, if an intervention is found to be efficacious in a particular setting or population of patients (such as those participating in an RCT), will the same results be found in a different setting, such as patients recruited outside RCTs?[12] Generally, external validity is judged on the basis of **proximal similarity** – results can be generalised 'with most confidence where treatment, setting, population, desired outcome, and year are closest in some overall way to the original program treatment'.[20]

However, if recruitment and participation in RCTs are biased because of preferences, external validity may be compromised, as the particular population of patients recruited to the RCT (who may have no preferences and agree to randomisation) may not be representative of the population to which the results are to be applied.

Preference effects on recruitment can act at the level of the patient, practitioner or centre.[6] The requirement of informed consent means that RCT participants almost always have information about the alternative interventions under test, and preferences for particular interventions may mean that patients refuse to enter the RCT and risk allocation to their non-preferred treatment.

Practitioners and centres may also have preferences concerning the competing interventions in an RCT, which may influence the

likelihood that they participate in the RCT at all. Even if they do participate, it may reduce the likelihood that they will recruit individual eligible patients.[21] This may impact on the external validity of the results of an RCT if non-participating practitioners and centres serve different populations, or practitioners recruit only certain subgroups of patients, and this leads to patients in RCTs differing in their capacity to benefit from other patient groups to which the results of the RCT are to be applied.

A previous HTA review provides a useful visual model of the possible effects of preference on internal and external validity.[6]

Chapter 2

Introduction to the role of preferences in randomised trials

Changing public and professional attitudes

Patients usually agree to participate in research at the request of their health professionals.[22] However, undergoing surgery, attending an outpatient clinic or psychotherapist or accepting admission to hospital are interventions that require an investment of time and energy by participants, who may hesitate to accept the lottery of randomisation. Furthermore, people might accept in principle the justification for a randomised experiment but prefer that it was others who ran the risk of receiving no treatment or missing out on the promising new intervention. They may also hesitate to take part in RCTs if, for whatever reason, they regard the new treatment(s) as dangerous or inferior to standard treatment(s) or, conversely, if they are eager to receive a new treatment. Even when the design of the RCT means that potential advances in treatment are only available in the experimental arms, patients may hesitate and request that the new intervention be available in some form to everyone. Powerful patient advocacy groups in the 1980s demanded that new drugs for AIDS were tested in RCTs in increasing doses rather than against a placebo, or covertly encouraged patients to share tablets in order that all would be exposed to at least some of the experimental drug. Patients with life-threatening illness and their families may not wish to risk reducing the QoL of their remaining days or weeks in an RCT of uncertain outcome. They may lack the altruism required to enter an RCT designed to improve the effectiveness of palliative care medicine for everyone but which may have an uncertain effect on their own lives.

Practitioners may not wish to randomise their patients to treatments in which they lack confidence or believe are not indicated. One ethical requirement for an RCT is that the trialists, in addition to the professionals recruiting patients, be in a state of clinical equipoise. This means that the balance of evidence could go either way, and the effectiveness of the new treatment is not known to be any better (or worse) than the standard. However, few practitioners have completely neutral views about the comparative efficacy of any two or more treatments. Many may be unaware of or unconvinced by the evidence for or against the intervention under examination.

Modifications of RCT design to take account of preferences

A number of adaptations have been developed to incorporate preferences into the traditional RCT. We identified these adaptations from (1) our previous knowledge of the field and (2) our conceptual framework and systematic review, which are reported in detail later. Many, but not all, of these have preference arms in which participants choose their treatment and randomised arms to which they are allocated randomly. In discussing these RCTs throughout the report, we shall use a short-hand that summarises the types of comparisons we are making.

If R1 and R2 represent the randomised arms and P1 and P2 the preference arms in a comparison of two interventions (two active or one active versus a placebo) then:

- Treatment effects are represented by:
  R1 versus R2
  P1 versus P2
  (R1 + P1) versus (R2 + P2)
- Preference effects are represented by:
  R1 vs P1
  R2 vs P2
  (R1 + R2) vs (P1 + P2).

RCTs that incorporate preference information[23,24]

1. Participants' preferences and expectancies are determined before and after randomisation and adjusted for as important baseline variables in the determination of outcome. Otherwise, the RCT follows the traditional model.
2. Preferences for each arm are measured and randomisation is stratified on this basis in order to achieve a balanced range of preferences within each arm.

Neither option (1) nor (2), however, prevents people with strong preferences from refusing to participate, but preferences must be known if their interaction with treatment outcomes is to be examined. Unfortunately, in many RCTs no attempt is made to measure participants' or professionals' preferences.

Comprehensive cohort (or patient preference) designs

Participants whose preferences for one or more treatment arms are so strong that randomisation is refused may be included as cohorts alongside the RCT by allocating them to the treatment of their choice and estimating outcome exactly as if they were a part of the RCT. This comprehensive cohort[25] or patient preference design[16] allows most eligible participants to be followed up under RCT conditions and comparisons of outcome made between randomised and non-randomised patients (*Figure 4*). However, important baseline differences between preference and randomised patients may compromise the value of this comparison. Modifications to the comprehensive cohort design have been developed:

1. When participants have strong preferences **against** one particular treatment arm in an RCT containing three or more arms, they may be prepared to be randomised between the remaining arms.[26,27] Although this comparison will not answer the main question posed by the RCT, it may provide greater power to examine differences in outcome between specific arms.
2. Where practitioners' preferences are the dominant factor reducing or biasing RCT recruitment, a design that mirrors the comprehensive cohort can be applied. The practitioner only randomises participants for whom he or she cannot decide which treatment arm is most indicated and the remainder are given the treatment of the clinician's choice.[28] Although this emphasis on clinical equipoise mirrors the reality of standard practice, it may restrict the external validity of the RCT if few participants are randomised. As discussed above, it is also important to be aware of the basis on which practitioners make their judgements. The fact that an RCT is planned at all suggests that current professional practice is not evidence based and is thus open to question.

Comprehensive cohort designs do not mean that **all** eligible participants enter the RCT. A small proportion of eligible participants will inevitably refuse to take part in any form of research. Furthermore, patients with preferences may temporarily agree to randomisation for altruistic reasons or if they perceive pressure from the researchers and thus the disadvantages of simple

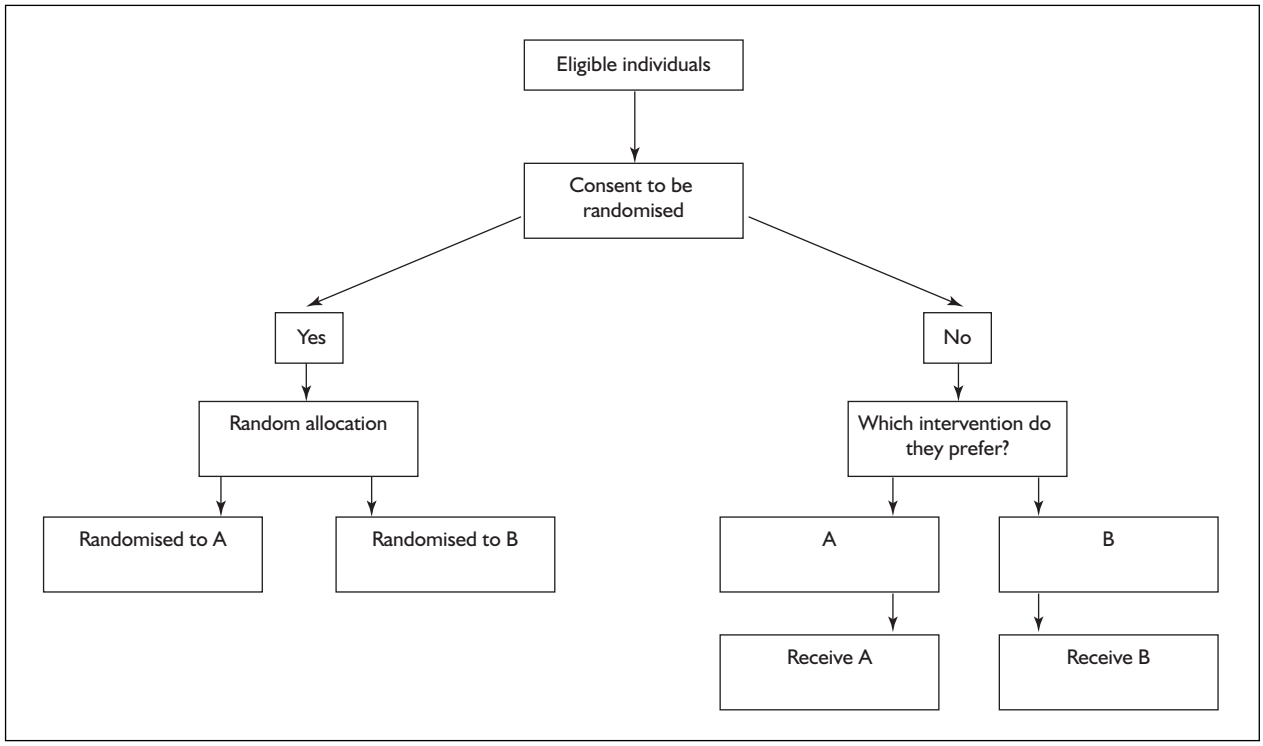

FIGURE 4 *Comprehensive cohort design*

RCTs may not all be avoided. Comprehensive cohort designs may also be expensive to conduct if a large proportion of suitable participants refuse to be randomised.[23,26,27] Recruiting only a random sample of participants with strong preferences into the preference arms may reduce such unnecessary costs.

Prerandomised designs

A further alternative to deal with preferences is the so-called Zelen design, in which participants are randomised **before** recruitment.[29] This design is radically different to the designs already described in that it aims to remove rather than measure or account for participants' preferences. In its original form, neither patients in the experimental nor those in the control arms were aware of the randomisation and merely gave consent to measurement of outcome. Because of ethical objections, later modifications were based on which participants were made aware of the randomisation.

1. In the single randomised consent design, only participants in the experimental group(s) are informed that they are in the RCT. Those in the usual or standard treatment arm are not informed on the grounds that this would have been their treatment if the RCT had not taken place. Participants in the experimental arm(s) who refuse to continue are allocated to the standard treatment arm but analysed on an intention-to-treat (ITT) basis as if they were a part of the experimental group. This is equivalent to treating them as protocol violations. Although widely written about, this design was rarely employed mainly because of ethical objections to research in which some participants are unaware they are part of an experiment. An RCT in which patients receiving the standard treatment (who are unaware they are participating) are more likely to complete treatment and undertake outcome assessments than those informed that they are in an experimental arm also violates the statistical assumption that participants are assigned at random to treatment arms. Finally, in some settings only informing patients in the experimental care arm may disadvantage potential participants in RCTs where the experimental treatment(s) may be more popular than usual care. One example is the RCT by King and colleagues, referred to above,[26,27] in which standard primary care treatment for depression was less popular than the brief talking therapies, whose effectiveness was the focus of the RCT.
2. In the double randomised consent design, **all** participants are informed of their group allocation and those who do not agree are

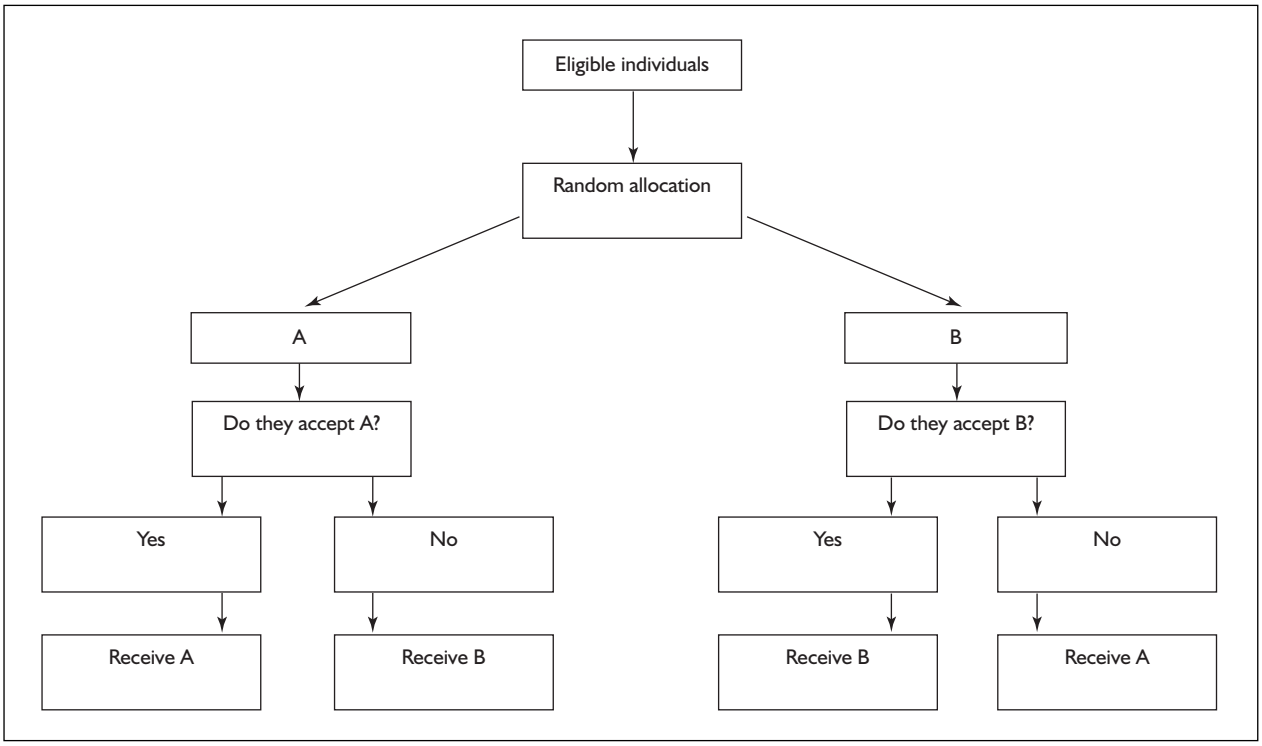

FIGURE 5 *Zelen's design with double randomised consent*

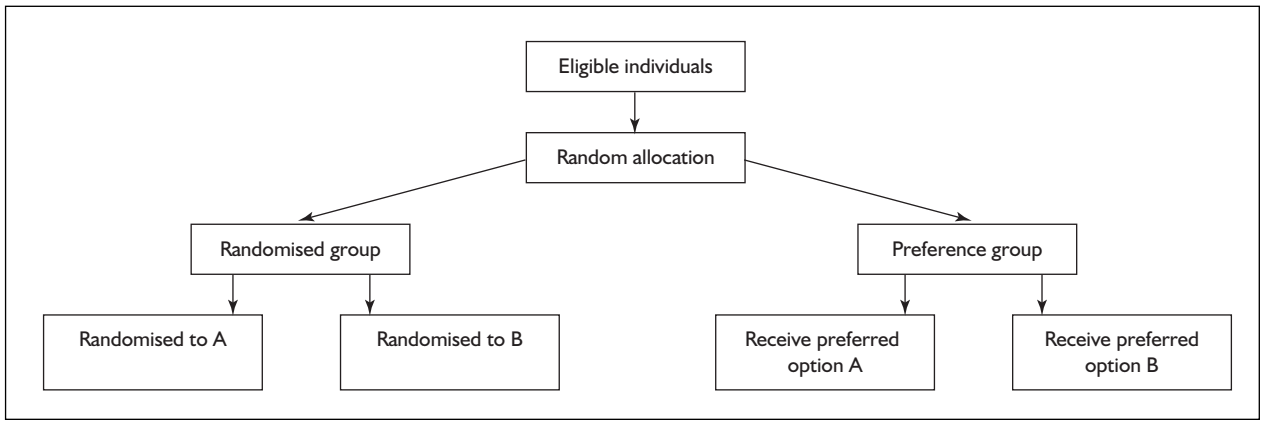

FIGURE 6 *Wennberg design*

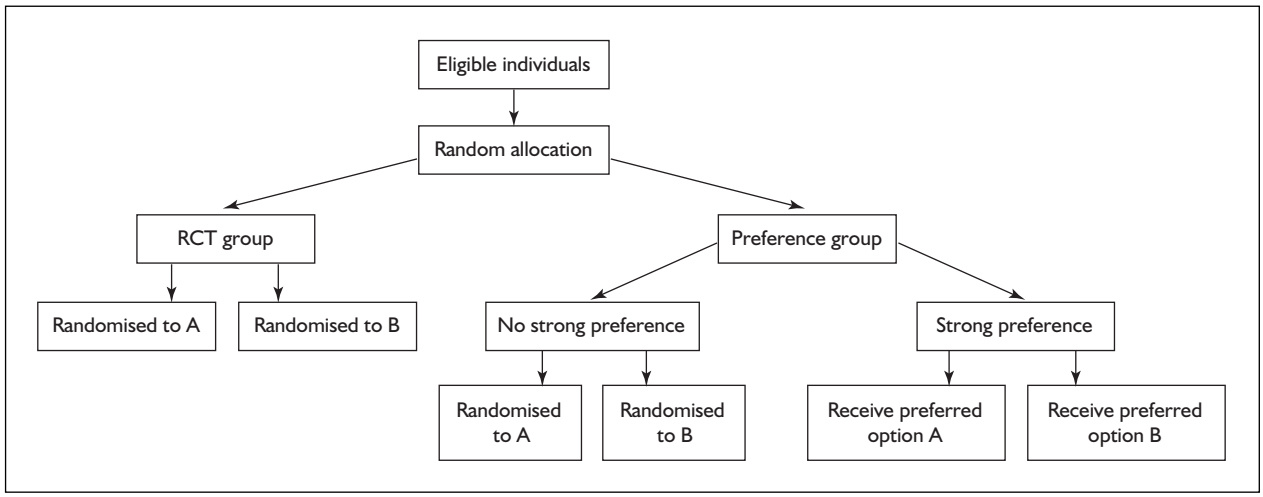

FIGURE 7 *Rücker design*

given the opportunity to transfer to an alternative treatment arm (*Figure 5*). However, this design runs a similar risk of compromise if too many participants do not accept their randomisation and offers little advantage over a standard RCT.[25]

The two-stage, randomised clinical design

1. Yet a further modification to take account of participants' preferences is the so-called Wennberg design, in which participants are initially randomised to two groups (*Figure 6*). In the first, they are offered a choice of treatment arm in the same way as the preference groups in a comprehensive cohort design. In the second, participants are randomised to treatment arms as in a classical RCT.[2,30]
2. In a modification of this design, participants randomised to the preference arm in the first randomisation, **who do not have a strong preference for a treatment arm**, are randomised a second time to a treatment arm (the Rücker design; *Figure 7*).[31]

These designs offer a more powerful method of determining the influence of preference on outcome than the comprehensive cohort design.[32] However, they do not circumvent the difficulty of recruiting patients who have such strong preferences that they are not prepared to risk the first randomisation procedure.

Advantages and disadvantages of each design

Each design has a number of different methodological advantages which are summarised in *Table 1*.

TABLE 1 *Advantages and disadvantages of each type of design*

Measurement of preference at baseline in a standard RCT
Internal validity – preference effects can be used as a stratification factor (to reduce the impact of preferences) or as a predictor of outcome
External validity – patients with very strong preferences may not enter the study, which may reduce or remove preference effects
Study administration – no increase in sample size, but large sample size may be required to detect preference interaction effects, valid and reliable measures of preferences required
Comprehensive cohort design
Internal validity – preference effects (e.g. R vs P) confounded, although can be controlled
External validity – almost all eligible patients enter the study and allows examination of characteristics of patients with all strengths of preferences
Study administration – potentially costly if large numbers of patients express a preference and not feasible if very few patients have a preference. *A priori* power calculations difficult if there is no prestudy estimate of the percentage accepting randomisation
Prerandomised (Zelen) design
Internal validity – all patients randomised but, depending on consent process, uneven drop-out may occur between intervention and control arms
External validity – all eligible patients enter study but ethical objections exist over lack of fully informed consent
Study administration – potentially low cost as all eligible patients will enter study but depending on later consent process, drop-out or switching between arms may make increased recruitment necessary. Ethical concerns in designs with partial or no patient consent
Two-stage, randomised designs (Wennberg and Rücker)
Internal validity – all patients randomised, increasing internal validity. However, P vs P and R vs P comparisons still subject to confounding as patients' characteristics may determine choice of treatment
External validity – reduced because only patients accepting randomisation enter the study Study administration – people with strong preferences may refuse randomisation

Conduct of RCTs

Ethical and public objections to randomisation have fuelled these imaginative attempts to modify RCTs so that effectiveness of treatments in challenging medical or social situations may be evaluated. There is no consensus, however, on which design fits which particular clinical or social circumstances, on how to standardise measurement of preferences in a variety of clinical or social environments or on inclusion and effect of preferences in the analysis of outcomes. These are the subjects of the review that follows.

Chapter 3

Hypotheses and methods

Hypotheses

Three principal hypotheses were tested in this review:

1. Preferences affect recruitment to RCTs.
2. Preferences are important effect modifiers in RCTs.
3. The magnitude of preference effects is larger for treatments that require greater effort and participation by participants.

Objectives

Our objectives were to:

1. Develop a conceptual framework of preferences for interventions in the context of RCTs.
2. Conduct a systematic review of RCTs incorporating participants' preferences, in order to examine the extent to which:
   (a) Preferences affect recruitment to RCTs.
   (b) Preferences modify the measured outcome in trials.
3. Make recommendations on the role of participants' preferences in the evaluation of health technologies.

Method

We searched scientific, psychological and medical databases for RCTs that accounted for patients' or professionals' preferences in their design. The databases searched were:

MEDLINE	1966–2001	CD-Plus Ovid Version
EMBASE	1980–2001	WebSpirs/SilverPlatter
PsycINFO	1984–2001	Ovid
CINAHL	1982–2001	Ovid
AMED	1985–2001	Ovid
Cochrane Library	1993–2001	

Compilation of the search strategy

Based on our previous knowledge of the subject area and after reading key papers, we generated keywords and phrases that we incorporated into the search strategies.

The two themes incorporated into the search strategy were (1) **preferences** and (2) p**ossible determinants of preferences**, such as socio-economic factors, decision-making, question framing and risk factors. The themes were translated into thesaurus terms and their equivalent text words and phrases. A lack of a precise thesaurus term to describe the concept of preference across all databases severely hindered the development of the search strategies. For example, when using the MEDLINE database, the searcher is directed to the term 'PATIENT SATISFACTION'. The scope note, however, for this term is as follows: 'The degree to which the individual regards the health service or product or the manner in which it is delivered by the provider as useful, effective or beneficial.' This definition emphasises satisfaction rather than preference.

The next filter incorporated in the search strategy was designed to capture RCTs, meta-analyses and systematic reviews. The aim was to capture RCTs that had attempted to account for and measure either practitioners' or patients' preferences and to assess the effect of these preferences on the outcome. Research designs specific to incorporating preferences into RCTs such as Zelen designs, two-stage randomisation (Wennberg and Rücker) designs, partially randomised studies and RCTs with preference arms were included in the search strategy so as to ensure identification of all types of preference studies.

We developed the search strategy initially in the MEDLINE database and undertook eight revisions. After running each version of the strategy, the research team assessed the outputs for their overall degree of relevance. Large numbers of records were retrieved in the early searches but, when these were examined, many papers of low relevance were found. The text words and thesaurus terms contributing to this excess were then excluded from the search strategy to improve the specificity of the search. For example, the text word 'decision' and the phrase 'quality of life' were eliminated from the search in order to provide a more focused strategy

(Appendix 8). This was translated and applied to the other databases listed above.

Extent of coverage

We searched four major biomedical databases, namely MEDLINE, EMBASE (renowned for its coverage of the European literature), PsycINFO and CINAHL alongside AMED (Allied and Complementary Medicine) and the Cochrane Library (see below). These databases ensured that a comprehensive search was achieved, considering the year spans, the range of journal titles and of publications types collectively indexed.

The Cochrane Library search included the Cochrane Database of Systematic Reviews, the Database of Abstracts of Reviews of Effectiveness and the Cochrane Controlled Trials Register, which were searched through the KA24 information gateway, using the following text words string: (preference$ OR choice$ OR utility$).ti. This cut-down version of the Cochrane Library does not have a thesaurus, hence we used text words. In addition to the three Databases listed above, the full version of the Cochrane Library also includes the Health Technology Assessment database, the NHS Economic Evaluation database, the Cochrane Database of Methodology Reviews and the Cochrane Methodology Register. The NHS Economic Evaluation database (NHSNEED) is the only database that focuses on economic evaluation of biomedical interventions. It selects studies for evaluation and inclusion by searching the MEDLINE, EMBASE and CINAHL databases. Hence our search strategy would have retrieved any studies examining the cost-effectiveness or any other form of economic evaluation relating to preference or choice of medical treatments.

Lastly, we supplemented the database searches with (1) hand searching of key journals for the years 2000–2003 on the basis that more recent publications are more likely to be missed by the search engines (see Appendix 9 for list of journal titles); (2) contacting experts who had published preference studies; and (3) searching through the reference lists of relevant journal articles and reports.

Systematic review

Study selection

Two researchers (MC and MM) examined the citations retrieved from each database search. Owing to the number of citations retrieved, it was not considered practical to have each researcher scan each citation independently. This process was adopted, however, for those citations where there was some doubt about relevance. If there were disagreements between the researchers, these were discussed at the steering group for a final decision. Citations that lacked sufficient information or did not have an abstract were retrieved in full text for further consideration. Once a final list of papers had been obtained we scanned the reference lists of each paper for relevant articles that might have been missed in the database searches or that provided more information on the particular study under consideration.

We applied the following inclusion/exclusion criteria to the review:

1. Inclusion:
   (a) Any RCT that **measured** or **recorded** patient or physician preference.
   (b) Any RCT that allocated participants based on preference.
   (c) RCTs with follow-ups on a non-randomised cohort (i.e. registry studies), **where patients received their or their physicians' preferred treatment**.
2. Exclusion:
   (a) No measurement or recording of patient or physician preference.
   (b) RCTs of decision aids or similar interventions.
   (c) When participants' preferences were measured after delivery of the intervention or after completion of the study.
   (d) Registry studies where patients who refused randomisation were merely followed up without reference to preferences
   (e) Experimental studies testing normal volunteers under laboratory conditions.

We did not apply any language restriction to our review and hence a number of potentially relevant foreign language citations were identified. In most cases, it was possible to determine from the English abstract found in the database searches whether or not to retrieve the full paper for consideration. If an abstract in English was not available, the full paper was retrieved where possible. Non-English language papers were translated and reviewed for possible inclusion.

Missing data

Where relevant data (e.g. in the comparison R1 versus P1) were not reported, we contacted authors

with requests for further information. In the event that the required data were not forthcoming, or the authors could not be contacted, the paper was included in the review but excluded from our calculations of preference effects.

Data extraction

Standard data extraction forms generally used for systematic reviews of randomised studies were not appropriate here, as they could not account for the preference arms of the RCT. Therefore, we drafted a form and piloted it by extracting data from two preference studies that were to be part of the systematic review (Appendix 10). The main sections of the form are listed below.

1. General study information, that is, the authors' names, study title and source (e.g. journal name, volume, year of publication).
2. Information on study population: a general description of the people in the sample (e.g. people with low back pain) and the location where the study was conducted (e.g. primary care, secondary care or unclear).
3. Details of the experimental and control interventions.
4. The study design (e.g. comprehensive cohort, Zelen design).
5. Quality of the randomised component of the study:
   (a) Randomisation – quality of allocation was assessed using The Cochrane Collaboration criteria as follows: **OK** = central randomisation or explicit statement about sealed opaque envelopes; **unclear** = insufficient information, studies did not report the concealment approach; **vulnerable** = use of alternation or date of birth.
   (b) Blinding/masking of participants, trialists and professionals – these are reported as **yes**, **no**, **not possible** or **not clear**.
   (c) How attrition was dealt with and whether an ITT analysis was used.
6. Baseline data reported in the following categories:
   (a) Demographic details, particularly age, gender, educational level and social class.
   (b) Data specific to the outcomes of interest.
   (c) Differences in baseline measures between randomised and preference groups.
7. Recruitment rates of eligible patients to the RCT and the proportion of recruited patients agreeing to being randomised.
8. The numbers of participants with data on the primary outcome in each arm of each RCT at baseline and all follow-up points.
9. The nature of the primary outcome measure and whether the authors clearly defined it in the text of their report or we inferred it for this review. We used the following checklist to ascertain whether the primary outcome was identified by the trialists:
   (a) Explicit statement by the authors in any section of the paper.
   (b) Implicit reference in the study hypotheses.
   (c) *A priori* power analysis based on a specific outcome of interest.
   If none of these were identified, the most likely and appropriate primary outcome was chosen by consensus by IN, MK and FL. When opinions were divided over two likely outcome measures, data were extracted for both outcomes of interest.
10. Data on the primary outcome(s) for each RCT at baseline and each follow-up point. This included descriptive data (means and standard deviations for continuous variables, proportions for categorical data), *p*-values and measurement of treatment effect.
11. Information on the analysis plan described by the trialists for each study and the results of the analysis presented in their publications.

Analysis

Quality

We assessed the RCTs found in terms of recognised criteria such as identification of primary outcome by the trialists, quality of randomisation and masking, but did not reject them on these bases, as they were not central to our questions. Neither did we judge them on aspects such as the proportions lost to follow-up as these are important considerations in the success or otherwise of alternative designs.

Methods to elicit and measure preference

The methods by which participants' preferences were elicited and measured were recorded for each RCT in the review. Details are presented in Chapter 4.

'Hierarchy' of preference study design

We considered but rejected an approach wherein study designs might be placed on a traditional 'hierarchy of evidence' in terms of answering our main question: do preferences affect outcome? All designs have advantages and disadvantages (see *Table 1*). Standard RCTs that measure preferences

at baseline are useful for examining the interaction of preference with treatment allocation, and RCTs that randomise all patients but later allocate on the basis of preference in at least some arms (e.g. Wennberg and Rücker designs) are useful designs for R versus P comparisons because they are balanced on known and unknown confounders at baseline. However, all such designs exclude patients who have strong preferences and will not risk randomisation in the first place. In addition, as discussed later, there is confounding for comparisons involving the preference groups in the Wennberg and Rücker designs. Comprehensive cohort designs recruit all or most eligible participants but R versus P and P versus P comparisons are subject to confounding. These are the only designs, however, that provide outcome data on patients who refuse randomisation. Hence it was not possible to say that any one design was superior to any other. All provided evidence for our main questions.

Analysis of recruitment rates

The recruitment rate is the percentage of eligible participants agreeing to enter the RCT. We collected information on three types of recruitment to these RCTs. The first, which applies to all, is the proportion of eligible people agreeing to enter the study (i.e. prior to allocation). The second type, which applies only to comprehensive cohort studies, is the numbers agreeing to be randomised rather than choose treatment. The third type occurs in designs where participants are randomised to random or preference cohorts (Wennberg and Rücker designs). Some may accept the first randomisation in the hope that they will eventually get a choice of treatment but drop out after allocation to the randomised cohort. This is a special type of attrition, which is unique to this design of RCT where patients initially have a 50:50 chance of being able to choose a treatment arm. We describe recruitment rates at all three levels.

It is possible to categorise preference designs according to how much participant preference is constricted. Comprehensive cohort designs offer the greatest choice to eligible participants and thus a hierarchy from highest to lowest would be:

- comprehensive cohort
- zelen design with double randomised consent
- two-stage randomisation (Wennberg and Rücker) designs
- conventional RCT (or Zelen design with no subsequent consent).

We related recruitment to this hierarchy, while conscious that the analysis is threatened by the possibility that particular designs may be used by trialists in different treatments and populations, precisely because of expected problems in recruitment.

Baseline differences between randomised and preference cohorts

We examined baseline differences in socio-demographic and clinical or health-related characteristics by considering the nature and number of baseline comparisons and significant differences reported for each comprehensive cohort study.

Participation in treatment

We suggested in Chapter 1 that the direct effect of preferences might be greater in those cases where an intervention required a significant degree of active participation by patients. For example, psychotherapy could be considered to require higher active participation than taking medication, as it involves a greater investment of time and energy. Hence patients could have greater influence over the benefit that they receive. We tested whether it would be possible to classify RCTs as including 'high' and 'low' participation' arms in which R versus P differences could be compared.

Loss to follow-up

We defined loss to follow-up as the proportion of participants for whom data on the primary outcome were missing. We compared proportions lost to follow-up between randomised and preference arms using a paired *t*-test pairing R and P arms within each study, weighted for sample size.

Analysis of preference effects on outcome: review of analyses conducted in each RCT report

We ascertained (1) the number of RCTs in which the authors had reported an analysis of treatment effects separately for randomised and preference patients (R1 versus R2 and P1 versus P2) and (2) the number that had compared outcomes between randomised and preference groups (R1 versus P1 and R2 versus P2). We summarised the results of these analyses. This was carried out separately for each type of RCT design as applicable. We also examined the proportion of analyses where adjustments had been made for (a) the baseline values of the outcomes (if applicable) and (b) other potential confounding factors.

Similarity of treatment effects between randomised and non-randomised groups does not necessarily

imply that there is no preference effect on outcome, as treatment outcome may be uniformly better or poorer in preference compared with randomised arms. Therefore we focused on treatment-specific comparisons of outcome between randomised and preference groups. We recognise, however, that these comparisons may be biased owing to differences between participants who choose their treatment compared with those who agree to be randomised.

The statistical significance of analyses of preference effects as reported by the trialists was described narratively (see Chapter 6). However, the statistical significance of such effects is a function of sample size, and therefore RCTs reporting the same magnitude of preference effects may report opposite conclusions based on statistical significance alone. In order partially to overcome this problem, outcome data from RCTs were converted into standardised effect sizes, which are independent of sample size.

Analysis of preference effects on outcome: reanalysis using effect sizes

We calculated 'preference effect sizes' by comparing outcome in randomised and preference groups. This analysis was conducted for all RCTs in which there was at least one preference arm, that is the comprehensive cohort and two-stage randomised (Wennberg and Rücker) designs. Whereas summary statistics in systematic reviews of RCTs are based on well-matched groups, this is not necessarily the case for the preference/randomised comparisons in this review. In order to address this issue, baseline differences in outcome variables were also taken into account.

For continuous outcome variables, we calculated treatment-specific effect sizes from the difference in means between randomised and preference groups divided by the pooled standard deviation. For binary outcome variables, log odds ratios comparing randomised and preference groups were calculated for each treatment. For comparison with the other studies, these log odds ratios were converted into approximate effect sizes by dividing by 1.81.[34] For RCTs in which baseline data for the primary outcome were applicable and available, baseline treatment-specific, randomised/preference effect sizes were calculated in the same way. All effect sizes were calculated such that a positive effect indicates a difference in favour of the preference group, whereas a negative effect indicates a difference in favour of the randomised group. Cohen[35] defined effect sizes as small = 0.2, medium = 0.5 and large = 0.8. For the purpose of this study, we created three categories: small effect sizes (0–0.2), medium effect sizes (0.21–0.79) and large effect sizes (≥ 0.8), according to absolute values of effect sizes.

In order to explore whether preference effects varied according to the type of treatments under investigation, we grouped the RCTs according to the disorders and treatments under examination. The groups included termination of pregnancy; treatments for ischaemic heart diseases; treatment of addiction disorders; treatment of depression; treatments for back pain; treatments for children with upper respiratory tract infection (URTI); educational programmes; and chemotherapy for cancer.

In conducting this analysis, we explored whether the degree to which preferences act on outcome depends on their nature and on the nature of the outcomes. For example, physicians' preferences may have a different effect to patients' preferences. Similarly, as noted in Chapter 1, preference effects may vary according to types of outcomes, with greater preference effects associated with subjective outcomes such as self-reported health or treatment satisfaction, compared with more objective outcomes such as survival or clinical events.

Chapter 4

Preferences: a conceptual framework and review of measurement methods

Introduction

In Chapter 1, we introduced the concept of preferences and discussed how preferences could impact on the design, analysis and interpretation of RCTs. One of the aims of the project was to review conceptual and measurement issues, in order to provide a framework to assist in the interpretation of the empirical findings of the main systematic review. The review is divided into two sections:

1. In the first part of this chapter we describe the conceptual framework and present a general model of the development of preferences. This model uses ideas developed from the wider literature on patient decision-making, and also the specific literature on the design of preference studies.
2. The second part of the chapter examines issues in the application of this model in RCTs, specifically the elicitation of preferences during informed consent procedures and the measurement of preferences for use in the analysis of RCTs.

Methods

Gathering information for the conceptual framework differed in form and function from the traditional systematic review of preference effects described in the previous chapter. Unlike the traditional review, the aim was not to provide an exhaustive and comprehensive analysis of all studies on this issue, but rather to examine the wider literature relevant to the concept of preference as applied to decision making about interventions in RCTs. Relevant literature concerning preferences were gathered from key papers already available to the authors, references identified in these key papers and the results of preliminary scoping searches undertaken for the main systematic review (detailed in the previous chapter). However, key processes used in systematic reviews (i.e. explicit and transparent literature-searching strategies, specific inclusion and exclusion criteria and the application of study quality criteria) were not appropriate for the development of a conceptual framework, where issues of logical consistency and utility were paramount. The need for the modification of traditional systematic review methods has been identified in a previous HTA review.[36]

Although a number of disciplines are of potential relevance to issues of preferences in RCTs, the bulk of the literature used in the development of the model derived from economics and psychology. Both are frequently used in relation to the design and analysis of RCTs, and their focus on methods of quantitative measurement is particularly pertinent in the current context. Although there are differences in exact definitions of concepts and theoretical emphases between these disciplines, it will be demonstrated that they are sufficiently similar at a broad descriptive level for the present purposes.

Definition of patient preferences

One dictionary definition of 'prefer' is 'to like something better than another: tend to choose',[37] which highlights the fact that preferences in the context of RCTs concern two key processes:

1. an **evaluation** of an intervention in terms of its desirability or attractiveness
2. a **choice** between alternative interventions based on that evaluation.

The model to be developed in this chapter makes a distinction between the evaluation and related decision-making. For the present purposes, the term 'preference' is restricted to the evaluation of desirability, and is defined as the difference in the perceived desirability of two (or more) interventions.

In addition, the definition of preference in the current context relates to a preference **between alternatives** (i.e. different intervention in an RCT). Therefore, preferences are understood as a relative quality (i.e. the desirability of one intervention compared with another), and cannot be understood in isolation.

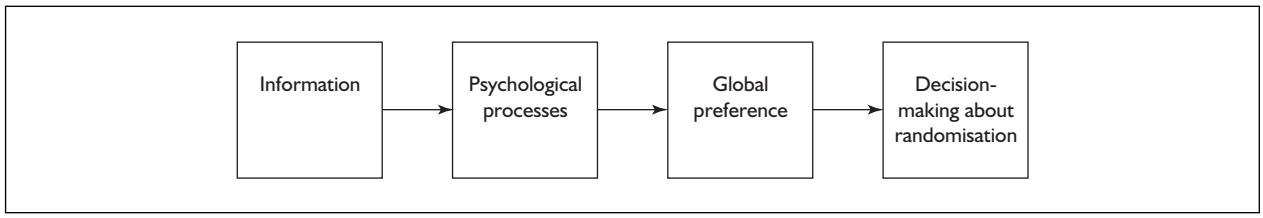

FIGURE 8 *General model of the development of preferences*

Given this broad definition of preferences, how does this link with the wider theoretical literature? In economics, the 'desirability' of an intervention can be understood in terms of the concept of **utility**, which refers to a measure of the satisfaction gained from the consumption of a good or service (such as healthcare).[38] In psychology, a global evaluation such as 'desirability' is best represented by the concept of **attitude**, defined as 'a disposition to respond favourably or unfavourably to an object, person, institution, or event'.[39]

Basing the current model of preferences on the concepts of utility and attitude suggests a number of key attributes. Both utility and attitude are concepts which are **global** and **unidimensional**. That is, both represent an overall evaluation of the desirability of an object, which does not describe the basis of that global evaluation. The psychological processes underlying this global evaluation are the subject of the next section.

In addition, both of these concepts are hypothesised to be **quantifiable**, that is, patients can be described as having a specific 'amount' or as displaying a particular 'strength' of preference. Strength of preference may vary from a slight preference which has little substantive importance, through to large preferences which have a major influence on behaviour. It would be expected that the strength of preference would be a key determinant of the causal mechanisms described in Chapter 1. Methods used to quantify the strength of preference will be discussed later in this chapter.

A model of preferences

Broadly, the model of preferences to be described is a stage model that involves four specific stages. The basic model is shown in *Figure 8*. The model is focused on those aspects of preferences that are most closely related to the conduct of RCTs. Wider influences will be discussed where appropriate, but are excluded from the model in *Figure 8* for the sake of simplicity.

The four stages of the model are as follows:

1. The first stage concerns the source of preferences. In the current context, this concerns **information** received about interventions in an RCT and the wider issues of informed consent procedures in RCTs.
2. The second stage concerns the **psychological processes** by which information about interventions is assimilated.
3. The third stage concerns the initial output of these psychological processes, which is a **global preference** for interventions in an RCT.
4. The fourth stage represents **patient decision-making**, related to whether or not they agree to be randomised. When patients are offered participation in a standard RCT, this concerns the decision whether or not to enter the RCT. In a comprehensive cohort design, it concerns whether patients will be randomised or choose to have a particular treatment.

Each stage is considered in greater detail below. The description above shows the temporal order of the stages, but the detailed description of the stages below follows the order of the theoretical development of the model. Hence the first section below concerns the nature of global preferences, followed by sections on psychological processes, information and informed consent and then the effects of preferences on decision-making about randomisation.

Decision-making processes

What are the processes underlying the development of preferences? Models of decision-making in both economics and psychology are similar in that both are broadly based on the paradigm of the rational, individual decision-maker. A brief description of decision-making models in each discipline is given below.

In economic theory, individuals have different utility functions, and the arguments of those functions represent those attributes of a commodity that contribute to its overall utility.[40] The overall utility related to a commodity depends on the utility associated with each argument, multiplied by the probability (either objective or subjective) of that argument. In decisions about treatments made under conditions of uncertainty, subjective expected utility is the normative model within economics,[38,41] which suggests that treatment preferences will be based on the individual utilities associated with the outcomes of each treatment, multiplied by the probabilities of those outcomes.

Similarly, models of the development of attitudes are generally based on expectancy–value theory, where 'a person's attitude towards an object is related to his beliefs that the object possesses certain attributes (expectancies) and his evaluations of those attributes (values)'.[39] This is broadly analogous to the economic approach described above.

It should be noted that both of these models involve **expectancies about a treatment**, which links with the candidate mechanism for the indirect effect of preferences on outcomes through the placebo effect described in Chapter 1 (*Figure 2*).[42]

The nature of expectancies

One major aspect of utility measurement in health economics is that the **arguments of the utility function** (i.e. the attributes of the treatment that are relevant to treatment decision-making) are generally limited to health status and those effects of treatment that impact on health status (e.g. side-effects). This is because healthcare is seen as something that has no value in use, but only through the benefits derived from it.[43] Therefore, the benefits of use of healthcare are restricted to associated health benefits in terms of mortality and morbidity and the utility associated with healthcare relates to judgements made about the utility of the patient's health state after receiving the intervention.[44,45]

Psychological models of expectancy also involve similar outcome expectancies, but also include expectancies about process,[42,46,47] such as financial cost, travel, discomfort and the invasive nature of the treatment. These process issues may be important *per se*, or patients may trade-off between process and outcome and prefer a less invasive treatment that is less effective. It should be noted that within economics, arguments have also been made for extending the utility function to consider process as well as outcomes, involving approaches to preference measurement such as conjoint analysis[48–50] and willingness to pay.[43,46] Expectancies concerning process issues will differ from outcomes in that the latter are uncertain, whereas there may be no formal uncertainty related to the process of an intervention.

One important process expectancy highlighted by psychological theory is **self-efficacy**.[51,52] This relates to the belief that particular behaviours required to use an intervention are within the capabilities of the individual. Self-efficacy (i.e. the belief that one can produce certain actions) is theoretically distinct from outcome expectancies (i.e. how one values the outcomes of those actions), and there is good evidence that self-efficacy is an important predictor of behaviour, including health-related decision-making.[53] As noted in Chapter 1, self-efficacy expectancies may be of greater relevance in those interventions where the patient has a role to play in compliance (e.g. taking medication) or where the treatment requires even more active involvement or participation with a clinician (e.g. psychological therapy) or is entirely patient directed (e.g. using a self-help treatment). Although the degree of patient involvement is not an easy thing to define or measure, variation in patients' beliefs that they can successfully undertake the behaviours required in the intervention may be an important predictor of their preference for that intervention and their eventual adherence,[54] irrespective of expectancies about outcome.

The nature of values

The second aspect of the expectancy–value calculation is the value that patients place on particular processes and outcomes. Compared with the amount of research on the expectancy aspect of decision-making, values have received relatively little attention. This is partly because they are inherently subjective: 'who knows better than an individual what he or she prefers'.[55]

However, the value aspect of the equation may be a critical determinant of variation in preferences. Although treatments within an RCT are standardised, patient perceptions of the value of different aspects will vary in relation to their own characteristics and experience. Thus, two patients offered the same intervention, and with the same expectancy of outcome and process, may still differ in their overall preferences if they value those processes and outcomes differently.

Information

The previous section described the psychological processes underlying preference development in terms of an expectancy value model, and described how expectancies could relate to outcomes, process attributes and self-efficacy. This section will consider the basis of expectancies about interventions in an RCT, that is, the information about the interventions. Information about interventions may derive from a number of sources (both outside the immediate context of the RCT and the RCT informed consent procedures) and patients may receive differing amounts of information. This information will lead to various expectancies about the process and outcome of an intervention.

A critical distinction in the present context concerns the **validity** of these different expectancies. Some authors have cautioned that a distinction needs to be made between **informed choices**, in which 'patients rely on the estimates of the size of risks and benefits of proposed interventions, as reported in reliable overviews' and **subjective preferences**, in which 'patients ignore the available evidence and prefer to rely on prayer, on a hunch or the advice of friends, relatives or seers for a decision'.[56]

Clearly, there is sufficient evidence of patient (and professional) difficulty in making sense of probabilities and statistics, to provide some support for such a crude distinction.[41] It is possible that preferences in the context of RCTs may be related to invalid prejudices about outcome expectancies related to 'new' treatments.[57] Strict distinctions between 'valid' and 'invalid' may be especially relevant for process expectancies, because treatments have certain objectively defined process attributes and patients' understanding and knowledge of these can be assessed relatively easily and compared against the actual process of treatment.

However, any judgement of the 'validity' of expectancies needs to be qualified. Any distinction between 'informed choices' and 'subjective preferences' does not influence the causal pathways described in Chapter 1. Both direct and indirect preference–treatment interactions are possible even when preferences are based on incorrect information or faulty reasoning, because is it the strength of the preference that may be of importance in determining the preference–intervention interaction, rather than the validity of the expectancies on which it is based. It is therefore useful to distinguish between information and expectancies (which can be described as 'valid' or 'invalid') and resulting preferences (which cannot).

RCTs are normally conducted where there is **clinical equipoise**, that is, where a rational, informed person has no preference between two available treatments[58] and where the choice between two interventions cannot be made on the basis of health outcomes. [It is possible that some trials might be conducted where differences in outcome exist, and are known, and the trial is being conducted to examine differences in costs, but these are likely to be rare. Another example would be differences between treatments arms in the types of outcomes (e.g. one treatment may improve social function more than an alternative, but have less of an impact on health status). Of course, clinical equipoise only relates to average outcomes, and there is the additional possibility that differences in outcomes between treatments might be hypothesised to occur in different types of patients.] However, preferences may still be 'reasonable' or 'rational' or 'appropriate' in the absence of objective evidence of differences in outcomes between interventions, for a number of reasons.

First, as indicated above, clinical equipoise does not take into account differences in the process of treatment, which are a reasonable basis for preferences. Second, patients' outcome expectancies may be based on **subjective** expectancies, such as the belief that a treatment is particularly suited to an individual (an implicit 'subgroup hypothesis'), which might contradict research evidence suggesting that the **average** effect is zero. This may be particularly important with practitioner preferences, as practitioners may perceive that they have particular knowledge or experience that contradicts the research evidence. Third, the expectancy–value model presented in this chapter has two sources for preferences: expectancies and values. The former may be amenable to objective description, but the latter are inherently subjective.[55] Therefore, ensuring that patients have comprehensive information and similar expectancies about an intervention will not ensure that their preferences are the same, if values differ significantly.

Preferences will therefore differ in terms of the nature of the expectancies on which they are based, which has important implications for changing preferences through information. Some preferences may be responsive to information, as factually incorrect expectancies may be relatively

easy to overcome. However, some expectancies may be based on subjective expectancies or values which may be less amenable to change through information.

In summary, the relationships between expectancies, values and preferences is potentially complex, and there are a number of cases where simple distinctions between 'informed choices' and 'subjective preferences' are unhelpful. It may make more sense to distinguish 'informed expectancies', where there is evidence that patients have received sufficient information, clear inaccuracies have been corrected and patients have had time to consider this information in order to make a judgement based on their expectancies and the values they place on them. Any such distinction is obviously complex, and the issue of 'informed expectancies' will be developed further later in the section on 'preference elicitation' (p. 26).

Decision-making about randomisation

The proposed model makes a distinction between patient's preferences (as a global evaluation of the relative desirability of two interventions), and actual decision making about agreement to randomisation (*Figure 9*). Thus patients will make a judgement as to their preference (based on the expectancy–value calculation outlined above), and this preference will **influence** their decision whether to agree to randomisation.

However, preferences for treatments are only one factor relating to the decision to agree to be randomised. For example, in one RCT, patients were asked if they were willing to be randomised to treatment, and 82% agreed to be randomised.[23] However, when completing a questionnaire after randomisation, 80% reported having a preference for treatment. In another RCT, patients reported willingness to 'tolerate' conservative management if access to surgery would still be available at the end of the RCT.[60] Such effects may be predictable from current theory. For example, within the psychological model of attitude, an additional influence on behaviour beyond the expectancy–value calculation is hypothesised to be **subjective norm**.[51] Decisions about interventions are not always made in isolation and are highly likely to be influenced by significant others, especially the patient's clinician. Subjective norms are hypothesised to be a function of two psychological constructs: the beliefs of significant others (e.g. the clinician or a close family member) as to whether the individual should perform that behaviour ('normative beliefs') and the patient's conviction that taking those beliefs into account are important ('motivation to comply'). Such effects may account for cases where preferences and decisions about randomisation are in opposition.

Distinguishing between patient preferences and final decisions is important, because patients may take decisions that conflict with their preferences, and in these cases preferences could still influence outcomes through the direct and indirect effects described in Chapter 1.

The current model proposes that the decision to agree to randomisation can be understood as a second expectancy–value calculation. Broadly, participation in the RCT will provide utility for the patient, which will depend on the strength of their overall preference, multiplied by the likelihood of receiving a particular intervention (usually 50% in a standard two-arm RCT).

An important contextual effect on this calculation concerns the availability of the preferred intervention outside the RCT. If the intervention is available outside the RCT, then the issue is one of potential loss if patients agree to randomisation, as patients must risk losing access to a treatment. If the intervention is not available outside the RCT, then the issue is one of potential gain through participation. Generally, the former is seen as the issue where preferences are most problematic.

It is beyond the scope of the present review to consider the wide variety of influences that may impact on patient or practitioner decisions to take part in an RCT more generally.[61–63] In the present context, it is simply important to note that any influence that potentially causes a patient to make a decision about participation in contradiction of preferences is a problem methodologically, as such patients may demonstrate the preference effects described in Chapter 1.

However, there is an important distinction to make from an ethical standpoint. Some influences on patient decisions about participation in an RCT may reflect influences which can be understood in terms of the expectancy–value framework, in that they may cause patients to make a decision in contradiction to preferences about the interventions, by compensating that loss with additional utility gains from elsewhere. For example, patients may gain utility through

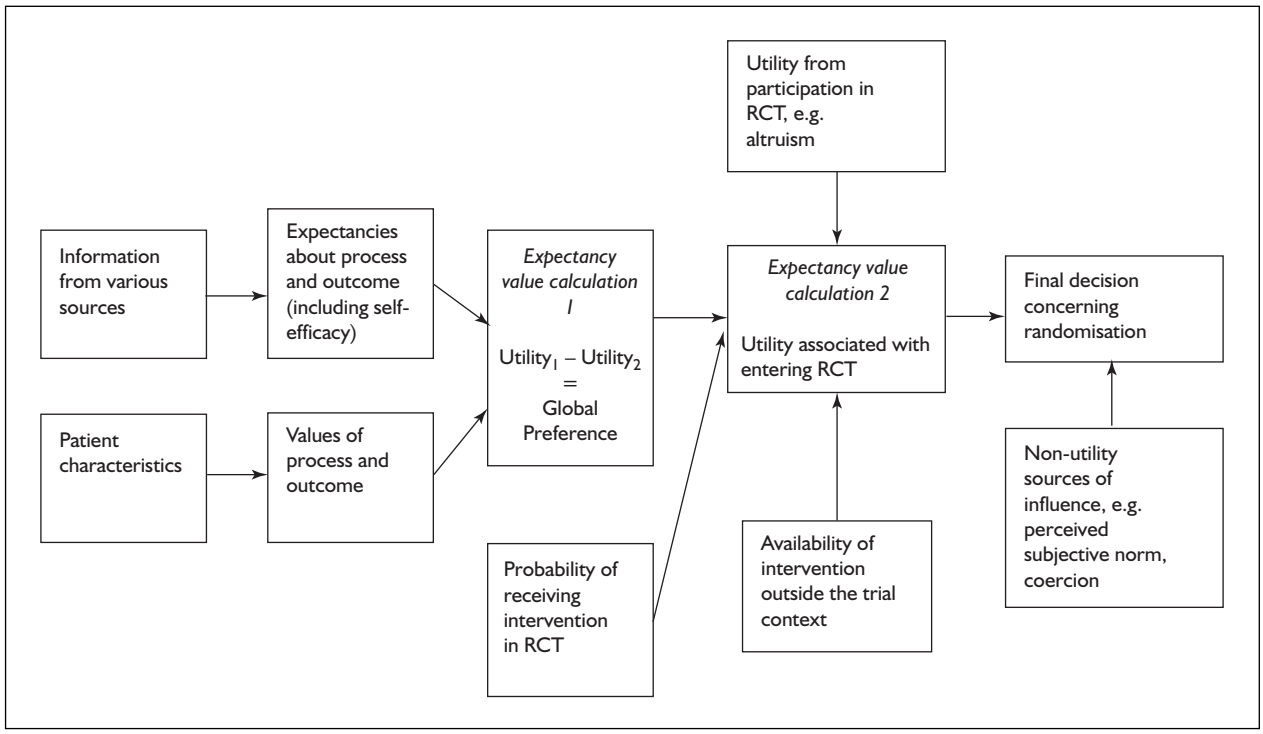

FIGURE 9 *Proposed preliminary model of intervention preference, participation preference and participation decision-making*

demonstrations of altruistic behaviour.[64] In such cases, their behaviour can be seen as informed and ethically appropriate, notwithstanding any methodological issues.

However, if patients' decisions about randomisation are not compensated with other utility gains, this has important implications.[64] From the point of view of RCT recruitment, there is no major problem with patients agreeing to randomisation in contradiction of their preferences, but ethical concerns are highlighted if that represents coercion rather than altruism.

Limitations of the proposed model

Figure 9 shows a more detailed version of the four-stage model, indicating the key factors and processes as outlined in the previous sections.

As noted in the introduction to this chapter, the proposed model is based on generic models of decision-making and judgements within economics and psychology. However, the model does not exhaust the full range of proposed models within those disciplines, nor does it consider the wider literature within other relevant disciplines.

For example, with economics and psychology, alternative decision-making models have been presented, such as prospect theory,[65] regret theory[66] and others,[67] which may have relevance in the current context. For example, regret theory suggests that decision-making is influenced by the probability of experiencing regret associated with a 'wrong' choice, where 'wrong is determined on the basis of actual outcomes rather than on the information available at the time of the decision'.[66] Individuals may trade off certain aspects in order to avoid regret. The current model may therefore require development in order to take into account the findings from alternative models. Particular extensions to the model may be relevant in relation to particular interventions and decision-making contexts.

Second, the models of decision-making developed within economics are generally normative models, i.e. they are concerned with how people **should** make decisions, not how they do. Psychological models also generally have this normative basis, although developments in these models (e.g. the addition of self-efficacy and subjective norm) represent an attempt to model better how people actually make decisions. It is likely that multidisciplinary studies of

TABLE 2 *A typology of patients in RCTs in relation to expectancies, preferences and decision-making*

Type	Informed expectancies?	Preference for intervention?	Agreement to randomise?	Implications
1	No	No	No	Ethical basis of decision problematic, decision inconsistent with preferences
2	No	Yes	No	Ethical basis of decision problematic, decision consistent with preferences
3	No	No	Yes	Ethical basis of decision problematic, decision consistent with preferences
4	No	Yes	Yes	Ethical basis of decision problematic, decision inconsistent with preferences
5	Yes	No	No	Decision inconsistent with preferences
6	Yes	Yes	No	Decision consistent with preferences
7	Yes	No	Yes	Decision consistent with preferences
8	Yes	Yes	Yes	Decision inconsistent with preferences

decision-making using a combination of quantitative and qualitative methods might further the development of a truly descriptive model of how patients actually make decisions in RCTs.

The proposed model in *Figure 9* should not therefore be considered complete or definitive, and is designed to function as a broad framework to assist in understanding patient preferences in the empirical systematic review. Later in this chapter, specific recommendations will be made concerning future research required to develop the model further.

Patient preferences in RCTs

The previous section described a broad model of the development of preferences, which has implications for the application of preferences in RCTs, through two main procedures. The first concerns the **elicitation** of preferences in RCT informed consent procedures, whereas the second concerns the **measurement** of patient preferences for use in later analyses. These issues will be considered from a theoretical perspective and in relation to the empirical data gathered on these issues on the studies included in the systematic review.

Preference elicitation and informed consent procedures

The model developed in the previous section suggests that patients entering an RCT can be characterised along three dimensions, relating to the four stages of the model in *Figure 8*:

1. First, whether the expectancies on which their preferences are based may be considered 'informed', that is, they are based on good-quality information combined with patient values.
2. Second, whether patients have preferences between the alternative interventions, or whether they are in a condition of equipoise.
3. Third, whether they are willing to accept randomisation.

Table 2 lists the eight possible states for an individual patient, based on these three dimensions. Each of these states has different implications for ethics, recruitment to RCTs and interpretation of the study results.

For example, any patient who does not have 'informed' preferences (types 1–4) is problematic from an ethical standpoint. However, if they are in equipoise and agree to randomisation (type 3), they are not problematic from the perspective of those concerned solely about recruitment and interpretation issues. Among those patients who have 'informed expectancies' (types 5–8), those who are not in equipoise but agree to randomisation are potentially problematic, both ethically and in terms of interpretation, but do not cause difficulties for recruitment.

The implications of the model in *Figure 9* and the typology in *Table 2* for informed consent procedures will now be considered.

Informed consent and 'informed expectancies'

As noted above, any informed consent procedure should ensure that patients at least meet the criteria for 'informed expectancies', as broadly defined earlier. However, the exact procedures required to ensure this are unclear.

The simplest case relates to those patients who enter RCTs with inaccurate expectancies (i.e. ideas about the process of treatment that do not reflect the actual intervention that they will receive). In these cases, such problems may be overcome by provision of more detailed information during informed consent procedures, with a focus on ensuring that the information provided meets current ethical requirements, is understandable to patients and is unbiased. A recent study described how ostensibly neutral descriptions (i.e. 'watchful waiting') could be interpreted negatively by patients in terms of an expectancy (in this case, of relative neglect),[68] and pilot qualitative research may assist by identifying the nature of expectancies which may be inaccurate.[69] Providing information that is 'balanced' may be easier in situations where two 'active' interventions are being compared, as compared with situations where a specific intervention is being compared with 'usual care' or 'no intervention', as there may be a general expectancy that any new intervention is more likely to be worthwhile than 'usual care'.[60,70] However, at least one author has cautioned against excessively positive expectancies being engendered in placebo control groups.[71]

The major technical limitation in such cases may relate to concerns about the ability of patients to comprehend and remember information relating to interventions.[72,73] There may be limits to the passive use of written information[74] and more innovative methods may be required. More complex approaches to the provision of information (such as multimedia or interactive presentations) might allow more effective provision of information concerning the exact process of an intervention, especially in interventions which are novel or where pre-RCT work has demonstrated a large gap between process expectancies and reality. In the psychological therapy RCT literature, significant effort was expended in developing placebo psychological therapy interventions, and it was deemed necessary to check that both the active therapy and the placebo were perceived by patients as equally potentially efficacious (i.e. as having equal 'treatment credibility').[74–76] Therefore, an additional step might involve checking the effects of information by assessing patient expectancies **after** the provision of information, to assess the impact of information provision and possibly encourage further discussion and intervention where written information has been ineffective.

However, there may be a case for a more fundamental change. A useful distinction from the wider literature on decision-making is between patient education materials (which seek to provide information and increase knowledge) and decision support aids (which seek to provide information, clarify values and augment skills in decision-making).[77]

Decision support in the context of preferences in RCTs might involve a specific, standardised preference elicitation interview with a researcher, which involves systematic consideration of the two decision-making processes described previously (i.e. stages 2 and 4 in the model in *Figure 8*). That is, in relation to stage 2, there would be a systematic assessment of the utility associated with each intervention, in which the expectancies of individual patients are assessed, checked for validity where relevant and the researcher assists the patients to integrate expectancies with their values explicitly. In addition, there would be a specific check that their decision about randomisation (stage 4) reflects these preferences, and has not been vulnerable to coercion or other influences. Such an approach begins to look like a formal decision analysis,[78,79] and there is no **theoretical** reason why such decision analyses could not be part of the consent procedure, although the practical barriers are significant. In such cases, only patients whose preferences for the interventions are identical, or at least highly similar within reasonable bounds,[58,80] would be judged as eligible for randomisation.

It is possible that elements of these more complex negotiations occur in some RCTs, but it seems unlikely that they are done systematically or in a standardised way. Core competencies required for shared decision-making have been defined and might be relevant for clinicians and researchers within RCTs.[81]

These considerations highlight the tension between ethical and administrative pressures in RCTs. There may be ethical concerns about modifications of informed consent procedures (e.g. specially designed information sheets) which have the **aim** (implicit or explicit) of increasing participation rates, because the assumption is that

participation in the RCT is the optimal decision for the patient and low rates of participation are suboptimal *per se*, as opposed to simply problematic for researchers.[80,82] The definition of a 'good' outcome in terms of the decision-making problems faced by patients in RCTs is complex[77] and there is no reason why the perspectives of clinicians, patients and researchers should agree. A recent RCT of a controversial intervention used qualitative research to examine the consent procedures, and iterative methods were used to refine further the information sources used during consent procedures and to train researchers involved in allocating patients.[68,83] This improved the percentage of patients willing to be randomised over time. The authors themselves raised the issue as to whether improved randomisation rates might reflect 'better' consent procedures and the effect of overcoming incorrect expectancies or increasing coercion. This relates to the model of preferences presented in this chapter, which highlights the fact that patients' final decision about randomisation may be related to something other than their actual preferences e.g. coercion may relate to subjectively perceived norms communicated (perhaps unconsciously) by recruitment staff. For this reason, the importance of encouraging expressions of equipoise in research staff has been highlighted.[71,77]

However, notwithstanding any practical difficulties associated with more complex informed consent procedures, adoption of more complex and involved procedures may mean that patients without pre-existing preferences develop preferences during this procedure. Although much of the relevant research has been conducted in laboratory settings and may not be generalisable, some research has indicated that preferences, and the values on which they are based, may not have an independent and static existence, but may be 'constructed' by the nature of their elicitation,[55,84,85] and vulnerable to so-called 'framing effects' relating to issues such as the order or context of presentation of information.[68] Although some have assumed that providing information would increase the proportion of well-informed people who have no strong preferences and would therefore be eligible to be randomised,[86] more complex preference elicitation techniques might have the untoward effect of **reducing** the sample willing to be randomised,[80,82] and there is some evidence from RCTs of informed consent procedures that there is an optimal level of information, and that increasing amounts of information can reduce recruitment rates.[87] Furthermore, the external validity of the study might be reduced. Patients in the RCT may have more information, and hence different preferences, from those who receive the intervention in routine care settings, where less information is given. Therefore, the results of the RCT might not generalise to routine care settings where the information provided is less comprehensive or where patients enter treatment with inaccurate expectancies. It should be noted that, although framing issues and other potential effects raise a tension between the methodological and ethical issues in an RCT, it is expected that ethical issues will generally take precedence.

Preference elicitation in the RCTs included in the review

Examination of the preference elicitation procedures in the RCTs included in the review indicated that 18 RCTs made no reference to how participants' preferences were elicited, three made a statement that preferences were elicited but gave no information on methods, 12 provided eligible participants with written descriptions of treatments and one used a video to explore the treatment options, which included actual demonstrations of therapy.

Preference measurement

The function of preference measurement

Preference measurement is required in RCTs that randomise all patients, in order to examine preference–intervention interactions in the analysis.[88,89] In some comprehensive cohort designs in the review, there were no quantitative measurements of preference. Rather, patients were asked to indicate whether they had a 'strong' preference for treatment, and then either randomised or allocated to their preferred intervention accordingly, that is, the study protocol involved preference elicitation only. Such a measure of decision-making behaviour might be seen as advantageous, in that the final treatment decision may be seen as the 'gold-standard' measure of preference, in comparison with measures based on self-reported attitudes.

However, there are two problems with this approach. First, as highlighted by the model presented earlier, preferences and decision-making are not always in exact agreement and some patients may agree to be randomised even if they have strong preferences for a particular

treatment. These patients would still be expected to suffer resentful demoralisation if randomised to their non-preferred treatment. Without preference measurement, these patients cannot be identified and dealt with in the analysis.

Second, having no measure of the strength of preference limits analyses that examine the relationship between preferences, baseline characteristics and outcome, in that analyses can only examine the general effect of having a preference, rather than a continuous measure of preference strength. Because information will be lost in using crude categories such as 'preference' and 'no preference', evidence for interesting associations may be lost. Designs without measurement of preference are also unable to empirically define 'strong preference'. Such an empirical definition may be useful in standardising preferences across groups and also be of use in making decisions about the likely impact of preferences. For example, all patients in an RCT might have a preference for one intervention over another but, if that preference is very slight, it might be expected to have few implications either ethically or methodologically.

Finally, measurement of preferences within an RCT may allow some analysis of factors and processes involved in patient preferences, in order to extend the general model in *Figure 9*.

Techniques for preference measurement

As quantitative sciences, both economic and psychological approaches to preferences provide methods of measurement that are useful in understanding the effects of preferences in RCTs. It was beyond the scope of the present review to conduct a comprehensive assessment of the different methods available to measure preferences. However, two previous HTA reviews[90,91] have already considered these techniques in terms of several key characteristics, such as the strength of evidence of validity and reliability, their feasibility (i.e. cost, time), theoretical basis and their properties.

Five approaches are of relevance to preference measurement, which represent four specific techniques from economics and one general approach from psychology.

1. Rating scales. Respondents provide a single global rating of the desirability of a particular health state on a Likert or visual analogue scale, relative to upper and lower extremes.
2. Standard gamble and time trade-off methods. Standard gamble approaches require that respondents choose between a lifetime in a certain health state or a gamble between different health states, whereas time trade-off requires respondents to choose between living for a period in less than perfect health, as opposed to a shorter period in perfect health.[90]
3. Willingness to pay methods. These methods are based on the idea that maximum willingness to pay for a commodity such as healthcare is a measure of preference for that commodity. Values can be elicited by a number of methods, such as open-ended approaches (simply asking for willingness to pay), payment cards and bidding techniques.[90,91]
4. Conjoint analysis. This approach measures the value placed on attributes of a commodity by requiring individuals to choose between different scenarios, where in each scenario the commodity in question has varying levels of different attributes.
5. Attitude measurement and other psychometric approaches. Attitude scales measure the global evaluation of an object through methods such as Likert rating scales or the semantic differential.[92] Additional psychological constructs such as self-efficacy and subjective norm can also be assessed through specially designed scales.[51]

The utility of different measures of preferences in RCTs

The model developed in the first part of this chapter suggested that expectancies about both process and outcome are of potential relevance to preferences. Expectancies concerning outcomes can be problematic within the context of clinical equipoise, therefore techniques such as rating scales, the standard gamble methods and time trade-off methods are of limited utility in the present context. Rating scales could be extended to cover issues of process, although in such cases there would be little to distinguish them from the general methods used in attitude measurement.

Conjoint analysis can examine process issues relating to interventions and can indicate the way in which patients 'weight' different criteria relating to treatments, such as different process criteria or process versus outcome.[90] This technique may therefore be of general use in developing understanding of the way in which patient expectancies are integrated with values in order to create their overall preference (*Figure 9*). However,

because conjoint analysis requires comparisons of different treatment 'scenarios', each involving different attributes, this technique may be of less use within the specific context of RCTs, which compare two standardised interventions with set levels of particular attributes.

Willingness to pay methods are perhaps the most useful of economic approaches in the present context. Process issues relating to interventions can be incorporated into measurement, and estimating willingness to pay for particular aspects of an intervention might provide some insight into patients' decision-making processes underlying their global evaluation. Willingness to pay is also simpler for patients than techniques such as the standard gamble, and the 'metric' being used to determine overall preference is relatively easy to understand, although ironically the very salience of the notion of paying for treatments may lead to 'protest' responses from respondents used to receiving healthcare free at the point of use.[43]

Attitude measurement and other psychometric techniques are the most flexible approach for preference measurement in RCTs, in that almost any issue relating to treatment preferences and decision-making can be scaled and quantified using standard psychometric techniques, and methods for the evaluation of constructs such as self-efficacy are already available.

What measures of preference measurement have been used in preference studies?

Three RCTs in the systematic review measured participants' global preferences, using methods such as a visual analogue or Likert scale.[93–95] One RCT measured participants' initial credibility of treatments using a four-item scale[96] and one measured outcome expectancy on a five-point scale.[88]

Such single-item scales have the advantage of simplicity and ease of use and may be considered appropriate for a global evaluation, but have a number of problems. Single-item scales are of low reliability and can only provide either a measure of global preference or a measure of one aspect of the constructs likely to be related to preference (e.g. outcome expectancy). Although they may increase respondent burden, multiple item scales avoid the problems associated with single-item scales in terms of reliability and allow consideration of a number of different constructs that might influence overall global evaluation of preferences.

All the RCTs included in the review measured preferences at baseline only. There is also an argument for measuring preferences using such scales at regular intervals throughout the study to assess the stability of expectancies and preferences and their interactions with blinding.[69]

Chapter summary

1. Preferences are broadly defined as the evaluation of the relative desirability of two (or more) interventions within an RCT.
2. The concept of preference can be understood in terms of the concept of utility in economics and attitude in psychology.
3. Literature in psychology and economics suggests that preferences are broadly based on (a) expectancies concerning the process and outcomes associated with the intervention (b) the perceived value placed on those outcomes and processes.
4. The development of preferences and their influence on decision-making can be described in terms of a four-stage model.
5. The model proposed is presented as a general framework for conceptualising preferences in RCTs, not a definitive model. Further theoretical and empirical work is required to extend and test each stage of the proposed model.
   (a) Future research could usefully examine the amount and source of information available to patients about interventions in RCTs, with special emphasis on the relationship between sources inside and outside the RCT context. Qualitative research undertaken as part of ongoing RCTs might be especially useful.
   (b) Additional research might examine the processes by which this information leads to preferences, which might develop or extend the proposed expectancy–value framework. A key question relates to the type of expectancies that enter into decision-making and the way in which different expectancies are valued by patients. Conjoint analysis may be especially useful in this regard.
6. RCTs differ in the amount of information provided to patients and the complexity of the techniques used to provide that information and to investigate patient preferences.

7. Where possible, RCT reports should contain information on the process by which preferences were elicited and the amount and nature of the information provided. Preference effects may differ in RCTs using different procedures, because of changes in the direction or strength of preferences associated with these procedures.
8. Comparisons of the feasibility and effectiveness of different informed consent procedures might be useful. Useful outcomes might include patient knowledge and satisfaction and recruitment rates.
9. A number of different methods are available for measuring preferences, but some are less applicable to preference measurement in RCTs. Willingness to pay methods and generic attitude measurement techniques may be most applicable in the current context.
10. Quantitative measures of preferences are largely absent from RCTs or restricted to single-item measures of global preferences or outcome expectancies alone. Multi-item measurements of preferences examining both outcome and process issues may be a useful addition to future RCTs.
11. The development of common approaches to the measurement of preferences in RCTs might increase comparability across RCTs. Although variation in interventions may limit the scope for standardisation, fruitful approaches might involve agreement over the scope of preference measurement, common item stems and standardised analysis.
12. Comparison of the results of preference measurement using different types of attitude scales (e.g. Likert scales, semantic differential) and between attitude scales and alternative approaches such as willingness to pay would be useful.
13. RCTs should measure preference quantitatively, even in comprehensive cohort RCTs where patients can choose their treatments. The standardised measurement of preferences within all RCTs (and analysis of the interaction with outcome) would allow the rapid development of a significant evidence base concerning patient preferences, albeit in relation to a single preference design.

Chapter 5

Results of the search strategy for systematic review

In this chapter we present the results of our search strategy and provide a brief account of each the studies included in the final systematic review.

Results of the search strategy

The search identified 10,023 citations, of which 37 papers were retrieved for inclusion (*Figure 10*). Many papers were not relevant and therefore excluded, mainly because this was not a narrow field focusing on one particular condition or treatment but one which required a broad search strategy. A search of the reference lists of these 37 papers identified seven more, making a total 44 that were included in the review. This covered 34 RCTs.

We wrote to all authors for information on other publications on the RCT identified and to obtain missing data where appropriate. Only one author provided information that assisted the review.[97]

The type of participant was extremely varied (Appendix 1). Four RCTs involved only children[98–101] and eight only women.[102–108,115,116] Twenty-eight took place in secondary care and five in primary care and one was an experimental study conducted on psychology students in a laboratory (Appendix 1).

The nature of the disorders and interventions involved in the RCTs also varied widely

- Four RCTs concerned the management of depression[26,27,96,109–112] and four concerned cancer therapy.[113–116]
- Three involved treatments for anxiety states, namely phobias, panic and/or agoraphobia[93,95,117]
- There were three on treatments for back pain.[118–120]
- Three involved surgical treatments for coronary artery disease.[121–125]
- There were two each on termination of pregnancy,[102,103,107,126] use of anaesthesia for procedures,[106,127] management of drug addiction[128,129] and management of recurrent otitis media.[97,99,101]
- There was one study on each of the following: alcohol addiction,[130] use of invasive investigations in the first trimester of pregnancy,[104] management of recurrent sore throat,[98] management of young children with cerebral palsy,[100] management of heavy menstrual bleeding,[105] prevention of cardiovascular diseases through educational programmes,[108] educational packages for the management of diabetes,[131] management of chronic pain[132] and obesity.[94]

Study designs

- Twenty-five RCTs employed comprehensive cohort designs. Three of these were a slight modification of a standard comprehensive cohort study[121–125] in which participants declining randomisation (because they or their physicians refused) were given the standard treatment of their choice and followed as 'registry' patients.
- Four studies used a two-stage randomised clinical design (Wennberg and Rücker design).
- Two measured preferences and/or credibility of treatments at baseline within a standard RCT.
- The remaining three RCTs involved unusual designs. One RCT entered patients according to preferences and crossed over allocation during the study[93] and two allocated participants to either their preferred or non-preferred treatment arms (Appendix 1).[94,95]

Description of RCTs identified by the search

In this section we describe the RCTs identified by the search according to their study design and report on the comparisons **conducted by the authors** (Appendix 6).

Comprehensive cohort studies

In the studies identified under this category, eligible participants who agreed were randomised to treatments (e.g. R1 and R2), whereas those who refused because they had strong preferences for one or other intervention were able to enter the arm of their choice (e.g. P1 and P2).

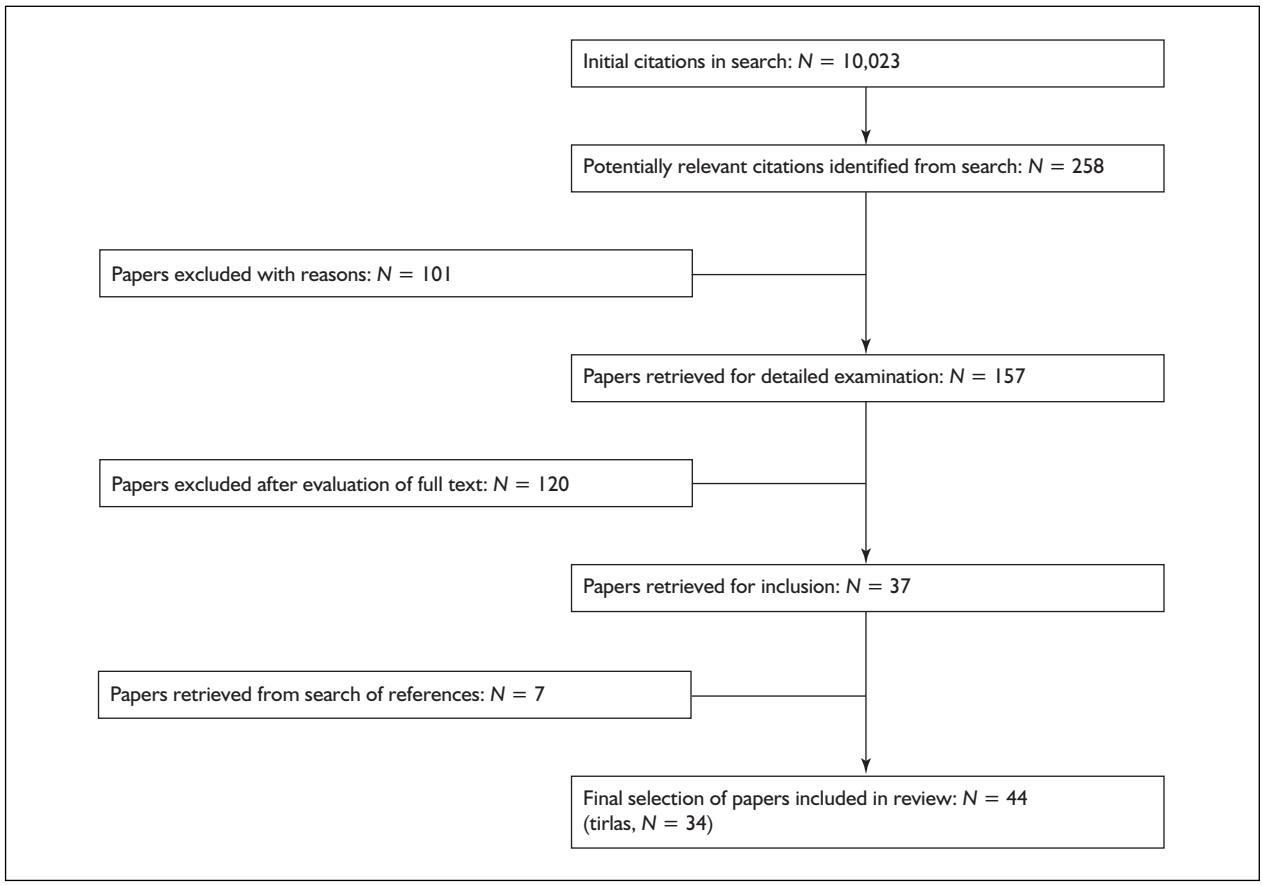

FIGURE 10 *Selection process and results*

***Ashok* et al., *2002*[107]**

Women requiring abortion at 10–13 weeks gestation were asked to take part in an RCT comparing medical termination (using mifepristone and misoprostol) with surgical termination of pregnancy. The researchers did not identify a primary outcome but of the outcomes listed, we identified the number of days of bleeding measured 2–3 days and 8 weeks after surgery of primary interest. Significant differences were reported between the randomised groups (R1 vs R2; surgical termination was associated with fewer days of bleeding). Although the comparison between preference arms (P1 vs P2) was similar in direction and magnitude, it was not statistically significant, possibly because of smaller numbers in this cohort.

***Bain* et al., *2001*[106]**

Women with dysfunctional uterine bleeding were invited to participate in an RCT that compared local and general anaesthesia for microwave ablation of the endometrium. All outcomes reported were related to participants' assessment of the treatment. The researchers failed to identify a primary outcome but we selected 'women's acceptability of the treatment' soon after surgery, as measured by a semantic differential scale. No significant difference was found for any of the 12 subscales of the primary outcome measure between the two arms of the preference cohort (P1 vs P2), but significant differences were reported for two of the 12 subscales (less attractiveness and less ease of anaesthetic procedure for local anaesthesia compared with general anaesthesia) between the randomised arms (R1 vs R2).

***Bakker* et al., *2000*[117]**

This four-arm RCT randomised patients with panic disorder, with or without associated agoraphobia, to one of three drug therapies or cognitive therapy. Potential participants who refused randomisation but expressed a strong preference for cognitive therapy were offered this treatment. The researchers did not identify a primary outcome but of the several outcomes reported in the paper, we identified frequency of panic attacks assessed retrospectively at baseline and at 12 weeks as primary. There were no significant differences in outcome between those randomised to cognitive therapy and those who chose it (R_1 vs P_1).

***Bedi* et al., *2000;*[109] *Chilvers* et al., *2001*[110]**
People consulting with depression in general practice were offered counselling (brief psychotherapy) or antidepressants. The primary outcome identified by the researchers was a measure of depression (Beck Depression Inventory), assessed at baseline and 8 weeks and 1 year after the delivery of the intervention. Unadjusted and adjusted comparisons of treatments between the randomised arms (R1 vs R2) were not significant at 8 weeks and 1 year. When randomised and preference patients were combined [(R1 + P1) vs (R2 + P2)] there were also no differences in outcome at 8 weeks or 1 year between antidepressants and counselling. Comparisons of outcomes between the randomised and preference cohorts (R1 vs P1 and R2 v P2) for each intervention were not significant at 8 weeks, but after 1 year, patients choosing counselling (P1) did significantly better than those randomised to counselling (R1 vs P1).

***Detre* et al., *1999*[121]**
People with diabetes and coronary heart diseases underwent coronary angioplasty or coronary artery bypass surgery as part of the 'BARI' trial. Eligible patients who were not randomised by choice entered an observational registry. The primary outcomes identified by the authors were death from any cause and death from cardiac causes measured 5 years after treatment. Comparisons of cardiac and all-cause mortality were made separately between the two randomised and the two preference arms. Analyses were adjusted for clinical, angiographic and QoL factors. The randomised comparison (R1 vs R2) revealed significantly lower mortality from all causes and cardiac causes in those who had coronary artery bypass surgery but these differences were weaker and not significant in the preference cohort (P1 vs P2).

***Gossop* et al., *1986*[128]**
People addicted to opiates were treated either as outpatients (counselling plus oral methadone) for 8 weeks or as inpatients (ward-based care plus oral methadone) for 3 weeks. Abstinence from opiates at discharge was the primary outcome. Withdrawal was more likely in those receiving inpatient treatments [(R1 + P1) vs (R2 + P2)]. However, no significant differences in outcome were found between randomised and preference arms for each treatment (R1 vs P1 and R2 vs P2) or irrespective of treatment (R vs P).

***Helsing* et al., *1998*[113]**
People with non-small cell lung cancer were entered into an RCT of chemotherapy or best supportive care. The primary outcome identified by the researchers was a global measure of QoL assessed at baseline and 4, 8, 12, 16, 20 and 24 weeks. The analysis was restricted to comparisons between the two randomised arms (R1 vs R2). QoL tended to improve more in the chemotherapy group compared with best supportive care. This was significant at the 20-week follow-up point.

***Henshaw* et al., *1993, 1994*[102,103] *Howie 1997*[126]**
Women requiring termination of pregnancy underwent either a medical (mifepristone and gemeprost) or surgical (using vacuum aspiration) procedure. The researchers identified two primary outcomes assessed 2 weeks after the treatments were received: (1) the procedure women would select for future termination of pregnancy and (2) the women's perception of the procedure measured from good to bad on a semantic differential scale. Other primary outcomes identified by the researchers were anxiety (anxiety subscale of the Hospital Anxiety and Depression scale), and self-esteem. Anxiety and self-esteem were assessed at baseline and 16 days whereas option for the future was also assessed 2 years after the intervention was delivered. The first two primary outcomes were compared separately between the two arms of the randomised (R1 vs R2) and the preference cohorts (P1 vs P2). Women randomised to a medical termination rated it worse than the surgical procedure (R1 v R2). They were also less likely to say that they would opt for the medical procedure in the future and this difference in opinion held at follow-up 2 years later. These differences were not observed between the preference arms (P1 vs P2). In an unusual analysis, change in anxiety or self-esteem was compared across the four RCT arms (R1 vs R2 vs P1 vs P2) and no significant differences were found.

***Kendrick* et al., *2001*[119]**
Patients in general practice with back pain were offered either a lumbar spine X-ray or no radiography in addition to standard care. The primary outcomes identified by the researchers were quantitative measures of pain and disability 3 and 9 months after the intervention. Participants randomised to radiography had higher disability scores at 3 but not at 9 months than those with no radiography (R1 vs R2). No differences occurred in pain scores. Patients who chose radiography had greater disability at baseline and at 3 but not 9 months than those randomised to radiography (P1 vs R1). No differences in pain were found. No

differences were reported between randomised and preference patients receiving no radiography (R2 vs P2).

***Kerry* et al., *2000*[118]**
Patients in general practice with low back pain were offered either lumbar spine X-rays with standard treatment or only standard treatment. The primary outcomes identified by the researchers were a quantitative measure of disability and number of consultations for back pain at baseline and 6 weeks and 1 year after the intervention. No differences were observed in disability scores or consultation rates at 6 weeks or 1 year between the two arms of the randomised cohort (R1 vs R2). Preference patients choosing radiography had greater disability at 1 year (but not 6 weeks) and were consulting more frequently by 6 weeks and 1 year (P1 vs P2). The latter finding held after adjustment for age, sex and duration of back pain at the time of recruitment to the study.

***SB King* et al., *1994, 1995, 1997*[122–124]**
Patients with multivessel coronary artery disease were treated either with percutaneous, transluminal coronary angioplasty or coronary artery bypass graft surgery ('EAST' trial). Where participants or their physicians refused randomisation, participants made a choice based on their physicians' advice and were followed as 'registry' patients. All-cause mortality and the occurrence of angina were primary outcomes. These outcomes were measured at 6, 12, 18, 24, 30 and 36 months after treatment for randomised but only at 12, 24 and 36 months in preference patients. In a somewhat complex analysis adjusted for clinical and angiographic variables, no differences in survival were found between preference arms (P1 vs P2). However, all-cause mortality was lower in the preference group compared to the randomised group in the CABG patients (R1 vs P1) and in a treatment-pooled analysis (R vs P). Comparisons between R1 and R2 were reported elsewhere. There was no significant difference between CABG and PTCA in terms of occurrence of the primary endpoint.

***M King* et al., *2000*[26] *Ward 2000*[27]**
Patients with depression or mixed anxiety and depression in general practice were randomised to cognitive behavioural therapy, non-directive counselling or usual general practice care. Participants who refused randomisation were able to choose an RCT arm but all but two participants chose cognitive behaviour therapy or non-directive counselling. The researchers identified a measure of depression (Beck Depression Inventory score) as a primary outcome. This was assessed at baseline and 4 and 12 months later. There were no significant differences between the two psychological therapies at either follow-up point in the randomised (R1 vs R2) or preference cohorts (P1 vs P2). There were, however, significant benefits after 4 months for those randomised to either psychological therapy compared with those randomised to usual general practice care (R1 vs R2 vs R3), but these effects were lost by 12 months.

***McKay* et al., *1995*[130]**
Patients addicted to alcohol attending two clinics in the USA were invited to take part in an RCT comparing day hospital with inpatient rehabilitation. The authors did not identify a primary outcome and we chose the number of drinking days in the preceding 30 days as the one of primary interest. This was assessed at baseline and 3, 6 and 12 months after delivery of the intervention. No significant differences were found in drinking days between those in the randomised and preference cohorts (R vs P) (adjusted for the intervention they received). Nor was there any difference between the two treatment types [(R1 + P1) vs (R2 + P2)], adjusted for treatment assignment.

***Mc Kay* et al., *1998*[129]**
People addicted to cocaine underwent 28 days of rehabilitation treatment as either outpatients or inpatients. In the absence of a primary outcome identified by the authors, we chose the number of days of cocaine use as the outcome of primary interest. This was assessed at baseline and 3, 6 and 12 months after treatment. Although participants in the randomised cohort were more likely to report higher usage of cocaine at baseline, this improved to a significantly greater extent than for those in the preference groups at 3 and 6 months (R vs P). At 12 months, however, both groups were very similar in terms of cocaine use. No comparisons were made between within randomised (R1 vs R2) and preference cohorts (P1 vs P2).

***Melchart* et al., *2002*[127]**
People undergoing gastroscopy in district hospitals were offered either acupuncture or intravenous midazolam for sedation. The primary outcome identified by the researchers was the participants' perception (recorded on a visual analogue scale) of the examination immediately after treatment. No significant differences were observed between the two treatment groups, irrespective of method

of allocation to the treatment (R1 + P1) vs (R2 + P2)].

***Nicolaides* et al., *1994*[104]**
This RCT compared chorionic villus sampling with early amniocentesis for foetal karyotyping. The primary outcome identified by the authors was foetal loss (total, induced and spontaneous). Results were difficult to interpret because of discrepancies in the published tables. Nevertheless, it appeared that spontaneous foetal loss was significantly greater in women randomised to early amniocentesis than in those randomised to chorionic villus sampling (R1 vs R2), but no significant differences were observed in outcomes between the two preference arms (P1 vs P2) for any of these outcomes. There were no significant differences in terms of induced or total foetal loss.

***Olschewski* et al., *1992*[125]**
This study (referred to as the CASS RCT) randomised people with mild and moderate, stable angina and a documented history of myocardial infarction to coronary artery bypass surgery or medical care. Those refusing randomisation either chose their own treatment or had their physician decide on the most appropriate treatment. The authors identified all-cause mortality at 5 years as the primary outcome. The non-randomised group had more severe coronary artery disease at baseline than the randomised group. No other difference was noted at baseline. At 5 years there was no difference in overall survival between those randomised to or those choosing medical (R1 vs P1) or surgical treatments (R2 vs P2).

***Paradise* et al., *1984*[98]**
Children with recurrent throat infections attending hospital underwent a surgical intervention (tonsillectomy with or without adenoidectomy) or standard non-surgical care. The researchers did not identify a primary outcome, hence we selected the number of episodes of throat infection per participant measured at 1, 2 and 3 years after the intervention as the outcome of primary interest. Significant differences in outcome were observed within both the randomised and preference cohorts at 1 and 2 years. Those receiving surgery had fewer episodes of throat infection than those in the control group (R1 vs R2 or P1 vs P2). However, after 3 years this difference remained significant only in the preference cohort (P1 vs P2).

***Paradise* et al., *1990*[99]**
Children with recurrent otitis media were offered either a surgical intervention (adenoidectomy) or standard treatment. In the absence of an identified primary outcome, we chose the number of episodes of, and proportion of time with, otitis media as the primary outcomes after 1, 2 and 3 years. Comparisons of outcomes at 1 and 2 years between the two arms of the randomised cohort (R1 vs R2) indicated significantly fewer days with otitis media in the group that received adenoidectomy, but this difference was somewhat smaller and not significant between the two arms of the preference cohort (P1 vs P2). Those randomised to adenoidectomy also reported significantly fewer episodes of otitis media at 2 years than those randomised to the control group (R1 vs R2). This difference was similar but non-significant in those choosing their treatment (P1 vs P2).

***Reddihough* et al., *1998*[100]**
Young children with cerebral palsy were offered either conducive education (an educational programme aimed at developing motor and cognitive skills) or traditional neuro-developmental programmes in which there was no emphasis on motor development. As the authors did not identify a primary outcome, we selected an assessment of cognitive (Vaple Assessment Battery) and gross motor function at baseline and 6 months as the primary outcomes. Improvements in cognitive function did not differ according to type of programme apart from for gross motor function in the preference cohort: the control group improved more (P1 vs P2).

***Reidl* et al., *2001*[114]**
People with carcinoma of the lower end of the oesophagus were offered selective bowel decontamination chemotherapy or intravenous antibiotics. The authors did not identify a primary outcome and hence we opted for the occurrence of pneumonia (a post-treatment complication) as the main outcome of interest. The numbers of patients in each arm were very small, and no significant differences were found between randomised arms (R1 vs R2) or in all patients receiving each treatment [(R1 + P1) vs (R2 + P2)].

***Rovers* et al., *2000*,[97] *2001*[101]**
Infants with bilateral otitis media with effusions were offered insertion of ventilation tubes or no treatment (watchful waiting). In the absence of the authors indicating a primary outcome, we selected the occurrence of otitis media and the total time spent with effusion by each child by 3, 6, 9 and 12 months. No significant differences were observed between the randomised and preference

group outcomes (R1 vs P1 and R2 vs P2) analysed separately for each of the treatments offered at 3, 6, 9 and 12 months.

***Schumacher* et al.;[115] *Schmoor* et al., *1996*[116]**
This study (the German Breast Cancer Group Study) involved three RCTs, of which only two were finally included. The RCT that we excluded was conducted in early breast cancer (node-negative patients with tumours <2 cm in diameter). Women were offered either mastectomy or lumpectomy with breast preservation. All patients received radiotherapy of the other breast. Only 6% of the women in this RCT were willing to accept randomisation and the authors did not analyse the results for the randomised cohort. Hence a direct comparison between the randomised and preference cohorts was not possible.

The second and third RCTs were included in the review. These involved women with node-positive breast cancer who had already been treated by mastectomy. In one RCT a 2 × 2 factorial design compared three and six cycles of combined chemotherapy for breast cancer and also investigated the effects of endocrine treatment (tamoxifen). The other investigated the effect of local radiotherapy in addition to six cycles of chemotherapy. The primary outcome identified by the authors was recurrence and all-cause mortality at 5 years. No differences in clinical prognostic factors were reported between the combined randomised and preference cohorts (R vs P). No significant differences in outcome (adjusted for these factors) were found in the randomised (R1 vs R2) or preference cohorts (P1 vs P2).

***Williams* et al., *1999*[132]**
Chronic pain sufferers were offered outpatient or inpatient cognitive behaviour therapy. The primary outcome was score on the Sickness Impact Profile. Comparisons within the randomised (R1 vs R2) and preference cohorts (P1 vs P2) revealed significant differences, suggesting that inpatients fared better than outpatients. Comparisons of outcomes between outpatients in the randomised and preference groups were non-significant (R1 vs P1). The same comparisons for inpatients (R2 vs P2) were difficult to interpret.

Two-stage RCT

Wennberg designs
In these RCTs, participants were randomly allocated to a randomised or preference cohort. Participants in the preference cohorts were given the opportunity to choose their treatments.

***Noel* et al., *1998*[131]**
Diabetic outpatients were given standard education or an innovative, educational programme for management of their illness. The authors identified the primary outcomes as the number of educational classes attended by the participants and their knowledge of diabetes at 6 months. Non-significant differences were reported for comparisons between the randomised and preference cohorts (R vs P), adjusted for treatment, and between the two treatment groups [(R1 + P1) vs (R2 + P2)].

***Rokke* et al., *1999*[111]**
Older people with depression were treated with (1) self-management therapy with a cognitive focus, (2) self-management therapy with a behavioural focus or (3) placement on the waiting list for future treatment. This RCT was a modification of the two-stage, randomised design. Participants were randomly assigned to a waiting list control group, a randomised cohort and a preference cohort. However, in the preference cohort, participants could not choose to be on a waiting list. The authors identified scores on the Hamilton Depression Scale, Beck Depression Inventory, a geriatric depression scale and a depression belief questionnaire as their primary outcomes but only reported an analysis of the first three. Any psychological treatment was significantly better than placement on a waiting list [(R1 + R2 + P1 + P2) vs R3)]. However, no significant differences were observed between the psychological treatments, irrespective of allocation [(R1 + P1) vs (R2 + P2)]. There were also no differences between randomised and preference cohorts irrespective of type of psychological treatments (R vs P).

Rücker designs
In these RCTs, participants were also randomly allocated to a randomised or preference cohort and participants in the preference cohort were given the opportunity to choose their treatments. However, in contrast to the Wennberg design, they were randomised if they had no particular treatment preference.

***Cooper* et al., *1997*[105]**
Women with heavy menstrual bleeding were allocated to transcervical resection of the endometrium or standard medical treatment chosen by the physician. As the authors did not identify a primary outcome, we identified treatment acceptability and willingness to opt for the same treatment in the future as primary. Each was measured 4 months after the interventions.

Women randomised to transcervical resection were more likely to find the treatment acceptable and would (hypothetically) opt for the same treatment in future than those randomised to usual medical care (R1 vs R2). No differences in outcome were found between the randomised and preference cohorts receiving transcervical resection (R1 vs P1). However, women who chose medical treatment were significantly more likely to find that treatment acceptable and to choose that treatment given the same circumstances than those randomised to medical treatment (P2 vs R2).

Janevic et al., *2003*[108]

Women with cardiovascular diseases were treated in three arms: a home-based, self-directed, disease management programme; one provided in a group led by a facilitator; or usual care. This RCT was described as a modification of the Rücker design. Although we cannot see how it differs from a Wennberg design, we include it here. First, participants were not contacted or told of the existence of the other arms until after randomisation. Second, all patients in the preference cohort were required to choose the intervention of their choice (as opposed to randomising those with no preference). The authors' primary outcomes were measures of adherence, namely participants' full attendance to at least **one unit** of training and completion of at least **one level** of training. This was assessed at 6 weeks, after the programme ended, in an adjusted analysis. There were no significant differences in outcome between those in the randomised and those in preference cohorts allocated to the self-directed management programme (R1 vs P1). However, participants in the preference arm of the group programme were much more likely to attend than those randomised to that arm (P2 vs R2). No comparisons were made within randomised and preference cohorts.

Preference measurement at baseline in a standard RCT

In both of these RCTs, all patients were randomised and participants' preferences were assessed prior to randomisation.

Hardy et al., *1995;*[96] *Shapiro* et al., *1994*[112]

Adults with depression were allocated to eight sessions of psychodynamic, interpersonal psychotherapy (IP), eight sessions of cognitive behaviour therapy (CBT), 16 sessions of IP or 16 sessions of CBT. Patients reported treatment credibility at baseline. The authors did not identify a primary outcome but we chose scores on the Beck Depression Inventory as the outcome of primary interest, measured at baseline and 16, 28 and 32 weeks. Patients randomised to eight or 16 sessions of CBT had lower depression scores than those randomised to eight or 16 sessions of IP [(R1 + R3) vs (R2 + R4)]. Duration, as distinct from type, of therapy was not associated with outcome [(R1 + R2) vs (R3 + R4)]. Participants rating CBT highly on credibility were more likely to do well with IP, but this had no effect on outcome given CBT. High ratings for credibility of IP appeared to play no role in the outcome of IP or CBT. However, when treatment was broken down by duration, initial credibility ratings and emergent credibility ratings were positively associated with recovery for the eight- but not for the 16-session treatment.

Moffett et al., *1999*[120]

People with low back pain of at least 1 month's duration were offered 1-hour, exercise classes twice weekly for 4 weeks (in addition to the usual general practice care) or to continue with usual general practice care alone. The primary outcome identified by the authors was a measure of disability (Roland disability scale) assessed at 6 weeks, 6 months and 1 year after the intervention was delivered. Although outcome at 6 weeks was not significantly different between RCT arms, there were changes in favour of those receiving exercise classes at 6 months and 1 year. Changes in disability were similar regardless of preference. There was no significant interaction between patient preference and treatment.

Other unusual designs

Devine et al., *1973*[95]

Psychology students who had an extreme fear of snakes were recruited to an unusual small comprehensive cohort preference RCT, in which one group of participants was allocated to their non-preferred treatment. The interventions were systematic desensitisation to snakes; an encounter with a snake; rationale emotive therapy designed to deal with the snake phobia; and a combination of modelling and behavioural rehearsal therapy. There were 12 RCT arms: four randomised, four preference arms and four arms in which students were allocated to a non-preferred treatment arm (NP). The method of allocation was not described. The primary outcome was a rating of snake phobia 1 week after the second therapy session. Comparisons of outcome between the systematic desensitisation groups and the encounter with a snake irrespective of allocation were non significant [(R1 + P1 + NP1) vs (R2 + P2 + NP2)]. However, comparison of outcome between the randomised, preference and non-preferred

treatment groups were significant (R vs P vs NP). The participants in the preference arms had less fear of snakes than those in either the randomised or non-preferred treatment arms for encounter and rational emotive techniques.

***Renjilian* et al., *2001*[94]**

This unusual RCT only randomised patients who had strong preferences for a particular treatment. Overweight people were recruited to individual or group therapy for weight reduction. The treatment package in both programmes combined cognitive behavioural, weight management therapy, a low-calorie diet and an exercise programme delivered over 26 weeks. Potential recruits were asked to indicate their preferences for individual or group therapy using a Likert scale. Those with no or only slight preferences were excluded from the study. The remainder with clear preferences for either treatment were stratified on the basis of their preference and percentage overweight before random assignment to either treatment arm. The main outcome of interest selected by the researchers was body weight measured before and after treatment. Multivariate analysis of change in outcome with treatment, preference and treatment by preference interaction suggested that group therapy patients lost more weight than individual therapy patients, but the effect of preference and treatment by preference interaction were not significant.

***Van Dyck and Spinhoven 1997*[93]**

People suffering from agoraphobia with panic were first asked to indicate their preference for one of two therapies. The first consisted of 4 hours of contact with a therapist over 4 weeks and homework consisting of 10 hours of practice with self-hypnosis and 7.5 hours of exposure to the factor(s) provoking anxiety. The alternative also comprised 4 hours of contact with the therapist but homework (17.5 hours) was dedicated entirely to exposure to the provoking factors. Having stated their preference, half of the patients were allocated to their preferred treatment and half to the non-preferred therapy. Subjects then switched treatments in a crossover design. Method of allocation was not given. The primary outcome was participants' tolerance of the trigger for the agoraphobia and panic symptoms, measured at baseline and 2 and 4 weeks after therapy. Analyses showed a significant effect of time with improvement from baseline to 4 weeks and from 4 weeks to post-treatment, but no significant effect of treatment or preference or treatment by preference interaction.

In the next chapter we summarise the analysis of the systematic review.

Chapter 6

Results of systematic review: evidence for preference effects

In this chapter we provide a systematic overview of the studies identified by our search in order to examine the evidence for preference effects. We describe the adequacy of the methodology used in each RCT according to conventional criteria, evaluate the published analyses and results and then present our analyses of those studies where adequate data were available.

Quality of the RCT methods

Identification of primary outcomes

The trialists identified at least one primary outcome in 21 RCTs and we were able to deduce a primary outcome in all of the remainder (Appendix 1). In 25 RCTs the primary outcome was a disease specific measure (e.g. scores on depression rating scales in RCTs of treatment for depression); in three a measure of participants' perception of, or satisfaction with, treatment; in two a QoL measure; in one service utilisation; in one patients' knowledge of the disease under treatment; in one a disease-specific outcome and also service use; and in one QoL and satisfaction with the intervention (Appendix 1). Seventeen RCTs reported *a priori* sample size estimations.

Adequacy of other aspects of RCT methodology

The quality of the concealment of randomisation using the Cochrane criteria was considered 'OK' in 14 RCTs, 'unclear' in 19 and 'vulnerable' in the one remaining.[131] It was not possible to mask either the participant or the health worker to the interventions in all RCTs, but it was potentially possible to mask the researcher undertaking the study evaluation in all cases. Thirteen RCTs referred to whether or not the researchers were masked to treatment allocation but in only five of these were explicit attempts made to mask them (Appendix 2).

Results relating to hypothesis one

Recruitment to RCTs

Comprehensive cohorts

In ten RCTs the numbers of people agreeing to enter the study prior to allocation were not reported. In the remaining 15 comprehensive cohort studies, eight reported either full or nearly full participation (>96%) and the rest reported participation of at least 60% of the eligible subjects (Appendix 2). Data on numbers accepting randomisation were reported in all the 25 comprehensive cohort studies and these rates varied from 26 to 88%. In only six RCTs ≥ 60% of the eligible people agreed to randomisation and in nine RCTs <40% agreed to be randomised. In three RCTs, it was not clear whether patients were aware that they could have a choice of treatment arm at the time they were invited to be randomised.[121–125] All three RCTs also depended heavily on physicians' preferences to guiding those participants who were prepared to participate in the preference arms. Mean acceptance of randomisation in these RCTs was 44% compared with 51% in the other 22 comprehensive cohort designs. Physicians' attitudes may have reduced the acceptance of randomisation.

It was difficult to discern a clear association between the types of treatment in each RCT and the proportion of people accepting randomisation (*Table 3*). The data might suggest that lower acceptance of randomisation occurred when unusual treatments were on offer (e.g. acupuncture), where there were differences in desirability between treatments (e.g. antidepressants versus psychotherapy) or where differences in time committed to treatment were required within the RCT (e.g. outpatient versus inpatient treatment).

Two-stage randomised RCT designs

The trialists conducting three of the two-stage RCTs implied that **all** eligible participants agreed to be randomised to the randomised or preference cohorts (Appendix 2).[105,111,131] In the remaining two-stage study by Janevic and colleagues randomisation was pre-consent.[108] The numbers refusing to take part in the next level of allocation (i.e. to R1 or R2 and P1 or P2) varied widely as follows:

- Cooper and colleagues[105] reported that 41/138 (30%) refused randomisation to the randomised

TABLE 3 *Proportions of participants accepting randomisation according to interventions in comprehensive cohort studies*

Clinical problem, study details	Intervention and comparison group(s)	Study design	Percentage accepting randomisation
Low back pain Kerry, 2000[118]	1. X-ray with standard treatment 2. Standard treatment alone	Comprehensive cohort	26
Sedation for gastroscopy Melchart, 2002[127]	1. Acupuncture 2. Midazolam: intravenous	Comprehensive cohort	26
Depression Bedi, 2000[109,110]	1. Counselling 2. Antidepressants	Comprehensive cohort	32
Lung cancer Helsing, 1998[113]	1. CT 2. BSC	Comprehensive cohort	32
Opiate addiction Gossop, 1986[128]	1. Outpatient (counselling and oral methadone) 2. Inpatient (oral methadone)	Comprehensive cohort	33
Alcohol addiction McKay, 1995[130]	1. Day hospital outpatients 2. Inpatient rehabilitation Both 28-day rehabilitation programmes	Comprehensive cohort	33
Dysfunction uterine bleeding Bain, 2001[106]	1. Local anaesthesia 2. General anaesthesia For microwave endometrial ablation	Comprehensive cohort	37
CHD Olschewski, 1992[125]	1. Medical treatment 2. Surgical treatment	Comprehensive cohort	37
Foetal Karyotyping procedures Nicolaides, 1994[104]	1. Early amniocentesis 2. Chorionic villus sampling	Comprehensive cohort	38
Oesophageal cancer Riedl, 2001[114]	1. SDD 2. Control/no SDD (intravenous antibiotic prophylactic only)	Comprehensive cohort	40
Otitis media in children Paradise, 1990[99]	1. Surgery: adenoidectomy in all cases. 2. Control: not defined. Assumed no treatment beyond initial medication	Comprehensive cohort	46
CHD SB King, 1994, 1995, 1997[122–124]	1. PCTA 2. CABG	Comprehensive cohort	47
CHD Detre, 1999[121]	1. PTCA 2. CABG	Comprehensive cohort	48
Recurrent throat infections in children Paradise, 1984[98]	1. Surgery: tonsillectomy or tonsillectomy with adenoidectomy 2. Control: not defined. Presumed no surgery	Comprehensive cohort	49
Chronic pain Williams, 1999[132]	1. Outpatient 2. Inpatient Both groups received CBT for chronic management of pain	Comprehensive cohort	49
Cerebral palsy in children Reddihough, 1998[100]	1. CE; educational programme to develop motor and cognitive skills in tandem 2. Traditional neuro-developmental programme. No emphasis on motor coordination (control)	Comprehensive cohort	52
Termination of pregnancy Henshaw, 1993, 1994, 1997[102,103,126]	1. Medical termination (mifepristone & gemeprost) 2. Surgical termination (vacuum aspiration)	Comprehensive cohort	54
Otitis media in children Rovers, 2000, 2001[97,101]	1. Ventilation tubes 2. Watchful waiting (control)	Comprehensive cohort	58

continued

TABLE 3 *Proportions of participants accepting randomisation according to interventions in comprehensive cohort studies (cont'd)*

Clinical problem, study details	Intervention and comparison group(s)	Study design	Percentage accepting randomisation
Breast cancer Schmoor, 1996[116]	1. 6 cycles chemotherapy 2. 6 cycles chemotherapy + radiotherapy	Comprehensive cohort	59
Breast cancer Schumacher, 1994[115]	Factorial design 3 cycles chemotherapy 3 cycle chemotherapy + tamoxifen 6 cycles chemotherapy 6 cycles chemotherapy + tamoxifen	Comprehensive cohort	66
Cocaine addiction Mckay, 1998[129]	1. Day hospital outpatient 2. Inpatient Both 28-day, 12-step oriented, rehabilitation programme	Comprehensive cohort	67
Depression M King, 2000[26,27]	1. CBT 2. NDC 3. GP care (usual care)	Comprehensive cohort	70
Panic disorders Bakker, 2000[117]	1. Cognitive therapy 2, 3, 4. Medical therapy groups	Comprehensive cohort	81
Termination of pregnancy Ashok, 2002[107]	1. Medical termination (mifepristone, misoprostol) 2. Surgical termination (vacuum aspiration)	Comprehensive cohort	83
Low back pain Kendrick, 2001[119]	1. Lumbar spine radiograph 2. No X-ray	Comprehensive cohort	88

BSC, best supportive care; CE, conductive education; CHD, coronary heart disease; CT, chemotherapy; MI, myocardial infarction; NDC, non-directive counselling; SDD, selective bowel decontamination.

cohort and 5/135 (4%) refused to either choose their treatment arm or accept randomisation.
- Janevic and colleagues[108] reported that 1038/1613 (64%) participants refused to be randomised and 913/1466 (62%) declined to choose a treatment arm.
- Noel and colleagues[131] and Rokke and colleagues[111] reported that all participants accepted the next level of allocation (randomised or chose treatment), but in each case a significant proportion failed to attend or complete the intervention programmes.

Preference measurement at baseline in a standard RCT

Of the two RCTs of this type identified, one did not provide information on recruitment[96] and the other reported 58% of potentially eligible patients were not randomised (Appendix 2).

Unusual designs

Of the RCTs with unusual designs, one reported 100% of eligible subjects agreed to take part in the trial.[95] The proportion agreeing to randomisation in the other two trials is unclear[93,94] (Appendix 2).

Relationship between extent of patient preference and recruitment

Our analysis of recruitment rates in terms of constriction of patient preference was limited by the fact that 25 of the 34 RCTs were comprehensive cohorts, only four contained a two-stage randomisation (Wennberg/Rücker). The two-stage randomisation design presented a particular difficulty as it involves two levels of recruitment. Three of the four RCTs of this type implied 100% recruitment of eligible patients at the first randomisation. However, recruitment at the next stage was much less complete. Although technically this refusal at the second level could be considered attrition, this RCT is unique in that patients have the chance of choice of treatment at the first but not the second randomisation and therefore we report it here as a form of recruitment rate. The mean recruitment rate for the comprehensive cohort designs (n = 15 RCTs) was 88.7%. On account of the small numbers of studies in other categories, we were unable to make meaningful comparisons. There were only four two-stage randomisation studies. One randomised pre-consent and the others reported

100% at first stage; at second stage recruitment ranged from 36.6 to 100%. Only one of the two standard RCTs that measured preference at baseline reported on recruitment, which was at least 42%. Finally, of the three studies with unusual design, one reported 100% recruitment.

Adherence to the intervention

Adherence to an intervention means that a patient completes the course of treatment as recommended for clinical and research purposes. However, adherence is a complex concept in this review of RCTs containing a wide variety of treatment types in terms of nature and intensity of the intervention. In many (usually surgical or chemotherapy) RCTs the intervention involved one procedure at baseline that all patients received, hence there was little variation possible in adherence. However, such treatments are rarely single in nature; many involve preintervention preparation and/or post-treatment convalescent management that are not often referred to in the RCT reports. In other RCTs (e.g. psychological therapies), patients were required to attend a recommended number of therapy or outpatient appointments and therefore there was considerable room for variation in adherence. 'Standard' or 'usual' care is also a treatment arm that rarely considered patients' degrees of adherence. Furthermore, few of the RCTs (and none involving drug therapies) provided data on adherence. Adherence to (or participation in) treatments was therefore not a concept that could be applied evenly across all 34 RCTs. This is detailed as follows:

1. Of the 25 comprehensive cohort studies:
   (a) Six RCTs[26,27,100,109,110,118,119,128] provided no data on adherence to treatments.
   (b) In ten RCTs it could be assumed that all people adhered to their core treatment because **both** treatment arms involved surgical procedures[102–104,107,121–124] or chemotherapy and/or radiotherapy/anaesthesia delivered at a hospital.[106,114,115,116,127] However, no information was given on adherence to pre- or postsurgical procedures/treatments.
   (c) In five RCTs treatment was a surgical procedure[98,99,101,125] or hospital-delivered chemotherapy[113] in **one** arm and once again an assumption was made of full adherence. In addition, the trialists made no explicit statement on adherence in the other arms.
   (d) An explicit statement on adherence was made in only four of the RCTs and no differences were observed between randomised and preference arms:
      (i) 9/35 (26%) participants did not adhere completely to treatment in the randomised compared with 7/31 (23%) in the preference arms in one RCT.[117]
      (ii) The other three RCTs compared outpatient and inpatient treatments. In the first, non-adherence to inpatient care was 13% in the randomised arm and 10% in the preference arm. Non-adherence to outpatient care was 29% in the randomised arm and 20% in the preference arm.[130] In the second RCT[129] non-adherence to inpatient care was 12% in the randomised patients and 20% in the preference patients and for outpatient care it was 47% in the randomised arm and 38% in the preference arm. In the third RCT, non-adherence to inpatient care was 5% in the randomised arm and 9% in the preference arm, and 1% in the randomised arm and 0% in the preference arm for outpatient care.[132]
2. Information on adherence was available in three of the two-stage RCT designs. Adherence to treatment was a primary outcome measure for two RCTs[108,131] (Appendices 5 and 6), one of which found greater adherence in the preference arm. The third study reported 80% adherence to therapy in the preference compared with 25% in the random cohort.[111] In the last study that compared surgical and medical treatments, 80% of those randomised to medical care adhered to their treatment but this information was not provided for the preference arm.[105]
3. Both RCTs that measured preference at baseline within a standard randomised design provided information on adherence to treatment. One reported 37% non-adherence to the course of treatment for back pain[120] in the intervention group, but no comment was made on adherence in the usual care arm. The other stated that there was complete adherence to psychotherapy in all four RCT arms.[96]

Two RCTs employing unusual designs provided no information on adherence to treatment[93,95] and one stated that non-adherence to treatment across the RCT ranged from 13 to16%.[94]

Baseline differences between randomised and preference cohorts

Baseline demographic measures

Of the 25 comprehensive cohort designs, 19 reported baseline demographic data broken down by allocation to preference and randomised

arms (Appendix 3). We summarised the numbers of baseline comparisons performed and whether there were differences in demographic characteristics between the randomised and preference arms (Appendix 4). At least one significant baseline difference was found in nine RCTs. Fifteen significant differences were reported out of 86 comparisons made overall, a finding greater than expected by chance. Although the likelihood of finding statistically significant baseline differences is related to sample size, the size of an RCT had no influence on whether any demographic differences were found. In four RCTs, preference patients were more educated; in two RCTs white people were more likely than others to be in the preference arms; and in three RCTs preference patients were more likely to be employed and not on benefits.

Baseline clinical and health measures

Twenty-one of the 25 comprehensive cohorts designs provided baseline data on clinical and health measures by allocation to randomised or preference arms (Appendix 5). A total of 257 comparisons were made, in 28 of which significant differences were reported, a finding once again somewhat greater than chance. Eight RCTs reported significant differences – preference patients appeared to have more severe clinical problems in two RCTs,[26,119] less severe problems in three RCTs,[121,129,130] while in the remaining three there was no consistent pattern.[118,122–125]

Results relating to hypothesis two

Loss to follow-up

In 28 of the 29 RCTs that had preference arms, the authors provided information on loss to follow-up in the preference and randomised arms (Appendix 2). In all but one RCT, follow-up data were available for each arm of the study. There was no difference in the mean percentage of patients with complete data on the primary outcome [89.7% in the preference group compared with 88.4% in the randomised group, difference –0.2 (95% CI –2.0 to 1.7), $p = 0.84$ using a paired t-test weighted for sample size].

Review of analyses conducted by trialists

All patients were included in the final analysis in 16 RCTs. In five RCTs the analysis plan was not reported or was unclear. Only four of the 13 remaining RCTs, which did not include all participants in the final analysis, investigated whether bias might have occurred on account of the missing data (Appendix 2). A summary of the analysis performed by the trialists is provided in *Tables 4* and *5*.

Comprehensive cohort studies

In comprehensive cohort studies, treatment effects can be examined within the randomised and preference cohorts or in the whole pooled cohort. Evidence for a difference in treatment effect in the randomised and preference cohorts can be formally assessed by a test for treatment by preference interaction. Evidence for an effect of preference on outcome can also be investigated by direct comparison of outcome between randomised and preference groups for each intervention separately. However, with the exception of the treatment effect in the randomised groups, all these comparisons are potentially confounded by differences in demographic and prognostic factors between patients who prefer one treatment and those who prefer another, and between patients who agree to be randomised and those who do not.

Reported analysis of treatment effects

Table 4 summarises the analysis of treatment effects in the 25 comprehensive studies that are detailed in Appendix 6. One of the 25 studies included only a single intervention arm in the preference cohort (i.e. there was no P2 group).[117] Of the remaining 24 studies:

- 14 compared treatments separately within the randomised (R1 vs R2) and preference arms (P1 vs P2).[26,27,98–100,102–104,106,107,115,116,118,121–124,132] Two of the 14 explicitly tested for a treatment by preference interaction.[115,116]
- Three pooled the data from random and preference cohorts [(R1 + P1) vs (R2 + P2)] and used an overall model that tested the effect of treatment, type of assignment (random/preference) and treatment by preference interaction.[127,129,130]
- Two performed the treatment comparison both within the randomised cohort (R1 vs R2) and within the whole pooled cohort [(R1 + P1) vs (R2 + P2)].[109,110,114] One of these additionally reported a test for treatment by preference interaction.[109,110]
- One pooled the data from random and non-random cohorts [(R1 + P1) vs (R2 + P2)] and ignored type of assignment in the treatment comparison.[128]
- Three examined the treatment effect within the randomised cohort only (R1 vs R2).[113,119,125]
- One did not present estimates of treatment effect for the outcome considered.[101]

TABLE 4 *Review of the analysis of treatment effects conducted by the trialists*

Type of analysis of treatment effect	**Comprehensive cohort studies *N* = 25[a]**	**Two-staged randomised design studies *N* = 4[a]**
Treatment effect assessed (any 1 vs 2)	24/25	3/4
Treatment effect assessed in R cohort: R1 vs R2	21/25	1/4
Treatment effect assessed in P cohort: P1 vs P2 (of *N* with P2 group)	14/24	0/4
Pooled treatment effect assessed with adjustment for R/P: (R1 + P1) vs (R2 + P2) (of *N* with P2 group)	3/24	1/4
Pooled treatment effect assessed **without** adjustment for R/P: (R1 + P1) vs (R2 + P2) (of *N* with P2 group)	3/24	1/4
Treatment by R/P assignment interaction tested (of *N* with P2 group)	6/24	1/4
Treatment effect in P groups adjusted for (any) baseline covariates (for those studies that assessed treatment effect using preference groups)	10/20	2/2

[a] Multiple analyses were performed in each study, hence column totals do not equal total numbers of studies. NB: These analyses do not apply to studies in which 'preferences were measured at baseline in a standard RCT' and the 'other design' studies.

TABLE 5 *Review of the analysis of preference effects conducted by the trialists*

Type of analysis of preference effect	**Comprehensive cohort studies *N* = 25[a]**	**Two-staged randomised design studies *N* = 4[a]**
Preference effect assessed (any R vs P)	13/25	4/4
Preference effect assessed for intervention: R1 vs P1	9/25	2/4
Preference effect assessed for other arm: R2 vs P2 (of the studies with P2 group)	8/24	2/4
Pooled preference effect assessed with adjustment for treatment: R vs P (of the studies with P2 group)	2/24	2/4
Pooled preference effect assessed **without** adjustment for treatment: R vs P (of the studies with P2 group)	5/24	0/4
Preference effect adjusted for (any) baseline covariates (of the studies that assessed preference effect)	5/13	3/4

[a] Multiple analyses were performed in each study, hence column totals do not equal total numbers of studies. NB: These analyses do not apply to studies in which 'preferences were measured at baseline in a standard RCT' and the 'other design' studies.

Therefore, overall, only six of the 24 studies performed a statistical test for interaction between treatment group and type of assignment (random/preference).

Of the 20 studies that examined treatment effects either within preference cohorts or in a pooled analysis, 10[26,27,100,109,110,115,116,118,121–124,129,130] made adjustments for at least one potentially confounding baseline disease and/or other characteristics, whereas the remaining 10 performed unadjusted analyses.

Of the 14 RCTs that examined treatment effects separately within randomised and non-randomised cohorts (R1 vs R2 compared with P1 vs P2):

- In 11 the treatment effects appeared broadly similar in the two cohorts;[26,27,98–100,104,106,107,115,116,122–124,132] however, the levels of statistical significance of the treatment effects were not necessarily similar between randomised and preference groups on account of different sample sizes. In addition, the comparison of treatment effects is difficult to interpret if unadjusted for baseline factors, as was the case for six of these 11 studies.
- In the remaining three RCTs, treatment effects appeared to differ between randomised and non-randomised cohorts. The BARI study report[121] compared coronary artery bypass grafting (CABG) and percutaneous transluminal coronary angioplasty (PTCA) in diabetic patients, using randomised and non-randomised populations. In the randomised comparison, CABG was associated with reduced cardiac and all-cause mortality compared with PTCA, whereas there was no such treatment effect in the registry group. The non-randomised patients who received PTCA and CABG represented the lower and upper extremes of angiographic disease severity, respectively. Although adjustment for baseline severity increased the preference group treatment effect in favour of CABG, the effect remained small and non-significant, suggesting that confounding by disease severity did not fully explain the differing treatment effects. Henshaw and colleagues compared medical with surgical abortion.[102,103] Among randomised patients, medical abortion received a poorer rating than surgical abortion, but among preference patients the two treatments were rated similarly. A result of this sort might be expected, given that the primary outcome was itself a measure of preference. Finally, Kerry and colleagues reported that patients with back pain randomised to receive X-ray or no X-ray examination had similar levels of disability at outcome and a similar likelihood of subsequent consultation, but in the preference cohort both disability and the likelihood of consultation were significantly increased for the X-ray compared with the no X-ray group.[118] Sample sizes were moderate to large in all three studies, but none reported a formal test for treatment by preference interaction.

Of the six studies that examined the treatment by preference interaction for the primary outcome variable, none found significant evidence to suggest that the effect of treatment differed in the randomised and non-randomised subgroups.[109,110,115,116,127,129,130] However, the power to assess such an interaction was low in most cases.

As discussed in Chapter 3, the similarity of treatment effects between randomised and non-randomised groups does not necessarily imply that preference has no influence on outcome, as treatment outcome may be uniformly better or poorer in preference compared with randomised groups.

Reported analysis of preference effects

Table 5 summarises the analysis of preference effects in the 25 comprehensive studies that are detailed in Appendix 6. Thirteen of the 25 comprehensive cohort studies examined evidence for a preference effect, i.e. performed treatment specific or pooled comparisons of outcome between randomised and non-randomised groups (R1 vs P1; R2 vs P2; R vs P).[101–103,109,110,117–119,122–125,127–130,132] Of seven studies that performed pooled assessments of preference effects (R vs P), two adjusted for treatment group.[129,130] Of the 13 studies that examined preference effects on outcome, adjustment was made for at least one baseline disease and/or other characteristics in only five cases.[117,118,122–124,129,130]

Five of the 13 RCTs reported evidence for a significant effect of preference on outcome.[102,103,109,110,119,122–124,129] In the EAST revascularisation RCT,[122–124] registry patients treated with CABG had lower mortality than patients randomised to CABG; this difference persisted even after careful adjustment for baseline factors. In this RCT (as in the BARI Study by Detre and colleagues[121]), physician rather than patient preference was the dominant factor in choice of treatment among non-randomised

patients. The authors concluded that physician selection of treatment might have played an important role in the outcomes observed. In a study of treatment for depression,[109,110] there was evidence of a preference effect in long-term follow-up: patients choosing counselling tended to do better than those randomised to it, as assessed by the Beck Depression Inventory, but this effect was not significant in early follow-up. There was no such effect in the antidepressant group. In a study on the termination of early pregnancy,[102,103] the proportion of patients having medical abortion who would choose the same method again was lower for the randomised than the preference group. Although the outcome of this RCT was in itself a measure of preference, there was no similar preference effect in the surgical abortion group. In an RCT of inpatient or outpatient treatment for cocaine users,[129] cocaine use at outcome was similar in the randomised and preference groups, but improvement over time was greater for the randomised patients. In a study of referral for X-ray for back pain patients,[119] disability at 3 months was higher for the group choosing to have an X-ray compared with those randomised to X-ray referral, but this unadjusted effect may well have reflected the higher baseline disability of preference patients who chose referral.

In summary, in the 13 studies that examined preference effects, five found significant differences in outcome between randomised and preference groups. In these five studies, outcome in the preference compared with the randomised group was better in three and worse in two. In four of these five studies, however, the effect was apparent for one intervention group only.

The results of analyses of treatment and preferences effects conducted by the trialists for the comprehensive cohort studies are summarised in *Figure 11*.

Two-stage RCTs (Rücker and Wennberg designs)

The approach to analysis in these four RCTs is broadly similar to that for comprehensive cohort studies. The main difference is that all patients are initially randomised and therefore the R and P groups are well matched at baseline. However, comparisons other than R1 vs R2 are still subject to confounding (albeit to a lesser extent than comparisons in comprehensive cohort studies), as patient characteristics may determine choice of treatment. The potential for any preference effect may also be reduced in this type of RCT, as all patients entering the RCT have agreed to the initial randomisation. Those with stronger preferences may not be represented, as they may have refused to take part. *Tables 4* and *5* summarise the reported analysis of treatment and preference effects in the four two-stage RCTs.

Reported analysis of treatment effects

One of the four studies performed a treatment comparison in the randomised arms (R1 vs R2) only,[105] one pooled the data from randomised and preference groups [(R1 + P1) vs (R2 + P2)] and used an overall model that incorporated treatment, type of assignment (random/non-random) and treatment by assignment interaction,[131] one pooled the data from randomised and preference groups [(R1 + P1) vs (R2 + P2)] and ignored type of assignment in the comparison,[111] and one did not perform any treatment comparisons.[108] In the one study that tested for treatment by assignment interaction, none was found.[131] Of the two studies that examined treatment effects using the preference groups,[111,131] both made adjustment for at least one potentially confounding baseline disease and/or other characteristics.

Reported analysis of preference effects

All four studies performed overall (R vs P) or treatment specific randomised/preference (R1 vs P1 or/and R2 vs P2) comparisons of the primary outcome. Three[108,111,131] of the four RCTs made adjustment for baseline differences in the outcome variables and one additionally adjusted for baseline socio-demographic factors and investigated confounding by other health characteristics.[108] In an RCT of treatments for menstrual bleeding,[105] treatment acceptability was higher for the preference medical group than the randomised medical group, but there was no similar effect in the surgical arms, where acceptability was high in both. In a study comparing two cardiovascular disease management programmes, attendance was higher for preference than randomised patients in the group programme arm, but there was no similar effect in the individual programme arm.[108] A diabetes management RCT with a similar attendance outcome found no significance differences between preference and randomised groups overall (R vs P).[131] Similarly, no evidence of a preference effect on attendance (R vs P) was found in a depression management RCT.[111]

In summary, two of the four RCTs found evidence for a positive effect of preference on outcome but in both cases this difference was apparent for one intervention only. Furthermore, the outcome variable in one of these two RCTs was in itself a

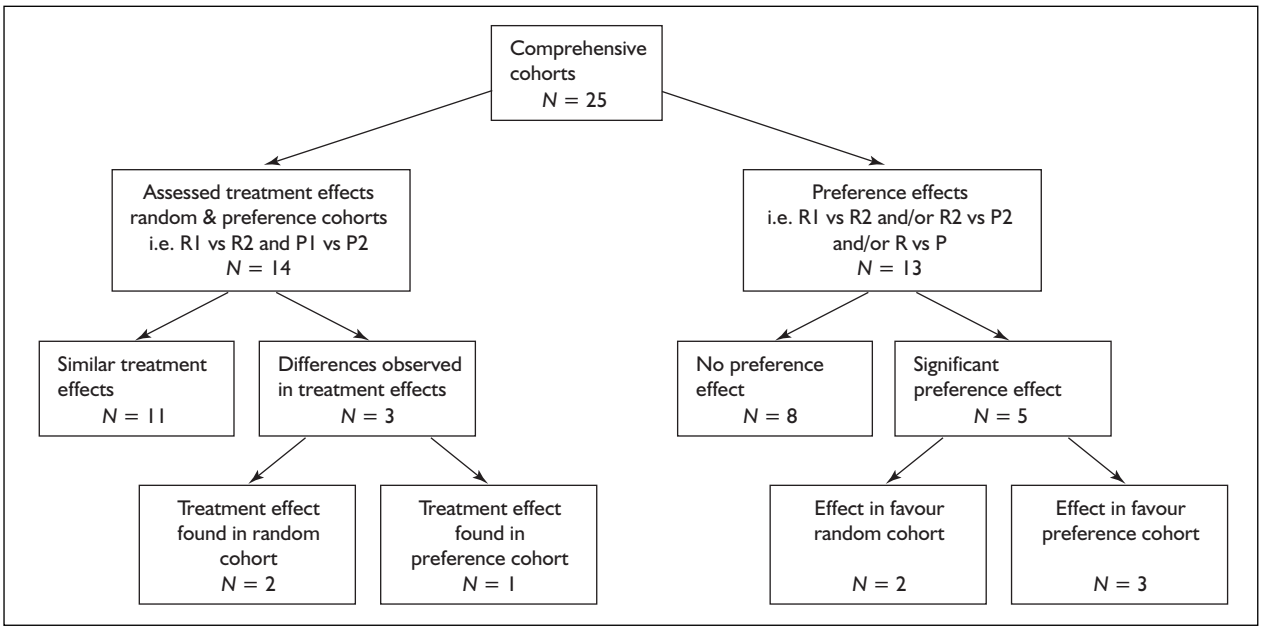

FIGURE 11 *Summary of analysis conducted by the trialists on comprehensive cohort studies*

measure of preference, whereas that in the other RCT was a measure of attendance rather than of disease. The remaining two RCTs found no evidence for a preference effect.

Preference measurement at baseline in a standard randomised controlled RCT

In this design there is no preference cohort, but the interaction between preference and treatment effect can be evaluated. As for the two-stage randomised designs, this design is limited by the constriction on patient choice and therefore preference effects may be reduced. In the first RCT comparing two approaches to the management of back pain,[120] improvement over time in disability was similar among those who did and did not prefer each treatment, and preference by treatment interaction was not significant (Appendix 6). The authors concluded that treatment preference did not appear to affect outcome, albeit that their study lacked power to detect a modest interaction. The other study of this type was Hardy and colleagues' study of treatment for depression.[96,112] The results of this study are difficult to interpret. There was some suggestion that credibility ratings were associated with outcome as measured by the Beck Depression Inventory, but the results were not consistent in that a high credibility rating for cognitive therapy was associated with a better outcome only in the interpersonal therapy group.

Other designs

The three RCTs with unusual designs differed from the remainder in that some or all patients were randomised to receive their non-preferred treatment. The preference effects that could play a role here are complex. On the one hand, the potential for any preference effect on outcome may be reduced, as all patients agreed to be randomised. On the other, it could be argued that the potential for a preference effect may be increased, because participants who wanted a particular treatment are compared with those who specifically did not want that treatment, rather than those who had no preference.

In Van Dyck and Spinhoven's RCT of treatment for panic disorder[93] and Renjilian and colleagues' RCT of obesity treatment,[94] the analyses used models that incorporated treatment and preference factors, and treatment by preference interactions, with adjustment for baseline variables (Appendix 6). Neither study found any evidence for a preference effect on outcome or a treatment by preference interaction. The final study compared treatments for fear of snakes[95] and found evidence that patients who received their preferred treatment had less fear at outcome than those who received their non-preferred treatment or those who were randomised to treatment. This effect appeared to be restricted to the encounter and rational emotive therapies only.

Analysis of preference effects conducted by reviewers using effect sizes

The details are summarised in Appendix 7 and the analyses are summarised in *Tables 6* and 7. We included all usable data from all the comprehensive

cohort and two-stage trials in the review. We calculated standardised intervention-specific differences between randomised and preference groups at outcome.

Outcome preference effect sizes should be interpreted together with the baseline preference effect size. For example, an RCT may show a strongly positive effect of preference on outcome but this would not imply that patient preference influences outcome if a similarly large difference in favour of the preference group already existed at baseline. Interpreting the overall pattern of these effect sizes is a subjective matter, particularly given that the majority of outcome preference effect sizes were small or medium. However, if precedence is given to medium or large effect sizes (>0.2 using absolute value), the following can be observed.

RCTs where analysis was not possible

There were seven studies for which preference effects could not be calculated: five provided no summary data for the preference arms[111,113,115,116,132] and standard deviations were not given or could not be approximated in the remaining two.[119,127]

RCTs where only outcome effect sizes could be calculated

There were 15 RCTs in which only **outcome** effect sizes could be calculated (Appendix 7 and *Table 7*). One of these RCTs reported outcome and baseline data for only one of the two primary outcomes, hence we calculated both an outcome effect size and a net effect size for the two primary outcomes.[102,103]

Three RCTs involved physicians' preferences, and in all the outcome was in favour of the preference arms.[121–125] For the remaining 12 RCTs:

- Five were in favour of preference arms, of which three involved clinical outcomes,[107,114,128] one perception of/satisfaction with treatment[102,103,126] and one the degree of adherence to treatment.[108]
- Two were in favour of the randomised arms[98,104] and both involved clinical outcomes.
- Five had no discernible pattern.[99,101,105,106,131] Four of these concerned clinical outcomes, two satisfaction with treatment and one adherence to treatment.

Three of the RCTs described above were two-stage RCT designs:

- One was in favour of preference arm.[108]
- Two had no discernable patterns.[105,131]

RCTs where the net effect size could be calculated

Baseline and outcome treatment effects were reported, or could be calculated, in eight RCTs (Appendix 7 and *Table 6*), in all of which the primary outcomes were clinical in nature. Patient-generated outcomes such as satisfaction or adherence with treatment cannot be measured before the treatment starts. Considering net effect sizes, of the eight RCTs, we found that:

- Two favoured the preference arm.[102,109]
- Two favoured the randomised arm.[26,129]
- In the remaining four there was no clear trend in favour of preference or randomised arms.[100,117,118,130]

None of the RCTs described in this section were two-stage, randomised designs.

Summary of the analyses conducted by reviewers

Figures 13–15 provide a graphical display with Forest plots of the effect sizes provided in *Tables 6* and *7* and Appendix 7. For studies with multiple outcome and/or time points, a single outcome variable and time point were chosen for each study; the decision was made by the reviewers according to clinical relevance.

Figure 13 shows baseline and outcome effect sizes for preference versus randomised arms by study and treatment group for the eight studies with baseline measurements of the outcome variable. The studies are shown in order of size, from the smallest at the top of the graph to the largest at the bottom. For each randomised/preference comparison, the open boxes show the baseline effect size and the closed boxes show the outcome effect size. Each of the studies has two randomised/preference comparisons (one for each treatment), with the exception of that of Bakker and colleagues, in which the pooled randomised/preference comparison is shown.[117] For each randomised/preference comparison, the proximity of the open and closed boxes gives information about the outcome effect size in relation to the baseline effect size.

Several observations can be noted from this graph. First, there is no clear indication of a systematic difference across the studies between randomised and preference groups either at baseline or at outcome. Second, few baseline or outcome randomised/preference effects are statistically significant. Third, the largest studies (which give the most precise estimates and therefore the strongest evidence) tend to have the smallest baseline and outcome effect sizes, which are

TABLE 6 *Effect size summary table for RCTs with baseline and end-point data*

Study details	Interventions	Variable	Net effect size (in favour of)
Reddihough, 1998[100] Children with cerebral palsy *N* = 66	1. Conducive education 2. Traditional programme	Cognitive measures[a] Motor measure	**1. Preference (medium)** **2. Random (medium)** **1. Random (medium)** 2. Random (small)
Bakker, 2000[117] CBT therapy *N* = 66	1. CBT	Frequency of panic	1. Preference (small)
McKay, 1995[130] Male alcoholics *N* = 144	1. Outpatient rehabilitation 2. Inpatient rehabilitation	Number of drinking days	*3 months* **1. Random (medium)** 2. Random (small) *6 months* **1. Random (medium)** **2. Preference (medium)** *12 months*[a] **1. Random (medium)** 2. Preference (small)
McKay, 1998[129] Cocaine users *N* = 171	1. Outpatient rehabilitation 2. In patient rehabilitation	Days of cocaine use	*3 months* **1. Random (large)** 2. Random (small) *6 months* **1. Random (medium)** **2. Random (medium)** *12 months*[a] **1. Random (medium)** 2. Random (small)
Bedi, 2000[109,110] Patients with depression *N* = 323	1. Counselling 2. Antidepressants	Beck's depression inventory	*8 weeks* 1. Random (small) 2. Random (small) *1 year*[a] **1. Preference (medium)** 2. Preference (small)
Henshaw, 1993, 1994, 1997[102,103,126] Women requiring abortion *N* = 363	1. Medical abortion 2. Surgical abortion	Anxiety subscale on HADS	**1. Preference (medium)** 2. Random (small)
M King, 2000[26,27] People with depression or mixed anxiety and depression *N* = 464	1. CBT 2. Non-directive counselling	Beck's depression inventory	*4 months* 1. Random (small) **2. Random (medium)** *12 months*[a] 1. Random (small) 2. Random (small)
Kerry, 2000[118] Patients with low back pain *N* = 580	1. X-ray with standard treatment 2. Standard treatment	Roland disability score[a] GP consults	*6 weeks* **1. Random (medium)** 2. Preference (small) *1 year*[a] Random (small) No difference *6 weeks* 1. Random (small) 2. Preference (small) *1 year* 1. Random (small) **2. Preference (medium)**

[a] In RCTs with more than one primary outcome, these outcomes and time points were selected from *Tables 6* and *7* for construction of the Forest plots (*Figures 13–15*).
Medium or large effect sizes (>0.2 using absolute value) are shown in bold.
HADS, Hospital Anxiety and Depression Scale.

TABLE 7 *Effect size summary table for trials with end-point data alone*

Study detail	Interventions	Outcomes	Effect size of outcome (in favour of)
Gossop, 1986[128] Opiate addicts *N* = 60	1. Outpatient care 2. Inpatient care	Withdrawal from opiates	1 and 2 combined **Preference (medium)**
Bain, 2001[106] Endometrial ablation *N* = 98	1. Local anaesthesia 2. General anaesthesia	Pain rating[a] Unpleasant rating	1. **Preference (medium)** 2. Random (small) 1. Preference (small) 2. **Random (medium)**
Riedl, 2001[114] Cancer of oesophagus *N* = 70	1. Selective bowel decontamination 2. No bowel decontamination	Occurrence of pneumonia	**1. Preference (medium)** **2. Preference (medium)**
Paradise, 1984[98] Children with recurrent throat infection *N* = 187	1. Surgery 2. Standard care	Number of episodes of recurrent throat infection	**12 months** **1. Random (medium)** 2. No difference
Paradise, 1990[99] Children with recurrent otitis media *N* = 213	1. Surgery 2. Standard care	Number of episodes of otitis media	*12 months* 1. Preference (small) 2. Preference (small)
Cooper, 1997[105] Women with heavy menstrual bleed *N* = 227	1. Transcervical resection 2. Medical care	Treatment acceptability[a] Opt for same treatment in the future	**1. Random (medium)** **2. Preference (medium)** **1. Random (medium)** **2. Preference (medium)**
Rovers, 2000, 2001[97,101] Infants with bilateral otitis media *N* = 320	1. Ventilation tubes 2. Watchful waiting	Occurrence of otitis media	*3 months* **1. Random (medium)** 2. Preference (small) *6 months* **1. Preference (medium)** 2. Random (small) *9 months* 1. Random (small) 2. Random (small) *12 months*[a] 1. Preference (small) 2. Random (small)
Henshaw, 1993, 1994, 1997[102,103,126] Women requiring abortion *N* = 363	1. Medical abortion 2. Surgical abortion	Rating good-bad Opt for same treatment	1. Preference (small) 2. Random (small) **1. Preference (large)** **2. Preference (small)**
Ashok, 2002[107] Women requiring abortion *N* = 445	1. Medical abortion 2. Surgical abortion	Number of bleeding days	**1. Preference (medium)** 2. Preference (small)
Noel, 1998[131] Diabetic patients *N* = 596	1. Standard education 2. Nutritional education	Attend any class	1 and 2. Preference (small)
Detre, 2000[121] Patients with coronary artery disease and diabetes *N* = 642	1. PTCA 2. CABG	Mortality all causes Cardiac mortality[a]	**1. Preference (medium)** 2. Preference (small) **1. Preference (medium)** 2. Preference (small)
Olschewski, 1992[125] People with coronary artery disease *N* = 2095	1. Medical treatment 2. Coronary artery surgery	Mortality all causes	1. Preference (small) 2. Preference (small)

continued

TABLE 7 *Effect size summary table for trials with end-point data alone (cont'd)*

Study detail	Interventions	Outcomes	Effect size of outcome (in favour of)
SB King, 1995[122–124] Patients with multi vessel coronary heart disease *N* = 830	1. PTCA 2. CABG	Mortality all causes	*1 year*[a] **1. Preference (large)** 2. Preference (small) *2 years* **1. Preference (medium)** **2. Preference (medium)** *3 years* **1. Preference (medium)** **2. Preference (medium)**
Janevic, 2003[108] Women with cardiovascular disease *N* = 1071	Disease management 1. Self-directed 2. Group management	Completion one unit[a] Completion all units	1. Preference (small) **2. Preference (large)** 1. Preference (small) **2. Preference (medium)**
Nicolaides, 1994[104] Pregnant women first trimester *N* = 1301	1. Amniocentesis 2. Chorionic villus biopsy	Total loss pregnancy Spontaneous loss pregnancy[a] Termination	1. Preference (small) **2. Random (medium)** 1. Preference (small) **2. Random (medium)** 1. Preference (small) 2. Random (small)

[a] In RCTs with more than one primary outcome, these outcomes and time points were selected from *Tables 6* and *7* for construction of the Forest plots (*Figures 13–15*).
Medium or large effect sizes (>0.2 using absolute value) are shown in bold.
HADS, Hospital Anxiety and Depression Scale.

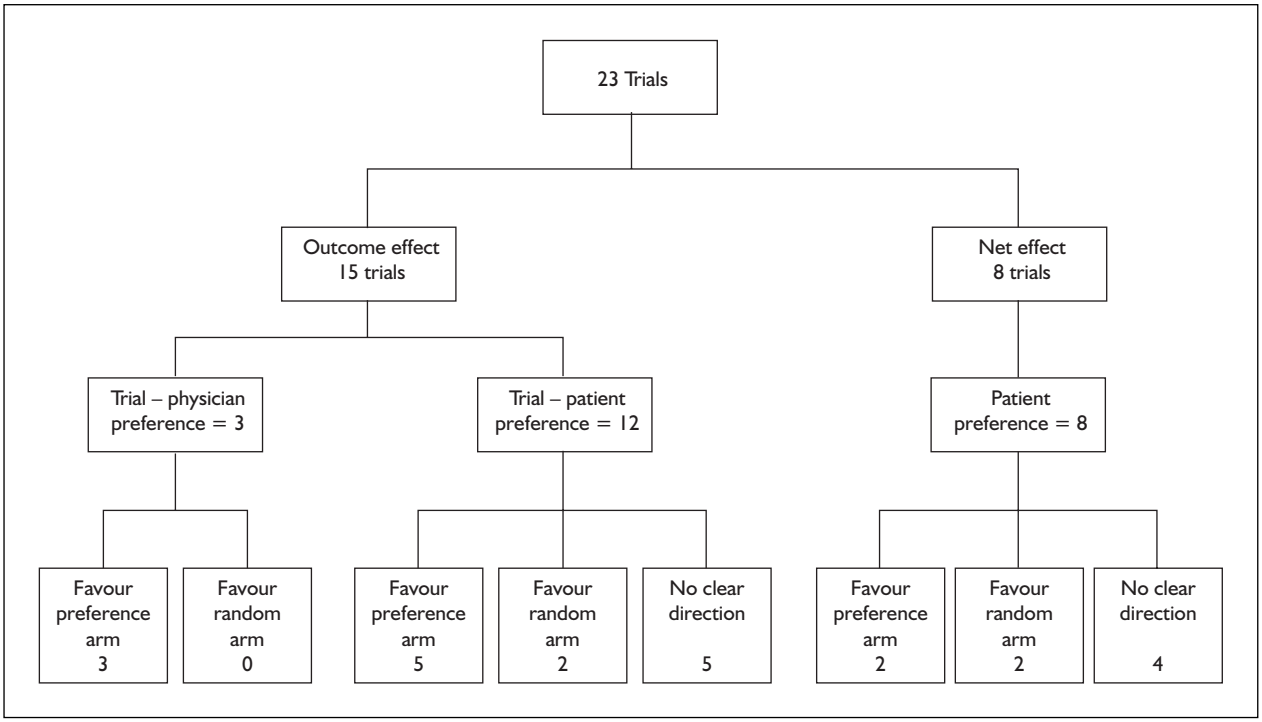

FIGURE 12 *Summary of calculated preference effects according to trial. Effect sizes >0.2 using absolute value are shown.*

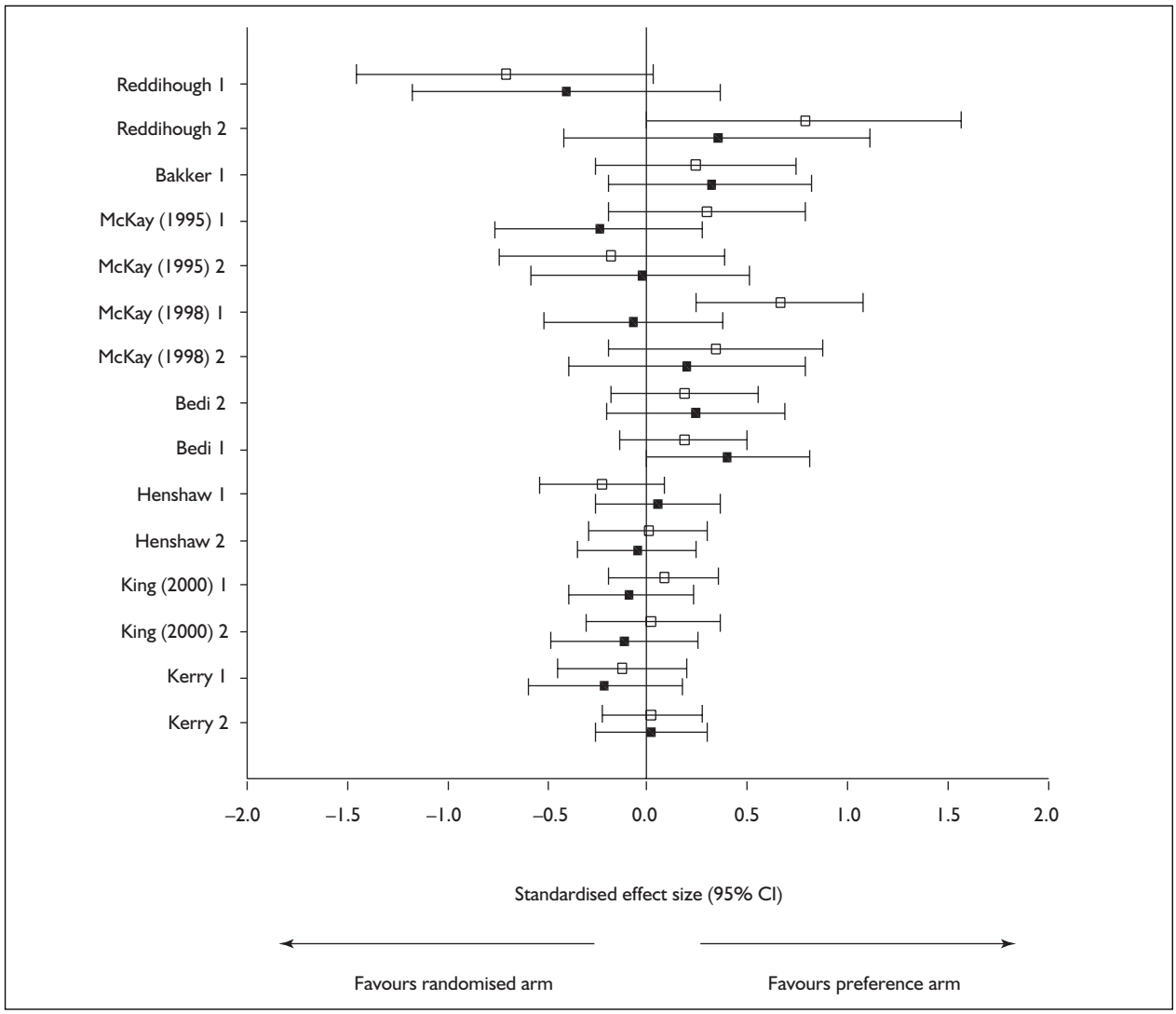

FIGURE 13 *Baseline and outcome effect sizes and 95% CIs for preference vs random cohorts by study and treatment group. Eight studies with baseline and outcome data are presented here. Studies ordered smallest to largest. 1 = Experimental intervention; 2 = control intervention. Single outcome and time point chosen for each study. Baseline effect sizes shown in open boxes; outcome effect sizes shown in filled boxes.*

clustered around zero. Fourth, the direction and magnitude of the baseline and outcome effect sizes for each comparison are similar in most cases, suggesting no clear effect of preference on outcome, over and above differences that existed at baseline. Notable exceptions to this are the following: the 'day hospital' arms in two studies of addiction treatment[129,130] and the traditional neuro-developmental programme arm in a study of patients with cerebral palsy[100] (in these studies moderate effects in favour of the preference groups that existed at baseline were not apparent at outcome), and the conductive education arm of the cerebral palsy study[100] and the medical treatment arm in an abortion study[102] (in these studies baseline effects against the preference group were not apparent at outcome). Net effect sizes (outcome minus baseline) for each of the eight studies are summarised in *Figure 14*. This figure again demonstrates that most net effect sizes were small and that there was no clear evidence in favour of randomised or preference groups.

Figure 15 shows the outcome randomised versus preference effect sizes for the remaining 14 studies that had only outcome information. These studies provide a weaker level of evidence than the studies discussed above, because outcome differences between randomised and preference cohorts are prone to bias. The figure shows that most effect sizes were small or moderate, with no clear pattern in favour of the randomised or preference groups. Three comparisons were statistically significant (all in favour of the preference group).

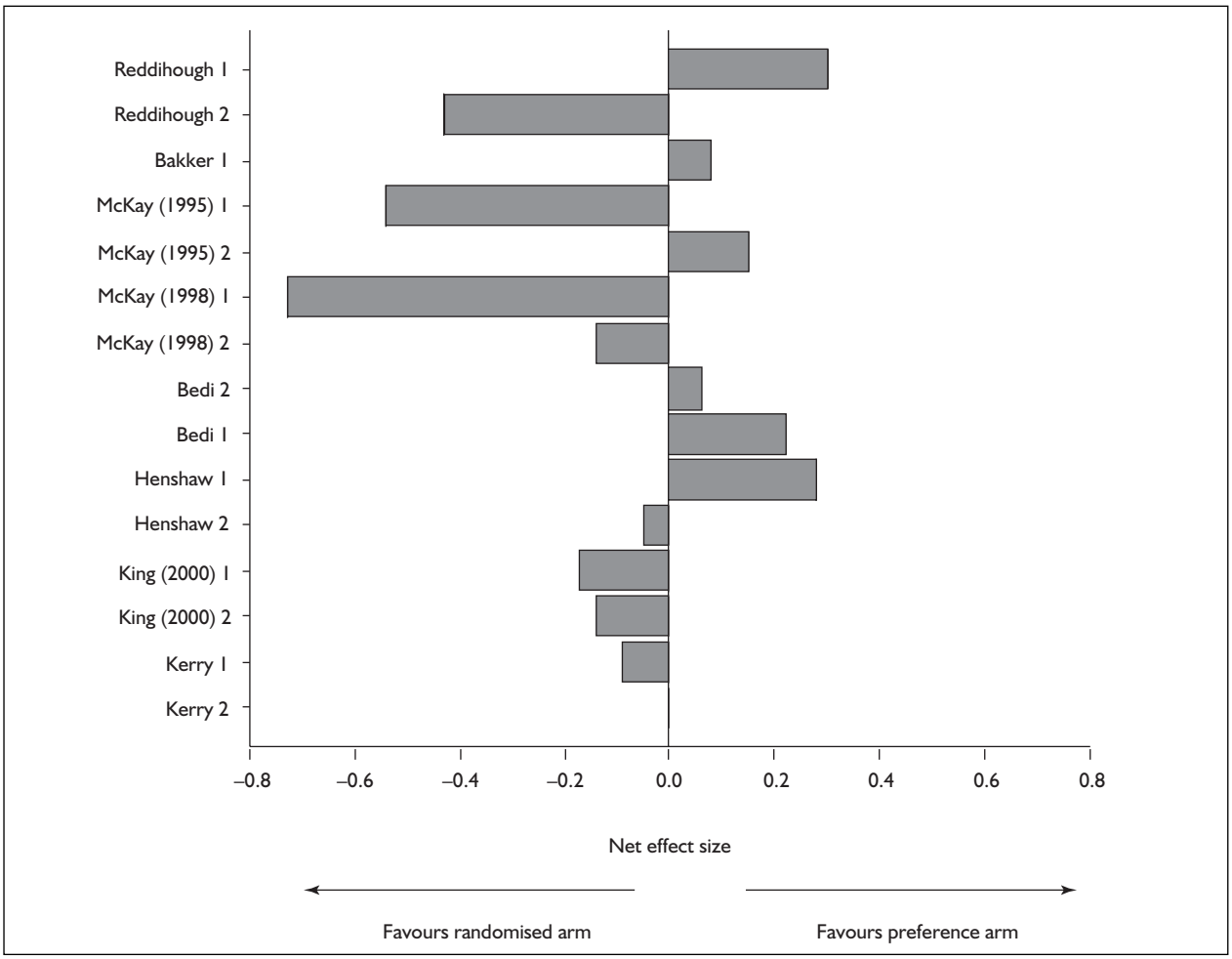

FIGURE 14 *Net effect sizes for preference vs random cohorts by study and treatment group. Eight studies with baseline measurements of outcome variable are presented here. Studies ordered smallest to largest. 1 = Experimental intervention; 2 = control intervention. Single outcome and time point chosen for each study.*

Preference effect sizes in each treatment type

When we considered only medium and large effect sizes for each of the treatments discussed below, we made the following observations:

Termination of pregnancy

The two RCTs (see *Table 8*) comparing medical and surgical approaches found positive effects of preference for those receiving medical treatments on two clinical measures, namely fewer days of vaginal bleeding[107] and reduced anxiety[102,103] after treatment. Additionally, patients who chose medical treatments were more likely than those randomised to it, to opt for the same treatment in the future. None of these results were observed for surgical abortion.

Treatments for ischaemic heart diseases

Data from two RCTs (see *Table 9*) that compared two surgical treatments (PTCA and CABG) for ischaemic heart diseases indicated a positive preference effect for all-cause mortality and/or death due to cardiac causes for people receiving percutaneous transluminal angiogram in both RCTs[121–124] and for patients receiving CABG in one of the RCTs.[122–125] These RCTs are distinct from others in this review in that allocation depended on clinicians' **and** patients' preferences.

Treatment of addiction disorders

Three RCTs compared outpatient and inpatient management of addiction disorders (see *Table 10*). An effect was observed on clinical outcomes (reduced substance use) in favour of the randomised arms for both inpatient and outpatient care in one RCTs.[129] In a second RCT, clinical outcome consistently favoured the randomised group for outpatient care only.[130] In the third RCT, outcome was in favour of the preference group.[128]

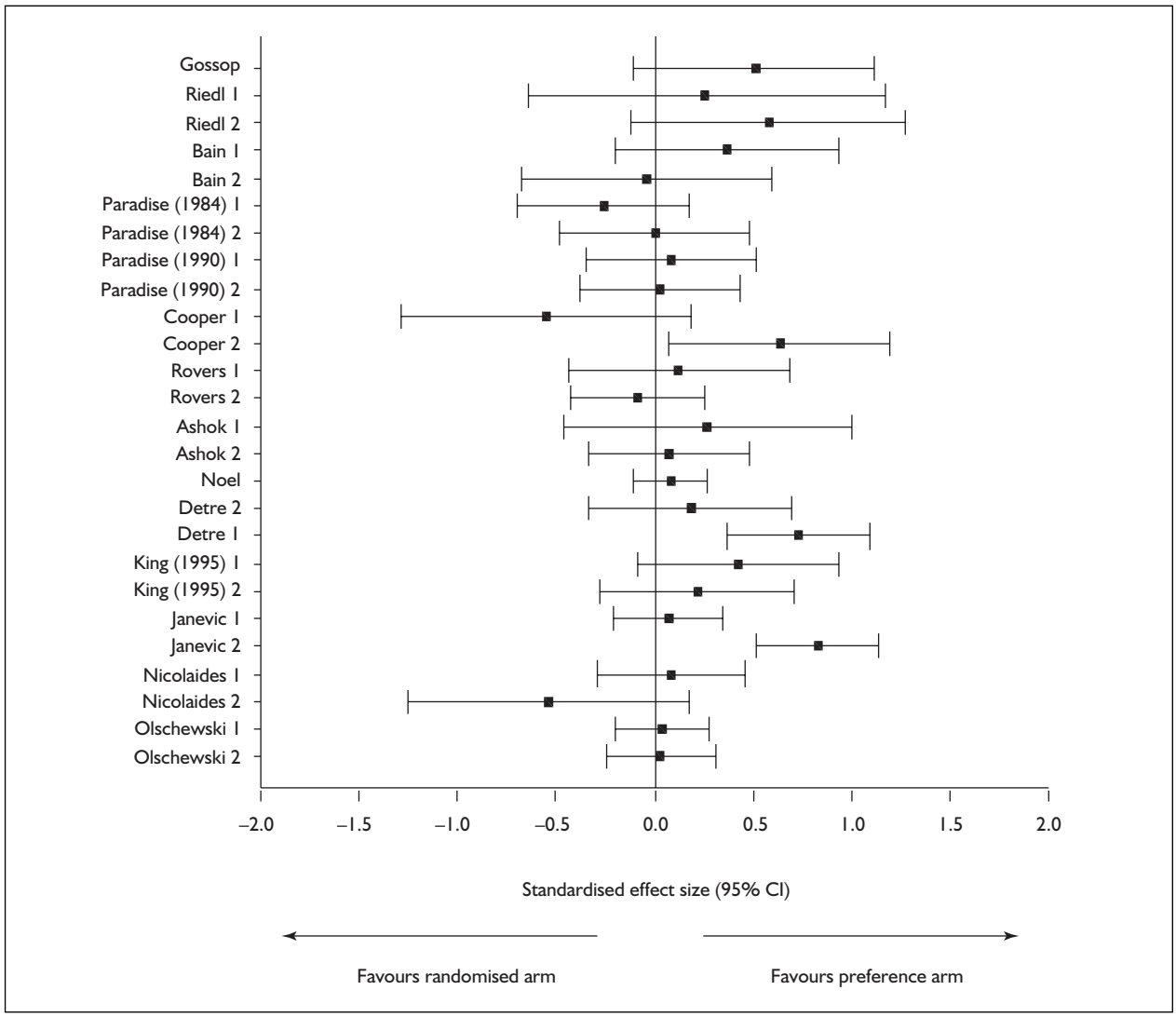

FIGURE 15 *Outcome effect sizes and 95% CIs for preference vs random cohort by study and treatment group. 14 studies without baseline measurements of outcome variable are presented here. Studies ordered smallest to largest. 1 = Experimental intervention; 2 = control intervention. Single outcome and time point chosen for each study.*

TABLE 8 *Two RCTs comparing medical and surgical approaches to termination of pregnancy*

Study detail	Interventions	Outcomes	Outcome in favour of (size of effect)
Henshaw, 1993 1994, 1997[102,103,126] Women requiring abortion *N* = 363	1. Medical abortion 2. Surgical abortion	Rating good-bad	1. Preference (small) 2. Random (small)
		Opt for same treatment	1. Preference (large) 2. Preference (small)
		Anxiety (HADs)	1. Preference (medium) 2. Random (small)
Ashok, 2002[107] Women requiring abortion *N* = 445	1. Medical abortion 2. Surgical abortion	Number of bleeding days	1. Preference (medium) 2. Preference (small)

HADS, Hospital Anxiety and Depression Scale.

TABLE 9 *Three RCTs comparing two surgical treatments (PTCA and CABG) for ischaemic heart disease*

Study detail	Interventions	Outcomes	Outcome in favour of (size of effect)
Detre, 1999[121] Ischaemic heart disease	1. PTCA 2. CABG	Mortality all causes Cardiac mortality	1. Preference (medium) 2. Preference (small) 1. Preference (medium) 2. Preference (small)
SB King, 1995[122–124] Patients with multi vessel coronary heart disease	1. PTCA 2. CABG	All-cause mortality	*1 year* 1. Preference (large) 2. Preference (small) *2 years* 1. Preference (medium) 2. Preference (medium) *3 years* 1. Preference (medium) 2. Preference (medium)
Olschewski, 1992[125] People with coronary heart disease	1. Medication 2. Coronary artery surgery	All-cause mortality at 5 years	1. Preference (small) 2. Preference (small)

TABLE 10 *Three RCTs comparing outpatient and inpatient management of addiction disorders*

Study detail	Interventions	Outcomes	Outcome in favour of (size of effect)
Gossop, 1986[128] Opiate addicts *N* = 60	1. Outpatient care 2. Inpatient care	Withdrawal from opiates	1 and 2 combined Preference (medium)
McKay, 1995[130] Male alcoholics *N* = 144	1. Outpatient rehabilitation 2. Inpatient rehabilitation	Number of drinking days	*3 months* 1. Random (medium) 2. Random (small) *6 months* 1. Random (medium) 2. Preference (medium) *12 months* 1. Random (medium) 2. Preference (small)
McKay, 1998[129] Cocaine users *N* = 171	1. Outpatient rehabilitation 2. Inpatient rehabilitation	Days of cocaine use	*3 months* 1. Random (large) 2. Random (small) *6 months* 1. Random (medium) 2. Random (medium) *12 months* 1. Random (medium) 2. Random (small)

Treatments for depression

The RCTs (see *Table 11*) that evaluated psychological treatments[26,27,111] or psychological therapy against drug therapy[109] for depression or mixed anxiety and depression showed no consistent effects in favour of the random or preference groups on depression scores.

Treatments for back pain

The three RCTs comparing exercise with or without X-rays for low back pain found no clear effect on disability scores in favour of either randomised or preference cohorts.[118–120]

Children with URTI

The three RCTs that compared surgery with standard care (conservative management) for otitis media[97,99] and sore throats[98] in children found no clear effects in favour of either randomised or preference arms on clinical measures such as recurrence rates of URTI.

TABLE 11 *Three RCTs evaluating treatments for depression*

Study detail	Interventions	Outcomes	Outcome in favour of (size of effect)
Bedi, 2000[109,110] Patients with depression *N* = 323	1. Counselling 2. Antidepressants	Depression rating scales	*8 weeks* 1. Random (small) 2. Random (small) *1 year* 1. Preference (medium) 2. Preference (medium)
M King, 2000[26,27] People with depression or mixed anxiety and depression *N* = 464	1. CBT 2. Non-directive counselling	Depression rating scales	*4 months* 1. Random (small) 2. Random (medium) *12 months* 1. Random (small) 2. Random (small)
Rokke, 1999[111] People with depression	1. Cognitive 2. Behavioural	Depression rating scales	No raw data provided but authors report R vs P No significant difference in outcome depression rating scales

Education programmes for diabetes and ischaemic heart diseases

No clear effects were noted in one RCT[131] but Janevic and colleagues reported that group education (rather than self directed learning) was associated with a positive preference effect on participants' adherence to the educational programme.[108]

Results relating to hypothesis three

Preference effects and degree of effort required for participation in the interventions

It proved difficult to make consistent comparisons across the 34 RCTs of the degree of effort required by patients to participate in each RCT arm or the degree of commitment that might influence outcome for each type of intervention. At first sight it would seem that in treatments such as psychological therapies, participants could determine their outcomes to a greater extent than in other treatments by their commitment to, and participation in, the therapy. However, when this principle was applied to the RCTs under review, it became apparent that the majority of interventions are complex and involve a degree of participant commitment. In all RCTs there is the potential for patients to alter the outcome depending on how they perceive the acceptability and efficacy of the treatment and cooperate. Accepting surgery involves commitment to the intervention in addition to the many pre- and postoperative procedures that are not always well described in the RCT reports.[121–124] Moreover, there was little difference in the level of participation required for two different types of interventions under investigation such as medical or surgical termination of pregnancy,[102,103,107] local or general anaesthesia procedures[106,127] or chorionic villus biopsy or amniocentesis for investigation of foetal defects in early pregnancy.[104] Many medical interventions are not simple. For example, chemotherapy for cancer[113–116] involves complex and sometimes unpleasant medication regimens that must be adhered to over many weeks or months, in addition to regular outpatient reviews. Finally, even antidepressant therapy requires at least some degree of confidence in the treatment and a willingness to tolerate possible side-effects, in order to adhere to it.[109,110] An added complication is that some primary outcomes are more vulnerable to influence than others, such as a self-report depression score[27] compared with survival.[121] For these reasons, we did not consider that a classification of the RCTs in terms of degree of participation required by patients was a sufficiently robust procedure by which we might further assess the impact of preferences on outcome.

Chapter 7
Discussion

Principal findings

Conceptual framework

Our conceptual framework defined preferences as a global evaluation relating to the difference in the perceived desirability of two (or more) interventions. Preferences can be usefully conceptualised as a utility, or an attitude.

A four-stage model of the development of preferences was proposed:

1. The first stage concerns the source of preferences. In the current context, this concerns **information** received about interventions in an RCT and the wider issues of informed consent procedures in RCTs.
2. The second stage concerns the psychological processes by which information about interventions is assimilated.
3. The third stage concerns the initial output of these psychological processes, which is a **global preference** for interventions in an RCT.
4. The fourth stage represents **patient decision-making**, related to whether or not they agree to be randomised. When patients are offered participation in a standard RCT, this concerns the decision whether or not to enter the RCT. In a comprehensive cohort design, this concerns whether patients will be randomised or choose to have a particular treatment.

There is no consensus as to how information on treatments is best provided to patients, and in the systematic review only 12/34 studies reported providing specific written information on the included interventions.

Because preferences can be conceptualised as utilities or attitudes, there are a number of appropriate measurement technologies available. Attitude measurement and other psychometric techniques may be the most pragmatic approach, although willingness to pay within economics may provide a useful alternative.

Measurement of preferences within RCTs in the systematic review was rare and few of the RCTs gave details of methods of preference measurement. Where preferences were measured, methods were generally unsophisticated.

Systematic review

There are a limited number of RCTs in the world literature that have taken account of patients' and/or professionals' preferences and utilised them in the design and analysis of outcomes.

- The majority were comprehensive cohort designs.
- Many of the RCTs identified had methodological weaknesses in terms of small sample size, failure to define a primary outcome(s), make a pre-RCT estimation of treatment effect, conceal randomisation or mask treatment groups.
- The quality of statistical analysis varied across the RCTs. Participants with missing data were excluded from analysis in many RCTs, which would have been a potential source of bias.
- There was no consistent approach to examining preference effects.

Hypothesis one

Our first hypothesis was that preferences affect recruitment to RCTs. In order to test this hypothesis we conducted analyses of:

- differences in baseline characteristics between randomised and preference arms in comprehensive cohort designs
- recruitment rates in relation to different levels of 'restriction' of patient preferences.

The significant numbers of patients in preference arms of comprehensive cohort studies indicate that preferences affect recruitment to RCTs if it is assumed that preference patients only participated because they were given a choice of treatment. In most of the comprehensive cohort studies there were no differences between preference and randomised patients at baseline. However, where differences did occur, there were indications that preference patients were more often white, educated to a higher level and employed than randomised patients. Differences in clinical characteristics showed no consistent pattern in favour of randomised or preference groups.

Recruitment to RCTs also varied with design and restriction on patient preferences. In the comprehensive cohort designs, patients were usually told at the outset that they could accept randomisation or choose a treatment arm. This generally led to high recruitment rates. In the two-stage randomisation designs, all patients were randomised initially and the authors reported that 100% agreed. However, refusal at the next stage in the cohort that was randomised was often very high. Hence, the two-stage designs appeared simply to postpone patient refusal. Recruitment was low in the one standard RCT with preference measurement that reported a recruitment rate. Hence it would seem that restriction of preference eventually reduces recruitment.

These findings give support to our first hypothesis, namely that patients' and professionals' preferences affect recruitment to those RCTs, but there is less evidence that external validity is seriously compromised.

Hypothesis two

Our second hypothesis proposed that preferences are important effect modifiers in RCTs. Analysis of the effects of preferences used comparisons of outcome between randomised and preference arms. We first examined the analyses reported in the published papers. However, many authors restricted their analyses to treatment effects rather than preference effects. We analysed the preference effect size for all RCTs where data were available.

No differences were reported in eight of the 13 comprehensive cohort designs that compared outcomes in the preference and randomised arms (R1 vs P1 and R2 vs P2). In the five RCTs where significant preference effects were found, better outcomes in the preference arms were reported in three and worse outcomes in two, but preference effects were usually seen in only one treatment arm. Two of the four two-stage RCTs found evidence for a positive effect of preference on outcome but, once again, this difference was apparent for one treatment arm only. Furthermore, in one RCT the outcome was itself a measure of preference, whereas in the other it was a measure of attendance rather than disease. Neither of the two standard RCTs with preference measurement reported significant preference by treatment interactions. Two of the unusual designs found no evidence of a preference effect on outcome or of a treatment by preference interaction and the third reported some evidence for a preference effect, although the design was complicated.

In our effect size reanalyses of randomised against preference arms (R1 vs P1 and R2 vs P2) for 20 comprehensive cohorts and three two-stage RCT designs, we found little evidence for an effect of preference if RCTs involving physician preference or process outcomes are discounted. Where net effects (considering baseline **and** outcome preference effects) could be calculated, few were of large magnitude. Evidence for moderate or large preference effects was weaker in larger trials.

There were insufficient RCTs in any one clinical disorder or discipline for us to reach definitive conclusions for any particular clinical field. However, significant preference effects on outcome were reported in treatments for termination of pregnancy and surgical treatments for ischaemic heart diseases, while significant effects in treatment of addictions were generally in favour of the randomised groups.

Hence although there was evidence of preference effects in a proportion of RCTs, these effects were inconsistent both between and within studies. Preference effects were also inconsistent in direction. This inconsistency was found in relation to all preference designs. Given this variation and the complexity of analysing and interpreting preference effects, the current evidence is insufficient to provide a strong test of this hypothesis, but the available evidence does not support the suggestion that preference effects are consistent or of significant magnitude.

Hypothesis three

Our third hypothesis was that the magnitude of preference effects is larger in RCTs that require greater effort and participation by participants. The nature of treatments in the studies varied widely and there was considerable variation in the amount of detail given about the exact processes involved in each one. Although we attempted to categorise the interventions with regard to degree of participation based on the limited information given and our own knowledge of such treatments, we found that it was difficult to do so in a consistent and reliable way. In our view, examining differential preference effects based on such unreliable categories would run the risk of drawing incorrect conclusions. Because of this, we refrained from testing our third hypothesis.

Strengths and limitations

One strength of the review was the combination of conceptual framework and systematic review of empirical data. The main limitations are:

1. We cannot be certain that the search strategy would have located all RCTs that made some estimate of concepts potentially related to preferences (e.g. treatment credibility) at baseline, especially where such measures were not a primary focus of the study.
2. Our review involved RCTs of a variety of interventions in diverse settings. This heterogeneity limited the extent to which we could pool data from individual studies. This was especially evident in the difficulties encountered in classifying RCTs according to the level of participation required by each treatment arm.
3. Most of the RCTs were comprehensive cohort designs, which limited the scope for comparisons between preference designs.
4. Seven RCTs with preference cohorts did not contain sufficient data for us to calculate effect sizes.
5. In order to be comprehensive we sought studies from as far back as 1966. This meant however, that RCTs were of variable quality.
6. We cannot be certain that RCTs accurately reported participation rates for all eligible participants. It seems intuitively unlikely that the two-stage randomised designs achieved 100% recruitment at first randomisation but had substantial refusal thereafter. However, our RCT reports are probably no better or worse in this respect than those in other reviews.
7. In evaluating differences in outcome between randomised and preference groups, the extent to which we could adjust for confounding variables was limited.

Non-randomised studies

Evidence from research conducted in the 1960s and 1970s led to the conventional wisdom that randomised experiments should be emphasised in the evaluation of effectiveness.[133–134] Since experimental and control groups are indistinguishable at baseline and the only subsequent difference between them is the intervention, that intervention must be responsible for the outcomes. Observational studies were considered subject to bias and to inflate positive treatment effects as compared with RCTs.[4,135,136] However, recent reviews of these claims based on current and better conducted cohort studies have challenged this view.[6,137] In a systematic review of randomised and cohort studies matched for medical discipline, nature of the intervention and outcome assessments, Benson and Hartz reported that estimates of treatment effects were similar in each.[137] They found 19 treatment comparisons that had been the subject of at least one observational study and one RCT. There were 53 observational studies and RCTs in the analysis. In only two of the 19 paired analyses did the combined magnitude of the effect in the observational studies lie outside 95% CIs for that in the RCTs. In a second review of 18 studies that made a comparison of RCTs and prospective, non-randomised studies, no obvious patterns of outcome emerged.[6] In a review of randomised and non-randomised studies of a single intervention, McKee and colleagues concluded that neither method gives consistently larger estimates of treatment effect.[138] Using a search of systematic reviews, Kunz and Oxman had reached a similar conclusion a year earlier.[139] However, both of the last two reviews included only a small number of RCTs that had allocated at least some participants on the basis of their preferences.

Our review is similar to that of Benson and Hartz[137] with one advantage: our cohort (preference) and randomised patients were evaluated under the same RCT conditions. Although they are not equivalent as they differed in their preferences for treatment arms at baseline, our results lend considerable weight to their conclusions that observational studies do not produce a large or consistent bias in results compared with RCTs.

Comparing randomised and non-randomised designs

Randomised and non-randomised trials both have drawbacks. RCTs may lead to greater treatment effects if patients receive a higher quality of treatment or are highly selected so as to benefit to a greater degree. The validity of RCTs may also be threatened if there are strong patient preferences against one or more treatment arms and outcomes are evaluated in an unmasked or partially masked fashion. Observational studies are perhaps most limited when evaluating treatments that doctors select for the sickest patients.[140] Although several of the comprehensive cohort studies that we reviewed had design or analysis problems, we found few important differences in baseline characteristics between the randomised and preference cohorts. This means that a common criticism of RCTs, namely that they have low external validity because of the self-selected nature of the patients agreeing to participate, was not upheld. We also found few differences between randomised and non-randomised cohorts in their loss to follow-up or treatment effects. Lack of randomisation or the operation of preferences may not be the major threats to internal validity that is often assumed.

Why might preferences have little impact as effect modifiers?

In a light-hearted attempt to assess preferences for red or white wine under survey, masked RCT and unmasked testing conditions, Pazart showed how participants' preferences are subject to influence and change in different settings.[141] This suggests why they have little impact on RCT outcomes, even if patients with strong baseline preferences receive their non-favoured intervention. Renjilian and colleagues' RCT is a striking example, which included only patients with strong preferences and systematically allocated half of them to the non-preferred intervention.[94] They found no evidence of a preference effect on outcome or of a treatment by preference interaction, but the study was small.

There are a number of potential reasons why preferences may not affect outcome:

1. Preferences may be based on inaccurate expectancies that change during the actual experience of treatment. Initial stated or measured preferences would therefore not be related to eventual outcome.
2. Patients may make decisions which are not in line with their preferences because of other perceived pressures on the decision-making process. This again would attenuate preference effects.
3. The effects of preferences may be restricted to particular outcomes, such as satisfaction or adherence, and these effects may be relatively minor in magnitude such that they do not translate into large effects on clinical outcomes or are not detectable within moderately sized RCTs.

Preferences may have greatest effect at the point of potential recruitment to the RCT rather than at any subsequent stage and hence potentially impact most on external validity.

Conclusions

Our conceptual framework has suggested that preferences are broadly based on expectancies concerning the process and outcomes associated with the intervention and the perceived value placed on those outcomes and processes. However, it is possible that participants' preferences may be based on insufficient or incorrect information. In addition, decisions about randomisation may not always accord with preferences and may be influenced by clinicians, relatives or friends. Where there is evidence that the basis of preferences is problematic, and these preferences are likely to affect the external validity of an RCT, it might be possible to improve recruitment rates by presenting potential participants with appropriate information to challenge or overcome preconceptions and ensure informed expectations. However, it is important that such information provision does not stray into coercion. We have suggested how preferences might best be measured. Once participants have been recruited, preferences may affect perceptions of the intervention and satisfaction but appear to exert few major effects on further participation or clinical outcome. Comprehensive cohort designs may still be worthwhile, but when a significant proportion of patients refuse to be randomised and (1) follow-up data are economical to collect, for example from routinely collected sources, or (2) when costs of follow-up are higher, a **random sub-sample of** participants are allocated to their preferred treatment and followed up. The case for a preference design would be strongest where:

- Strong preferences exist that are based on informed expectations concerning interventions.
- There is good evidence that preferences will impact on recruitment.

RCTs remain the gold standard for evidence of effectiveness. However, our review adds to the growing evidence that where strong preferences or ethical objections to an RCT exist, observational methods are a valuable alternative. The outcomes of non-randomised studies best approximate to those of RCTs when exclusion criteria are carefully defined and when prognostic factors and patients and professionals' preferences are well understood. These methods could be used to exploit available clinical databases[137] and also to investigate the effectiveness of interventions where randomisation is very difficult, undesirable or unacceptable to clinicians and/or patients and their families.[142]

Data from observational studies may be valuable in situations where:

- There are strong, well-informed preferences on the part of eligible participants or physicians and when only a small proportion of them will accept randomisation.
- Known confounders of treatment outcome (including strength of preference) are measured and taken account in the analysis.
- There are strong ethical or legal objections to undertaking an RCT.

Methodological tool kit

All RCTs in which participants and/or professionals

cannot be masked to treatment arms should estimate participants' preferences. In this way we shall increase the amount of evidence available to answer questions about the effect of treatment preferences within and outwith RCTs. If preferences are consistently measured, their relationships with outcomes can be analysed readily with no major modification to RCT design or sample size. Furthermore, RCTs should routinely attempt to report the proportion of eligible patients who refused to take part because of their preferences for treatment. Beyond these two general recommendations, our findings also indicate a number of approaches to the design and analysis of RCTs that take account of participants' and/or professionals' preferences.

Design

1. Explanations of the interventions to participants should use written information but alternative methods (e.g. video presentations, explicit preference elicitation) may be useful in some contexts.
2. A careful balance must be struck between providing information on which participants can make informed choices and reducing the sample willing to be randomised.
3. Measurement of preferences should take account of (a) expectancies concerning the process and outcomes associated with the intervention and (b) the perceived value placed on those outcomes and processes.
4. Preferences should be measured (not just elicited) even in comprehensive cohort studies where patients can choose their treatments.
5. Preferences are best measured at baseline and each follow-up point using quantitative methods such as willingness to pay and psychometric approaches based on attitudes.
6. Two-stage randomisation designs are superior to comprehensive cohort designs in that the randomised and preference groups are balanced at baseline. However, they may suffer from poor recruitment because of lack of participant choice and hold no advantage over comprehensive cohort designs if assessment of external validity is paramount.
7. Zelen designs with double randomised consent hold no advantage over comprehensive cohort designs and, because of added complexity and, ethical doubts, should not be used.

Reporting results

1. Baseline data on all important participant characteristics and outcomes must be reported separately for each randomised and preference arm.
2. All participants in all randomised and preference arms must be accounted for at each follow-up.
3. Outcomes for participants in all randomised or preference arms should be reported fully for each outcome at each follow-up point.
4. Where possible, adherence to the intervention should be reported in each RCT arm as this may vary with the form of allocation and provide important extra information on the interaction of preference with treatment arm.

Analysis

1. Analysis of RCTs with preference arms can address two questions:
   (a) Does preference for a particular treatment influence outcome? This means comparing outcome between the randomised and preference patients for each treatment arm.
   (b) Is the treatment effect similar among those with or without preferences? This means comparing the treatment effect in the randomised patients with that in the preference patients.
2. These analyses may be confounded and should be adjusted for baseline values of the outcome variable if applicable and other known prognostic factors.

Recommendations for research

Future research could usefully:

1. Examine the amount and source of information available to patients about interventions in RCTs, with special emphasis on the relationship between sources inside and outside the RCT context. Qualitative research undertaken as part of ongoing RCTs might be especially useful.
2. Examine the processes by which this information leads to preferences in order to develop or extend the proposed expectancy–value framework. Key questions relate to the type of expectancies that enter into decision-making and the way in which different expectancies are valued by patients. Conjoint analysis may be especially useful in this regard.
3. Examine how information about interventions changes participants' preferences and compare the feasibility and effectiveness of different informed consent procedures.
4. Examine how strength of preference varies for different interventions within the same RCT and how these differences can be taken account of in the analysis.
5. Explore the differential effects of patients' and professionals' preferences on evidence arising

from RCTs. Our findings would suggest that patients' preferences act mainly on recruitment. Professionals' preferences appear to affect external and internal validity but the number of RCTs in which professionals' preferences were reported was very small.

6. Assess whether the standardised measurement of preferences within all RCTs (and analysis of the interaction with outcome) would allow the rapid development of a significant evidence base concerning patient preferences, albeit in relation to a single preference design.

Relevance to the NHS

Patient choice

The material discussed in this report highlights many of the current tensions at the heart of the NHS. Although the RCT is widely used to assess the efficacy and cost-effectiveness of new treatments, the design runs counter to the NHS' current emphasis on patient choice. Giving patients more choice about how, when and where they receive treatment is one cornerstone of the Government's health strategy. Furthermore, consumer participation in research is advocated by the Government and most research councils and charities. Although patients must have a choice to enter RCTs, it is often forgotten that they also have preferences about the treatments on offer in the RCT arms. However, provision of effective treatments is also in the interests of consumers. We need to find a balance between offering patient choice, encouraging patient participation in research design and planning and ensuring that treatments offered are based upon the best possible evidence.

The views of professionals

Most RCTs in the UK occur in NHS services. Although exhortations are made to clinicians to increase the proportion of their patients in RCTs, lack of consideration of professionals' preferences hamper these efforts. As we discussed, there are many ways in which professionals can influence recruitment and participation in RCTs. For example, many RCTs continue to be reported, where it is not clear how many people refused to consider recruitment or where professionals avoided discussing the possibility of an RCT at all. This is not because trialists do not make every effort to collect such data; professionals' preferences may not always be obvious. Recruitment to RCTs is often regarded as a burden that interferes with their professional and empathic relationships with patients and their families. One example is seen in RCTs conducted in primary care, where only a proportion of GPs may refer potential participants to the RCT,[143] on occasion leading to the RCT's collapse.[144]

A solution

In addition to understanding more about how professionals' preferences affect internal validity of RCTs and informing professionals and patients about the need for good evidence of efficacy, we need greater application of information systems within the NHS to make use of routine data collection as one source of evidence on effectiveness.

References

1. Sackett D, Rosenberg W, Gray J, Haynes B, Richardson W. Evidence-based medicine: what it is and what it is not. *BMJ* 1996;**312**:71–2.
2. Jadad A. *Randomised controlled trials*. London: BMJ Publishing; 1998. pp. 24–5.
3. Altman D, Bland M. Treatment allocation in controlled trials: why randomise? *BMJ* 1999; **318**:1209.
4. Colditz G, Miller J, Mosteller F. How study design affects outcomes in comparisons of therapy I: medical. *Stat Med* 1989;**8**:441–54.
5. Schulz K, Chalmers I, Hayes R, Altman D. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; **273**:408–12.
6. Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Choosing between randomised and non-randomised studies: a systematic review. *Health Technol Assess* 1998;**2**(13).
7. MacLehose R, Reeves B, Harvey I, Sheldon T, Russell I, Black A. A systematic review of comparisons of effect sizes derived from randomised and non-randomised studies. *Health Technol Assess* 2000;**4**(34).
8. Crow R, Gage H, Hampson S, Hart J, Kimber A, Thomas H. The role of expectancies in the placebo effect and their use in the delivery of health care: a systematic review. *Health Technol Assess* 1999;**3**(3).
9. Orne M. On the social psychology of the psychological experiment: with particular reference to demand characteristics and their implications. *Am Psychol* 1962;**17**:776–83.
10. Adair J. The Hawthorne effect: a reconsideration of the methodological artifact. *J Appl Psychol* 1984;**69**:334–45.
11. Adair J, Sharpe D, Huynh C. Hawthorne control procedures in educational experiments: a reconsideration of their use and effectiveness. *Rev Educ Res* 1989;**59**:215–28.
12. Cook T, Campbell D. *Quasi-experimentation – design and analysis issues for field settings*. Chicago, IL: Rand McNally; 1979.
13. Schulz K, Grimes D. Blinding in randomised trials: hiding who got what. *Lancet* 2002; **359**:696–700.
14. Day S, Altman D. Blinding in clinical trials and other studies. *BMJ* 2000;**321**:504.
15. Moncrieff J, Wessely S, Hardy R. Meta-analysis of trials comparing antidepressants with active placebos. *Br J Psychiatry* 1998;**172**:227–31.
16. Brewin CR, Bradley C. Patient preferences and randomised clinical trials. *BMJ* 1989;**299**:313–15.
17. Parloff, M. Placebo controls in psychotherapy research: a sine qua non or a placebo for research problems? *J Consult Clin Psychol* 1986;**54**:79–87.
18. Basham R. Scientific and practical advantages of comparative design in psychotherapy outcome research. *J Consult Clin Psychol* 1986;**54**:88–94.
19. McPherson K. The best and the enemy of the good: randomised controlled trials, uncertainty, and assessing the role of patient choice in medical decision making. *J Epidemiol Commun Health* 1994; **48**:6–15.
20. Campbell D. Relabeling internal and external validity for applied social scientists. In Trochim W, editor. *Advances in quasi-experimental design and analysis*. San Francisco, CA: Jossey-Bass; 1986. pp. 67–77.
21. Fairhurst K, Dowrick C. Problems with recruitment in a randomised controlled trial of counselling in general practice: causes and implications. *J Health Serv Res Policy* 1996;**1**:77–80.
22. Gotay CC. Accrual to cancer clinical trials: directions from the research literature. *Soc Sci Med* 1991;**33**:569–77.
23. Torgerson D, Klaber-Moffett J, Russell I. Patient preferences in randomised trials: threat or opportunity? *J Health Serv Res Policy* 1996;**1**:194–7.
24. Torgerson D, Sibbald, B. What is a patient preference trial? *BMJ* 1998;**316**:360.
25. Olschewski M, Scheurlen H. Comprehensive cohort study: an alternative to randomised consent design in a breast preservation trial. *Methods Inf Med* 1985;**24**:131–4.
26. King M, Sibbald B, Ward E, Bower P, Lloyd M, Gabbay M, Byford S. Randomised controlled trial of non-directive counselling, cognitive–behaviour therapy and usual general practitioner care in the management of depression as well as mixed anxiety and depression in primary care. *Health Technol Assess* 2000;**4**(19).

27. Ward E, King M, Lloyd M, Bower P, Sibbald B, Farrelly S, *et al.* Randomised controlled trial of non-directive counselling, cognitive-behaviour therapy and usual general practitioner care for patients with depression. I. Clinical effectiveness. *BMJ* 2000;**321**:1383–8.

28. Korn EL, Baumrind S. Randomised clinical trials with clinician-preferred treatment. *Lancet* 1991; **337**:149–52.

29. Zelen M. A new design for randomised clinical trials. *N Engl J Med* 1979;**300**:1242–5.

30. Wennberg JE, Barry MJ, Fowler FJ, Mulley A. Outcomes research, PORTS, and health care reform. *Ann N Y Acad Sci* 1993;**703**:52–62.

31. Rücker G. A two stage trial design for testing treatment, self selection and treatment preference effects. *Stat Med* 1989;**8**:477–85.

32. Feine JS, Awad MA, Lund JP. The impact of patient preference on the design and interpretation of clinical trials. *Commun Dent Oral Epidemiol* 1998;**26**:70–4.

33. McPherson K, Britton AR, Wennberg JE. Are randomised controlled trials controlled? Patient preferences and unblind trials. *J R Soc Med* 1997; **90**:652–6.

34. Chinn S. A simple method for converting an odds ratio to an effect size for use in meta-analysis. *Stat Med* 2000;**19**:3127–31.

35. Cohen J. *Statistical power analysis for the behavioural sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum; 1988.

36. Ashcroft R, Chadwick D, Clark S, Edwards R, Frith L, Hutton J. Implications of socio-cultural contexts for the ethics of controlled trials. *Health Technol Assess* 1997;**1**(9).

37. Pearsall J, editor. *Concise Oxford Dictionary*. Oxford: Oxford University Press; 2001.

38. Drummond M, Stoddart G, Torrance G. *Methods for the economic evaluation of health care programmes*. Oxford: Oxford Medical Publications; 1997.

39. Ajzen I. *Attitudes, personality and behaviour*. Milton Keynes: Open University Press; 1988.

40. Mooney G. *Key issues in health economics*. Hemel Hempstead: Harvester Wheatsheaf; 1994.

41. Wright G. *Behavioural decision theory: an introduction*. Harmondsworth: Penguin; 1984.

42. Crow R, Gage H, Hampson S, Hart J, Kimber A, Thomas H. The role of expectancies in the placebo effect and their use in the delivery of health care: a systematic review. *Health Technol Assess* 1999;**3**(3).

43. Donaldson C, Shackley P. Does 'process utility' exist? a case study of willingness to pay for laparoscopic cholecystectomy. *Soc Sci Med* 1997; **44**:699–707.

44. Sculpher M. The cost-effectiveness of preference-based treatment allocation: the case of hysterectomy versus endometrial resection in the treatment of menorrhagia. *Health Econ* 1998; **7**:129–42.

45. Sculpher M, Gafni A. Recognizing diversity in public preferences: the use of preference sub-groups in cost-effectiveness analysis. *Health Econ* 2001;**10**:317–24.

46. Donaldson C. Eliciting patients' values by use of 'willingness to pay': letting the theory drive the method. *Health Expect* 2001;**4**:180–8.

47. Holmes-Rovner M, Llewellyn-Thomas H, Entwistle V, Coulter A, O'Connor A, Rovner D. Patient choice modules for summaries of clinical effectiveness: a proposal. *BMJ* 2001;**322**:664–7.

48. Scott A, Vick S. Patients, doctors and contracts: an application of principal-agent theory to the doctor–patient relationship. *Scott J Political Econ* 1999;**46**:111–34.

49. Vick S, Scott A. Agency in health care. Examining patients' preferences for attributes of the doctor–patient relationship. *J Health Econ* 1998; **17**:587–605.

50. Ryan M, Farrar S. Using conjoint analysis to elicit preferences for health care. *BMJ* 2000;**320**:1533.

51. Conner M, Sparks P. The theory of planned behaviour and health behaviours. In Conner M, Norman P, editors. *Predicting health behaviour: research and practice with social cognition models*. Buckingham: Open University Press; 1996. pp. 121–62.

52. Bandura A. *Self-efficacy: the exercise of control*. New York: Freeman; 1997.

53. Longo D, Lent R, Brown S. Social cognitive variables in the prediction of client motivation and attrition. *J Couns Psychol* 1992;**39**:447–52.

54. Schwarzer R, Fuchs R. Self efficacy and health behaviours. In: Conner M, Norman P, editors. *Predicting health behaviour: research and practice with social cognition models*. Buckingham: Open University Press; 1996. pp. 163–96.

55. Fischoff B, Goitein B, Shapira Z. The experienced utility of expected utility approaches. In Feather N, editor. *Expectations and actions: expectancy–value models in psychology*. Hillsdale, NJ: Lawrence Erlbaum, 1982. pp. 315–39.

56. Silverman W, Altman D. Patients' preferences and randomised trials. *Lancet* 1996;**347**:171–4.

57. Kerr C, Robinson E, Lilford R, Edwards S, Braunholtz D, Stevens A. The impact of describing clinical trial treatments as new or standard. *Pat Educ Couns* 2003;**53**:107–13.

58. Chard J, Lilford R. The use of equipoise in clinical trials. *Soc Sci Med* 1998;**47**:891–8.

59. Ashok P, Kidd A, Flett G, Fitzmaurice A, Graham W, Templeton A. A randomized comparison of medical abortion and surgical vacuum aspiration at 10–13 weeks gestation. *Hum Reprod* 2003;**17**:92–8.

60. Brookes S, Peters T, Campbell R, Featherstone K, Neal D, Abrams P, *et al*. Including a 'no active intervention' arm in surgical trials is possible: evidence from the CLasP randomised trial. *J Health Serv Res Policy* 2003;**8**:209–14.

61. Hjorth M, Holmberg E, Rodjer S, Taube A, Westin J. Physicians' attitudes towards clinical trials and their relationship to patient accrual in a Nordic multicenter study on myeloma. *Control Clin Trials* 1996;**17**:372–86.

62. Verheggen F, Nieman F, Jonkers R. Determinants of patient participation in clinical studies requiring informed consent: why patients enter a clinical trial. *Pat Educ Couns* 1998;**35**:111–25.

63. Ross S, Grant A, Counsell C, Gillespie W, Russell I, Prescott R. Barriers to participation in randomised controlled trials: a systematic review. *J Clin Epidemiol* 1999;**52**:1143–56.

64. Holmes-Rovner M, Wills C. Improving informed consent: insights from behavioural decision research. *Med Care* 2002;**40**:V-30–V-38.

65. Kahneman D, Tversky A. Prospect theory: analysis of decisions under risk. *Econometrica* 1979; **47**:263–91.

66. Loomes G, Sugden R. Regret theory: an alternative theory of rational choice under uncertainty. *Econ J* 1982;**92**:805–24.

67. Jacobson N, Roberts L, Berns S, McGlinchey J. Methods for defining and determining the clinical significance of treatment effects: description, application, and alternatives. *J Consult Clin Psychol* 1999;**67**:300–7.

68. Donovan J, Mills N, Smith M, Brindle L, Jacoby A, Peters T, *et al*. Improving design and conduct of randomised trials by embedding them in qualitative research: ProtecT (prostate testing for cancer and treatment) study. *BMJ* 2002; **325**:766–70.

69. Halpern S. Prospective preference assessment: a method to enhance the ethics and efficiency of randomized controlled trials. *Control Clin Trials* 2002;**23**:274–88.

70. Chalmers I. What is the prior probability of proposed new treatment being superior to an established treatment? *BMJ* 1997;**314**:74–5.

71. Zimbroff D. Patient and rater education of expectations in clinical trials (PREECT). *J Clin Psychopharmacol* 2001;**21**:251–2.

72. Ubel P. Is information always a good thing? Helping patients make 'good' decisions. *Med Care* 2002;**40**:V-39–V-44.

73. Say R, Thomson R. The importance of patient preferences in treatment decisions – challenges for doctors. *BMJ* 2003;**327**:542–5.

74. Tinsley H, Bowman S, Ray S. Manipulation of expectancies about counselling and psychotherapy: review and analysis of expectancy manipulation strategies and results. *J Couns Psychol* 1988;**35**:99–108.

75. Parloff M. Placebo controls in psychotherapy research: a sine qua non or a placebo for research problems? *J Consult Clin Psychol* 1986;**54**:79–87.

76. Shapiro D, Shapiro D. Meta-analysis of comparative therapy outcome studies: a replication and refinement. *Psychol Bull* 1983; **92**:581–604.

77. O'Connor A, Tugwell P, Wells G, Elmslie T, Jolly E, Hollingworth G, *et al*. A decision aid for women considering hormone therapy after menopause: decision support framework and evaluation. *Pat Educ Couns* 2002;**33**:267–79.

78. O'Connor A, Stacey D, Entwistle V, Llewellyn-Thomas H, Rovner D, Holmes-Rovner M, *et al*. Decision aids for people facing health treatment or screening decisions. *The Cochrane Library*, Issue 4, 2001. Update Software.

79. Ubel P, Loewenstein G. The role of decision analysis in informed consent: choosing between intuition and systemasticity. *Soc Sci Med* 1997; **44**:647–56.

80. Lilford R, Edwards S, Braunholtz D, Jackson J, Thornton J, Hewison J. Ethical issues in the design and conduct of randomised controlled trials. In Stevens A, Abrams K, Brazier J, Fitzpatrick R, Lilford R, editors. *The advanced handbook of methods in evidence based healthcare*. London: Sage Publications, 2001. pp. 11–24.

81. Elwyn G, Charles C. Shared decision making: the principles and the competences. In Edwards A, Elwyn G, editors. *Evidence-based patient choice*. Oxford: Oxford University Press; 2001. pp. 118–43.

82. Lilford R. Ethics of clinical trials from a bayesian and decision analytic perspective: whose equipoise is it anyway? *BMJ* 2003;**326**:980–1.

83. Mills N, Donovan J, Smith M, Jacoby A, Neal D, Hamdy F. Perceptions of equipoise are crucial to trial participation: a qualitative study of men in the ProtecT study. *Control Clin Trials* 2003; **24**:272–83.

84. Fischoff B, Slovic P, Lichtenstein S. Knowing what you want: measuring labile values. In Wallsten T, editor. *Cognitive processes in choice and decision behaviour*. Hillsdale, NJ: Lawrence Erlbaum, 1980; 117–41.

85. Shafir E, LeBoeuf R. Rationality. *Annu Rev Psychol* 2002;**53**:491–517.

86. McPherson K, Chalmers I. Information about patients' preferences must be obtained first [letter]. *BMJ* 1998;**317**:78.

87. Edwards S, Lilford R, Thornton J, Hewison J. Informed consent for clinical trials: in search of the 'best' method. *Soc Sci Med* 2003;**47**:1825–40.

88. Torgerson D, Klaber-Moffett J, Russell I. Patient preferences in randomised trials: threat or opportunity? *J Health Serv Res Policy* 1996;**1**:194–7.

89. Klaber-Moffett J, Torgerson D, Bell-Syer S, Jackson D, Llewlyn-Phillips H, Farrin A, *et al*. Randomised controlled trial of exercise for low back pain: clinical outcomes, costs and preferences. *BMJ* 1999;**319**:279–83.

90. Ryan M, Scott D, Reeves C, Bate A, van Teijlingen ER, Russell EM, *et al*. Eliciting public preferences for healthcare: a systematic review of techniques. *Health Technol Assess* 2001; **5**(5).

91. Brazier J, Deverill M, Green C, Harper R, Booth A. A review of the use health status measures in economic evaluation. *Health Technol Assess* 1999;**3**(9).

92. Streiner D, Norman G. *Health measurement scales – a practical guide to their development and use*. Oxford: Oxford Medical Publications; 1989.

93. Van Dyck R, Spinhoven P. Does preference for treatment matter? A study of exposure in vivo with or without hypnosis in the treatment of panic disorder with agoraphobia. *Behav Modif* 1997; **21**:172–86.

94. Renjilian D, Nezu A, Shermer R, Perri M, McKelvey W, Anton S. Individual versus group therapy for obesity: effects of matching participants to their treatment preferences. *J Consult Clin Psychol* 2001;**69**:717–21.

95. Devine D, Fernald P. Outcome effects of receiving a preferred, randomly assigned or nonpreferred therapy. *J Consult Clin Psychol* 2003;**41**:104–7.

96. Hardy G, Barkham M, Shapiro D, Reynolds S, Rees A. Credibility and outcome of cognitive–behavioural and psychodynamic–interpersonal psychotherapy. *Br J Clin Psychol* 1995;**34**:555–69.

97. Rovers MM. Otitis media with effusion in infants – the effects of ventilation tubes. Proefschrift, 2000; **1**:2.

98. Paradise JL, Bluestone CD, Bachman RZ, Colborn K, Bernard BS, Taylor FH, *et al*. Efficacy of tonsillectomy for recurrent throat infection in severely affected children. *N Engl J Med* 1984; **310**:674–83.

99. Paradise JL, Bluestone CD, Rogers KD, Taylor FH, Colborn K, Bachman RZ, *et al*. Efficacy of adenoidectomy for recurrent otitis media in children previously treated with tympanostomy-tube placement. Results of parallel randomised and nonrandomised trials. *JAMA* 1990; **263**:2066–73.

100. Reddihough DS, King J, Coleman G, Catanese T. Efficacy of programmes based on conductive education for young children with cerebral palsy. *Dev Med Child Neurol* 1998;**40**,763–70.

101. Rovers MM, Straatman H, Ingels K. van der Wilt GJ, van den Broek P, Zielhuis GA. Generalisability of trial results based on randomised versus non-randomised allocation of OME infants to ventilation tubes or watchful waiting. *J Clin Epidemiol* 2001;**54**:789–94.

102. Henshaw RC, Naji SA, Russell IT, Templeton AA. Comparison of medical abortion with surgical vacuum aspiration: women's preferences and acceptability of treatment. *BMJ* 1993;**307**:714–7.

103. Henshaw RC, Naji SA, Russell IT, Templeton AA. A comparison of medical abortion (using miferpristone and gemeprost) with surgical vacuum aspiration: efficacy and early medical sequelae. *Hum Repro*d 1994;**9**:2167–72.

104. Nicolaides K, Brizot ML, Patel F, Snijders R. Comparison of chorionic villus sampling and amniocentesis for fetal karyotyping at 10–13 weeks gestation. *Lancet* 1994;**344**:435–9.

105. Cooper KG, Grant AM, Garratt AM. The impact of using a partially randomised patient preference design when evaluating alternative managements for heavy menstrual bleeding. *Br J Obstet Gynaecol* 1997;**104**:1367–73.

106. Bain C, Cooper KG, Parkin ED. A partially randomised patient preference trial of microwave endometrial ablation using local anaesthesia and intravenous sedation or general anaesthesia: a pilot study. *Gynaecol Endosc* 2001;**10**:223–8

107. Ashok P, Kidd A, Flett GMM, Fitzmaurice A, Graham W, Templeton A. A randomised comparison of medical abortion and surgical vacuum aspiration at 10–13 weeks gestation. *Hum Reprod* 2002;**17**:92–8.

108. Janevic M, Janz N, Dodge J, Lin X, Pan W, Sinco B, Clark N. The role of choice in health education intervention trials: a review and case study. *Soc Sci Med* 2003;**56**:1581–94.

109. Bedi N, Chilvers C, Churchill R, Dewey M, Duggan C, Fielding K, *et al*. Assessing effectiveness of treatment of depression in primary care: a partially randomised preference trial. *Br J Psychiatry* 2000;**177**:312–18.

110. Chilvers C, Dewey M, Fielding K, Gretton V, Miller P, Palmer B, *et al*. Antidepressant drugs and

generic counselling for treatment of major depression in primary care: randomised trial with patient preference arms. *BMJ* 2001;**322**:1–5 [follow-up paper for Bedi study[109]].

111. Rokke PD, Tomhave JA, Jocic Z. The role of client choice and target selection in self-management therapy for depression in older adults. *Psychol Aging* 1999;**14**:155–69.

112 Shapiro DA, Barkham M, Rees A, Hardy GE, Reynolds S, Startup M. Effects of treatment duration and severity of depression on the effectiveness of cognitive–behaviour and psychodynamic–interpersonal psychotherapy. *J Consult Clin Psychother* 1994;**62**:522–34.

113. Helsing M, Bergman B, Thaning L, Hero U. Quality of life and survival in patients with advanced non-small cell lung cancer receiving supportive care plus chemotherapy with carboplatin and etoposide or supportive care only. A multicentre randomised phase III trial. *Eur J Cancer* 1998;**34**:1036–44.

114. Riedl St, Peter B, Geiss H, Aulmann M, Bach A, Lehnert Th. Mikrobiologische und klinische Wirksamkeit der selektiven Darmdekontamination bei der trandthorakalen Resektion von Oesophagus-und Kardiacarcinomen. *Chirug* 2001:**72**:1160–70.

115. Schumacher M, Bastert G, Bojar H, Hubner K, Olschewski M, Beyerle C, *et al*. for the German Breast Cancer Study Group. Randomised "X" trial evaluating hormonal treatment and the duration of chemotherapy in node positive breast cancer patients. *J Clin Oncol* 1994:**12**:2086–93.

116. Schmoor C, Olschewski M, Schumacher M. Randomized and non-randomized patients in clinical trials: experience with comprehensive cohort studies. *Stat Med* 1996:**15**:263–71.

117. Bakker A, Spinhoven P, Balkom AJLM, Vleugel L, van Dyck R. Cognitive therapy by allocation versus cognitive therapy by preference in the treatment of panic disorder. *Psychother Psychosom* 2000; **69**:240–3.

118. Kerry S, Dundas D, Hilton S, Rink E, Patel S, Lord J. Routine referral for radiography of patients presenting with low back pain: is patients outcome influenced by GP's referral for plain radiography? *Health Technol Assess* 2000;**4**(20).

119. Kendrick D, Fielding K, Bentley E, Miller P, Kerslake R, Pringle M. The role of radiography in primary care patients with low back pain of at least 6 weeks duration: a randomised (unblinded) controlled trial. *Health Technol Assess* 2001;**5**(30).

120. Moffett JK, Torgerson D, Bell-Syer S, Jackson D, Llewlyn-Philips H, Farrin A, *et al*. Randomised controlled trial of exercise for low back pain: clinical outcomes, costs and preferences *BMJ* 1999;**319**:279–83.

121. Detre KM, Guo P, Holubkov R, Califf RM, Sopko G, Bach R, *et al*. Coronary revascularisation in diabetic patients: a comparison of the randomised and observational components of the bypass angioplasty revascularisation (BARI). *Circulation* 1999;**99**:633–40.

122. King SB, Lembo NJ, Weintraub WS, Kosinski AS, Barnhart HX, Kutner MH, *et al*. for the Emory Angioplasty versus Surgical Trial (EAST). A randomised trial comparing coronary angioplasty with coronary bypass surgery. *N Engl J Med* 1994: **331**:1044–50.

123. King SB, Lembo NJ, Weintraub WS, Kasinski AS, Barnhart HX, Kutner MH. Emory Angioplasty Versus Surgery Trial (EAST): design, recruitment, and baseline description of patients. *Am J Cardiol* 1995;**70**:42C–59C.

124. King SB, Bernhart HX, Kosinski AS, Weintraub WS, Lembo NJ, Petersen JY, *et al*. Angioplasty or surgery for multivessel coronary artery disease: comparison of eligible registry and randomised patients in the EAST trial and influence of treatment selection on outcomes. *Am J Cardiol* 1997;**79**:1453–9.

125. Olschewski M, Schumacher M, Davis K. Analysis of randomised and nonrandomised patients in clinical trial using the comprehensive cohort follow up study design. *Control Clin Trials* 1992; **13**:226–39.

126. Howie FL, Henshaw RC, Naji SA, Russell IT, Templeton AA. Medical abortion or vacuum aspiration? two year follow up of a patient preference trial. *Br J Obstet Gynaecol* 1997; **104**:829–33 [follow-up to Henshaw paper[103]].

127. Melchart D. Steger HG, Linde K, Makarian K, Hatahet Z, Brenke R, Saller R. Integrating patient preferences in clinical trials: a pilot study of acupuncture vs midazolam for gastroscopy. *J Altern Complement Med* 2002;**8**:265–74.

128. Gossop M, Johns A, Green, L. Opiate withdrawal: inpatient versus outpatient programmes and preferred versus random assignment to treatment. *BMJ* 1986;**293**:103–4.

129. McKay JR, Alterman AI, McLellan T, Boardman CR, Mulvaney FD, O'Brien, CP. Random versus non-random assignment in the evaluation of treatment for cocaine abusers. *J Consult Clin Psychol* 1998;**66**:697–1.

130. McKay JR, Alterman AI, McLellan T, Snider EC, O'Brien CP. Effect of random versus non random assignment in a comparison of inpatient and day hospital rehabilitation for male alcoholics. *J Consult Clin Psychol* 1995;**63**:70–8.

131. Noel PH, Larme AC., Meyer J, Marsh G, Correa A, Pugh JA. Patient choice in diabetes education curriculum. *Diabetes Care* 1998; **21**:896–01.

132. Williams AC, Nicholas MK, Richardson PH, Pither CE, Fernandes J. Generalising from a controlled trial: the effects of patient preference versus randomisation on the outcome of inpatient versus outpatient chronic pain management. *Pain* 1999;**83**:57–65.

133. Campbell D, Stanley J. *Experimental and quasi-experimental evaluations in social research*. Chicago, IL: Rand McNally; 1963.

134. Chalmers TC, Celano P, Sachs HS, Smith H Jr. Bias in treatment assignment in controlled clinical trials. *N Engl J Med* 1983;**309**:1358–61.

135. Sachs H, Chalmers TC, Smith H Jr. Randomised versus historical controls for clinical trials. *Am J Med* 1982;**72**:233–40.

136. Miller J N, Colditz G A, Mosteller F. How study design affects outcome in comparison to therapy. II. Surgical. *Stat Med* 1989;**8**:455–66.

137. Benson K, Hartz A. A comparison of observational studies and randomised controlled trials. *N Engl J Med* 2000;**342**:1878–86.

138. McKee M, Britton A, Black N, McPherson K, Sanderson C, Bain C. Interpreting the evidence: choosing between randomised and non-randomised studies. *BMJ* 1999;**319**:312–15.

139. Kunz R, Oxman A D. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* 1998;**317**:1185–90.

140. Green S B, Byar D P. Using observational data from registries to compare treatments: the fallacy of omnimetrics. *Stat Med* 1984;**3**: 361–73.

141. Pazart L. Will you have a glass of red or white wine? *Lancet* 1999;**353**:1448.

142. Pawson R, Tilley N. *Realistic evaluation*. Sage: London; 1997.

143. Fairhurst K, Dowrick C. Problems with recruitment to a randomised controlled trial of counselling in general practice: causes and implications. *J Health Serv Res Policy* 1996;**1**:77–80.

144. Friedli K, King M, Lloyd M, Horder J. Randomised controlled assessment of non-directive psychotherapy versus routine general practitioner care. *Lancet* 1997;**350**: 1662–5.

Appendix 1

Trial descriptions

No.	Study	Location[a]	Population	Intervention and comparison group(s)[b]	Study design	Primary outcome	Primary outcome identified by	Follow-up points for primary outcome	Description of how preference elicited	Description of how preference quantified
1	Ashok, 2002[107]	Hospital (2)	Women of 10–13 weeks gestation	1. Medical termination (mifepristone, misoprostol) 2. Surgical termination (vacuum aspiration)	Comprehensive cohort	Number of days bleeding	Reviewer	After intervention, 2–3 weeks, 8 weeks	Participants expressing strong preferences for a treatment were given it	Questionnaire (no details reported)
2	Bain, 2001[106]	Hospital (2)	Women with heavy menstrual bleeding	1. Local anaesthesia 2. General anaesthesia For microwave endometrial ablation	Comprehensive cohort	Treatment acceptability: semantic differential scale (12 items)	Reviewer	Postoperation	Women informed about both methods and chose if preference expressed	Not reported
3	Bakker, 2000[117]	Outpatients' clinics (2)	Patients with panic with or without agoraphobia	1. Cognitive therapy[c] 2. Paroxetine 3. Clomipramine 4. Placebo	Comprehensive cohort	Panic frequency	Reviewer	12 weeks	Patients who refused randomisation given preferred treatment	Not reported
4	Bedi, 2000 Chilvers, 2001[109,110]	GP practices (1)	Patients with depression	1. Counselling 2. Antidepressants	Comprehensive cohort	BDI	Authors	8 weeks 1 year	Patients who refused randomisation given preferred treatment	Not reported
5	Detre, 1999[121]	Hospitals (2)	Coronary disease patients with diabetes	1. PTCA 2. CABG	Comprehensive cohort	1. All-cause mortality 2. Cardiac cause mortality	Authors	5 years	Eligible but non-consenting patients entered into registry. Physician preference noted as factor	Not reported
6	Gossop, 1986[128]	Hospital (2)	Opiate users	1. Outpatient (counselling & oral methadone) 2. Inpatient (oral methadone)	Comprehensive cohort	Abstinence from opiates	Authors	*Inpatient* 21 days *Outpatient* 8 weeks	Patients asked for treatment preference and their preference offered first	Not reported

continued

No.	Study	Location[a]	Population	Intervention and comparison group(s)[b]	Study design	Primary outcome	Primary outcome identified by	Follow-up points for primary outcome	Description of how preference elicited	Description of how preference quantified
7	Helsing, 1998[113]	Hospital (2)	Non-small cell lung cancer patients	1. CT 2. BSC	Comprehensive cohort	Global QoL: scale	Authors	4, 8, 12, 16, 20, 24 weeks	Some centres unable to obtain patient consent allocated on preference basis	Not reported
8	Henshaw, 1993, 1994, Howie, 1997[102,103,126]	Hospital (2)	Women <64 days gestation	1. Medical termination (mifepristone and gemeprost) 2. Surgical termination (vacuum aspiration)	Comprehensive cohort	1. Method that women would choose to undergo in future 2. Perception of procedure on semantic differential scale ('good' to 'bad') 3. Anxiety subscale of HADS 4. Fall in self-esteem measure	Authors	16 days, 2 years	Written information. Women expressing a clear preference offered their choice	Not reported
9	Kendrick, 2001[119]	GP practice (1)	GP consulters with low back pain	1. Lumbar spine radiograph 2. No X-ray	Comprehensive cohort	1. Roland disability score 2. Pain score	Authors	3, 9 months	Patients who did not consent to randomisation could choose treatment	Not reported

continued

No.	Study	Location[a]	Population	Intervention and comparison group(s)[b]	Study design	Primary outcome	Primary outcome identified by	Follow-up points for primary outcome	Description of how preference elicited	Description of how preference quantified
10	Kerry, 2000[118]	GP practices (1)	GP consulters with low back pain	1. X-ray with standard treatment 2. Standard treatment alone	Comprehensive cohort	1. Roland disability score 2. No. of consultations for back pain	Authors	6 weeks, 1 year	Patients or physicians who refused randomisation invited to participate in observational study	Not reported
11	SB King, 1994,1995, 1997[122–124]	Hospital (2)	Patients with multivessel coronary artery disease	1. PTCA 2. CABG	Comprehensive cohort	1. All-cause mortality 2. Occurrence of angina	Reviewer	6, 12, 18, 24, 30, 36 months – randomised cohort 12, 24, 36 months – preference cohorts	Physician preference main factor in randomisation refusal	Not reported
12	M King, 2000[26] Ward 2000[27]	GP practices (1)	Patients with depression or mixed anxiety and depression	1. Cognitive behavioural[d] therapy 2. Non-directive counselling[d] 3. GP care (usual care)	Comprehensive cohort Two randomised cohorts: 1. Randomisation 3-way (CBT, NDC, GP) 2. Randomisation 2-way (CBT, NDC) 3. Preference cohort in 3 groups (CBT, NDC, GP[e])	BDI	Authors	4, 12 months	Written information Offered choice or randomisation Encouraged to accept randomisation	Not reported

continued

No.	Study	Location[a]	Population	Intervention and comparison group(s)[b]	Study design	Primary outcome	Primary outcome identified by	Follow-up points for primary outcome	Description of how preference elicited	Description of how preference quantified
13	McKay, 1995[130]	Hospital (2)	Male alcoholics	1. Day hospital outpatients. 2. Inpatient rehabilitation Both 28-day rehabilitation programmes	Comprehensive cohort	Number of drinking days in last 30 days	Reviewer	3, 6, 12 months	Participants offered randomisation Refusers allocated to preferred treatment	Not reported
14	McKay, 1998[129]	Hospital (2)	Cocaine users	1. Day hospital outpatient 2. Inpatient Both 28-day, 12-step oriented, rehabilitation programmes	Comprehensive cohort	Days of cocaine use	Reviewer	3, 6, 12 months	Not reported	Not reported
15	Melchart, 2002[127]	Hospital (2)	Gastroscopy patients	1. Acupuncture 2. Intravenous midazolam	Comprehensive cohort	Perception of the examination: visual analogue scale	Authors	Immediately post-treatment	Patients given detailed information First offered randomisation	Not reported
16	Nicolaides, 1994[104]	Hospital (2)	Pregnant women (1st trimester)	1. Early amniocentesis 2. Chorionic villus sampling	Comprehensive cohort	Total foetal loss Spontaneous loss Termination	Authors	End of pregnancy	Procedures explained and women offered randomisation or choice	Not reported
17	Olschewski, 1992[125]	Hospital (2)	People with Ischaemic heart disease	1. Medical treatment 2. Surgical treatment	Comprehensive cohort	All-cause mortality	Authors	5 years	Physician preference noted as factor in randomisation refusal	Not reported
18	Paradise, 1984[98]	Children's hospital (2)	Children with recurrent throat infection	1. Surgery: tonsillectomy or tonsillectomy with adenoidectomy 2. Control: not defined. presumed no surgery	Comprehensive cohort	Number of throat infections: episodes	Reviewer	1, 2, 3 years	Parent (or child) given detailed information. Those withholding consent for randomisation given choice	Not reported

continued

No.	Study	Location[a]	Population	Intervention and comparison group(s)[b]	Study design	Primary outcome	Primary outcome identified by	Follow-up points for primary outcome	Description of how preference elicited	Description of how preference quantified
19	Paradise, 1990[99]	Children's hospital (2)	Children with recurrent otitis media	1. Surgery: adenoidectomy in all cases. 2. Control: not defined. Assumed no treatment beyond initial medication	Comprehensive cohort	Occurrence of otitis media: 1. Proportion of time with otitis media 2. No. of episodes	Reviewer	1, 2, 3 years	Parents (or child) given written information First offered randomisation	Not reported
20	Reddihough, 1998[100]	Outpatient clinic (2)	Young children with cerebral palsy	1. CE; educational programme to develop motor and cognitive skills in tandem 2. Traditional neuro-developmental programme. No emphasis on motor coordination (control)	Comprehensive cohort	1. Cognitive function: VAB 2. Total score of the GMFM	Reviewer	6 months	Children's parents who refused randomisation given choice	Not reported
21	Riedl, 2001[114]	Hospital (2)	Patients with carcinoma of oesophagus	1. SDD 2. Control/no SDD (intravenous antibiotic prophylactic only)	Comprehensive cohort	Occurrence of pneumonia	Reviewer	Post-operation	Patients given choice. If unable to decide were randomised	Not reported
22	Rovers, 2000,[97] 2001[101]	Outpatients' clinics (2)	Infants with persistent bilateral otitis media with effusion	1. Ventilation tubes 2. Watchful waiting (control)	Comprehensive cohort	1. Occurrence of bilateral otitis media during the trial period 2. Time spent with effusion over 1-year period	Reviewer	3, 6, 9, 12 months	Parents who refused randomisation given treatment of their or surgeon's choice	Not reported

continued

No.	Study	Location[a]	Population	Intervention and comparison group(s)[b]	Study design	Primary outcome	Primary outcome identified by	Follow-up points for primary outcome	Description of how preference elicited	Description of how preference quantified
23	Schumacher, 1994[115]	Hospital (2)	Women with primary breast cancer	Factorial design 1. 3 cycles of CT 2. 3 cycles of CT + tamoxifen 3. 6 cycles of CT 4. 6 cycles of CT + tamoxifen	Comprehensive cohort	1. Tumour recurrence 2. All-cause mortality	Author	5 years	Women who refused randomisation given choice	Not reported
24	Schmoor, 1996[116]	Hospital (2)	Women with primary breast cancer	1. 6 cycles of CT 2. 6 cycles of CT + radiotherapy	Comprehensive cohort	All-cause mortality	Author	5 years	Women who refused randomisation given choice	Not reported
25	Williams, 1999[132]	Hospital (2)	Chronic pain sufferers	1. Outpatient 2. Inpatient Both groups received CBT for chronic management of pain	Comprehensive cohort	Sickness impact profile	Author	1, 12 months	Patients who refused randomisation given choice	Not reported
26	Cooper, 1997[105]	Hospital (2)	Women with heavy menstrual bleeding	1. Transcervical resection of the endometrium 2. Medical care: standard medical treatment, clinician prescribed medication he/she felt appropriate	Two-stage randomised clinical trial (modified)	1. Treatment acceptability 2. Willingness to continue same treatment	Reviewer	4 months	Written information given. Women in choice group asked if had strong preference	Not reported
27	Janevic, 2003[108]	Hospital (2)	Women with cardiovascular conditions	Disease management programmes: 1. Self-directed, home-based[d] 2. Group programme[d] 3. Usual care	Two-stage randomised clinical trial (modified)	1. Attendance of at least one unit of training 2. Completion of at least one level	Authors	6 weeks (end of programme)	Randomisation pre-consent. Written information, followed by scripted telephone call Patients in preference cohort asked to choose a programme	Not reported

continued

No.	Study	Location[a]	Population	Intervention and comparison group(s)[b]	Study design	Primary outcome	Primary outcome identified by	Follow-up points for primary outcome	Description of how preference elicited	Description of how preference quantified
28	Noel, 1998[131]	Ambulatory care centre (2)	Diabetic patients	Educational packages: 1. Nutritional: 60% nutritional, 40% non-nutritional 2. Standard: (usual care) 60% non-nutritional management, 40% nutritional	Two-stage randomised clinical trial	1. Class attendance 2. Knowledge of diabetes	Authors	6 months	Written, neutral descriptions of interventions given to those in 'Choice' group and they were allowed to choose	Not reported
29	Rokke, 1999[111]	Ambulatory care (2)	Depressed older adults	1. SMT – cognitive focus[d] 2. SMT – behavioural focus[d] 3. Waiting list controls	Two-stage randomised clinical trial	1. Hamilton Rating Scale for Depression 2. BDI 3. Geriatric Depression Scale 4. Expectancy measures (credibility questionnaire, depression beliefs questionnaire)	Author	10 weeks, 1, 3 months, 1 year (for subset only) For expectancy: 5, 10 weeks	For choice condition group therapist offered detailed description and discussed treatment choice	Not reported
30	Hardy, 1995[96] Shapiro, 1994[112]	Secondary care (2)	Patients with depression	1. Psychodynamic – interpersonal psychotherapy, 8 or 16 sessions. 2. Cognitive–behavioural therapy, 8 or 16 sessions	Preference measurement at baseline	BDI	Reviewer	16, 28, 32 weeks	Participants rated initial credibility of treatments	A four-item treatment credibility measure (each item 1 = not at all to 7 = very)

continued

No.	Study	Location[a]	Population	Intervention and comparison group(s)[b]	Study design	Primary outcome	Primary outcome identified by	Follow-up points for primary outcome	Description of how preference elicited	Description of how preference quantified
31	Moffett, 1999[120]	GP practices (1)	Patients with mechanical low back pain of >4 weeks or <6 months' duration	1. Exercise classes (in groups of 10), eight 1-h sessions over 4 weeks (a cognitive–behavioural approach was used) 2 Standard GP care for low back pain (control)	Preference measurement at baseline	1. Score on RDS 2. Improvement by at least 3 points on the RDS	Author	6 weeks, 6 months, 1 year	Patients asked to state their preference before being told the treatment condition to which they had been randomised	
32	Devine, 1973[95]	Psychology laboratory	Psychology students with extreme fear of snakes	1. Systematic desensitisation 2. Encounter 3. Rational emotive 4. Combination of modelling and behavioural rehearsal	Other unusual designs	14-point behavioural rating scale indicating fear of snakes	Author	1 week after the second therapy session	*For preference and non-preference cohorts only*: participants viewed videotaped descriptions (5 minutes) and demonstrations (5 minutes) of how the four therapies work	After viewing tapes, each therapy rated on 5-point scale (results not provided). Subsequently interviewed to ensure accurate expression of preferences
33	Renjilian, 2001[94]	Psychology clinic (2)	Obese adults	1. Individual therapy 2. Group therapy Both 26 weekly sessions of standard cognitive behavioural weight management Low-calorie diet and exercise programme	Other unusual designs	1. Body weight 2. Body mass index	Authors	26 weeks	Written information. Patients asked to indicate preference before random assignment	Six-point Likert Scale 1 = 'strongly prefer group' and 6 = 'strongly prefer individual'

continued

No.	Study	Location[a]	Population	Intervention and comparison group(s)[b]	Study design	Primary outcome	Primary outcome identified by	Follow-up points for primary outcome	Description of how preference elicited	Description of how preference quantified
34	Van Dyck, 1997[93]	Outpatient clinic (2)	Agoraphobics with panic disorder	1. Combined: 4 h therapist contact, 10 h self-hypnosis practice, 7.5 h exposure *in vivo* 2. *In vivo*: 21.5 h therapy over 4 weeks, 4 h with therapist, 17.5 h exposure practice	Other unusual designs	Quantitative measurement of tolerance to trigger	Authors	2 weeks (intermediate), 4 weeks (post-test)	Written information. Patients indicated their preference prior to randomisation	Visual analogue scale to indicate strength of preference

BDI, Beck Depression Inventory; BSC, best supportive care; CE, conductive education; CT, chemotherapy; GMFM, Gross Motor Function Measure; HADS, Hospital Anxiety and Depression Scale; NDC, non-directive counselling; RDS, Roland Disability Score; SDD, selective bowel decontamination; SMT, self-management therapy; VAB, Vaple Assessment Battery.

[a] Primary (1) or secondary (2) care.

[b] Intervention group is listed first.

[c] Preference group received cognitive therapy only.

[d] Preference group received first two interventions only.

[e] Excluded from analyses as too small a sample.

Appendix 2

Quality of trials

No.	Study	Primary outcome identified by	Power calculation	Adequate concealment of randomisation[a]	Blinded to treatment[b]	Participation rate	Numbers in each arm (R vs P)	Percentage accepting randomisation	All participants accounted for	% with missing data on primary outcome	All patients included in the analyses of primary outcome
1	Ashok, 2002[107]	Reviewer	Yes	OK	P = not possible HW = not possible RA = not clear	579 eligible, 134 declined participation 445 (77%) participated	R = 368 P = 77	83	Yes	*At 2 weeks* R1 = 37 R2 = 38 P1 = 40 P2 = 42	No
2	Bain, 2001[106]	Reviewer	No	OK	P = not possible HW = not possible RA = not clear	114 eligible, 98 (86%) participated	R = 36 P = 62	37	Yes	None	N/A
3	Bakker, 2000[117]	Reviewer	No	Unclear	P = not possible HW = not possible RA = No	Not clear	R = 131 P = 31	81	Yes	R = 26 P = 23	Yes (last observation carried forward)
4	Bedi, 2000 Chilvers, 2001[109,110]	Authors	Yes	OK	P = not possible HW = not possible RA = not clear	425 general practices approached 31 (7.3%) enrolled >1 patient	R = 103 P = 220	32	Yes	*At 8 weeks* R1 = 14 R2 = 25 P1 = 30 P2 = 23 *At 1 year* R1 = 39 R2 = 35 P1 = 42 P2 =31	No (but bias due to missing data investigated)
5	Detre, 1999[121]	Authors	Yes	Unclear	P = not possible HW = not possible RA = not clear	100% of identified eligible patients	R = 1829 P = 2010 (R = 353 and P = 339 for diabetic patients)	48	Yes	Not clear (likely to be none)	N/A

continued

No.	Study	Primary outcome identified by	Power calculation	Adequate concealment of randomisation[a]	Blinded to treatment[b]	Participation rate	Numbers in each arm (R vs P)	Percentage accepting randomisation	All participants accounted for	% with missing data on primary outcome	All patients included in the analyses of primary outcome
6	Gossop, 1986[128]	Authors	No	Unclear	P = not possible HW = not possible RA = not clear	Not reported	R = 20 P = 40	33	Yes	None	N/A
7	Helsing, 1998[113]	Authors	No	Unclear	P = not possible HW = not possible RA = not clear	Not reported	R = 49 P = 102	32	Yes	*4 weeks* R1 = 19 R2 = 18 *8 weeks* R1 = 42 R2 = 32 *12 weeks* R1 = 62 R2 = 32 *16 weeks* R1 = 65 R2 = 41 *20 weeks* R1 = 65 R2 = 41 *24 weeks* R1 = 77 R2 = 55	No
8	Henshaw, 1993, 1994, Howie, 1997[102,103,126]	Authors	Yes	OK	P = not possible HW = not possible RA = yes	373 eligible 363 (97.3%) took part	R = 195 P = 168	54	Yes	*At 2 weeks* None *At 2 years* R1 = 61 R2 = 60 P1 = 63 P2 = 62	N/A No

continued

No.	Study	Primary outcome identified by	Power calculation	Adequate concealment of randomisation[a]	Blinded to treatment[b]	Participation rate	Numbers in each arm (R vs P)	Percentage accepting randomisation	All participants accounted for	% with missing data on primary outcome	All patients included in the analyses of primary outcome
9	Kendrick, 2001[119]	Authors	Yes	OK	P = not possible HW = not possible RA = no	52/73 (71%) GP practices recruited patients 476 participated	R = 421 P = 55	88	Yes	*At 3 months* R1 = 5 R2 = 4 P1 = 6 P2 = 4 *At 9 months* R1 = 7 R2 = 6 P1 = 6 P2 = 9	No
10	Kerry, 2000[118]	Authors	Yes	OK	P = not possible HW = not possible RA = no	303 GP practices approached 126 (41.6%) took part 659 eligible patients 580 (88%) participated	R = 153 P = 427	26	No	*Disability at 6 weeks* R1 = 19 R2 = 16 P1 = 20 P2 = 17 *Disability at 1 year* R1 = 37 R2 = 29 P1 = 34 P2 = 23 *Consults at 6 weeks* R1 = 5 R2 = 11 P1 = 4 P2 = 6 *Consults at 1 year* R1 = 5 R2 = 11 P1 = 4 P2 = 6	Not clear

continued

No.	Study	Primary outcome identified by	Power calculation	Adequate concealment of randomisation[a]	Blinded to treatment[b]	Participation rate	Numbers in each arm (R vs P)	Percentage accepting randomisation	All participants accounted for	% with missing data on primary outcome	All patients included in the analyses of primary outcome
11	SB King, 1994,1995, 1997[122–124]	Reviewer	Yes	Unclear	P = not possible HW = not possible RA = not clear	Not reported	R = 392 P = 450	47	Yes	None	N/A
12	M King, 2000[26] Ward, 2000[27]	Authors	Yes	OK	P = not possible HW = not possible RA = no	624 referred 464 (74.4%) participated	R = 327 P = 137	70	Yes	*At 4 months* R1 = 13 R2 = 11 P1 = 16 P2 = 4 *At 12 months* R1 = 20 R2 = 19 P1 = 19 P2 = 26	Yes (last observation carried forward)
13	McKay, 1995[130]	Reviewer	No	Unclear	P = not possible HW = not possible RA = not clear	144 eligible All (100%) participated	R = 48 P = 96	33	Yes	*At 3 months* R1 = 21 R2 = 17 P1 = 7 P2 = 35 *At 6 months* R1 = 17 R2 = 12 P1 = 11 P2 = 23 *At 12 months* R1 = 17 R2 = 17 P1 = 12 P2 = 26	No

continued

No.	Study	Primary outcome identified by	Power calculation	Adequate concealment of randomisation[a]	Blinded to treatment[b]	Participation rate	Numbers in each arm (R vs P)	Percentage accepting randomisation	All participants accounted for	% with missing data on primary outcome	All patients included in the analyses of primary outcome
14	Mckay, 1998[129]	Reviewer	No	Unclear	P = not possible HW = not possible RA = not clear	Not clear	R = 115 P = 56	67	No	*At 3 months* R1 = 3 R2 = 11 P1 = 8 P2 = 5 At 6 *months* R1 = 9 R2 = 7 P1 = 0 P2 = 11 *At 12 months* R1 = 11 R2 = 14 P1 = 11 P2 = 5	Not clear
15	Melchart, 2002[127]	Authors	Yes	OK	P = not possible HW = not possible RA = no	108 eligible All (100%) participated	R = 28 P = 80	26	Yes	R1 = 14 R2 = 0 R3 = 0 R4 = 0	No
16	Nicolaides, 1994[104]	Authors	Yes	OK	P = not possible HW = not possible RA = not clear	1870 eligible 1301 (70%) participated	R = 488 P = 813	38	Yes	R1 = 0 R2 = 0 P1 = 0 P2 = 0.2	N/A
17	Olschewski, 1992[125]	Authors	Yes	OK	P = not possible HW = not possible RA = not clear	2099 were eligible to take part all (100%) participated	R = 780 P = 1319	37	Yes	None	N/A

continued

No.	Study	Primary outcome identified by	Power calculation	Adequate concealment of randomisation[a]	Blinded to treatment[b]	Participation rate	Numbers in each arm (R vs P)	Percentage accepting randomisation	All participants accounted for	% with missing data on primary outcome	All patients included in the analyses of primary outcome
18	Paradise, 1984[98]	Reviewer	No	Unclear	P = not possible HW = not possible RA = not clear	187 eligible All (100%) all took part	R = 91 P = 96	49	Yes	*Year 1* R1 = 12 R2 = 27 P1 = 15 P2 = 23 *Year 2* R1 = 28 R2 = 40 P1 = 35 P2 = 36 *Year 3* R1 = 49 R2 = 58 P1 = 71 P2 = 70	No (but bias due to missing data investigated)
19	Paradise, 1990[99]	Reviewer	No	Unclear	P = not possible HW = not possible RA = no	213 eligible All (100%) took part	R = 99 P = 114	46	Yes	*Year 1* R1 = 8 R2 = 19 P1 = 17 P2 = 9 *Year 2* R1 = 13 R2 = 43 P1 = 32 P2 = 24 *Year 3* R1 = 29 R2 = 68 P1, P2 = not reported	No (but bias due to missing data investigated)

continued

No.	Study	Primary outcome identified by	Power calculation	Adequate concealment of randomisation[a]	Blinded to treatment[b]	Participation rate	Numbers in each arm (R vs P)	Percentage accepting randomisation	All participants accounted for	% with missing data on primary outcome	All patients included in the analyses of primary outcome
20	Reddihough, 1998[100]	Reviewer	No	Unclear	P = not possible HW = not possible RA = yes	69 eligible 66(96%) participated	R = 34 P = 32	52	Yes	*VAB* R1 = 0 R2 = 0 P1 = 13 P2 = 24 *GMFM* R1 = 47 R2 = 47 P1 = 47 P2 = 35	No
21	Riedl, 2001[114]	Reviewer	Yes	Unclear	P = not possible HW = not possible RA = not clear	Not reported	R = 28 P = 42	40	Yes	None	N/A
22	Rovers, 2000, 2001[97,101]	Reviewer	No	OK	P = not possible HW = not possible RA = not clear	386 eligible 320 (83%) participated	R = 187 P = 133	58	Yes	*At 3 months* R1 = 0 R2 = 1 P1 = 3 P2 = 3 *At 6 months* R1 = 4% R2 = 3% P1 = 14% P2 = 9% *At 9 months* R1 = 9% R2 = 13% P1 = 23% P2 = 18% *At 12 months* R1 = 3% R2 = 9% P1 = 23% P2 = 20%	No

continued

No.	Study	Primary outcome identified by	Power calculation	Adequate concealment of randomisation[a]	Blinded to treatment[b]	Participation rate	Numbers in each arm (R vs P)	Percentage accepting randomisation	All participants accounted for	% with missing data on primary outcome	All patients included in the analyses of primary outcome
23	Schumacher, 1994[115]	Authors	Yes	Unclear	P = not possible HW = not possible RA = not clear	Not reported	R = 473 P = 247	66	Yes	None	N/A
24	Schmoor, 1996[116]	Authors	Yes	Unclear	P = not possible HW = not possible RA = not clear	Not reported	R = 189 P = 129	59	Yes	None	N/A
25	Williams, 1999[132]	Authors	No	Vulnerable	P = not possible HW = not possible RA = not clear	412 eligible 249 (60%) participated	R = 121 P = 128	49	No	Not clear	Not clear
26	Cooper, 1997[105]	Reviewer	Yes	OK	P = not possible HW = not possible RA = not clear	273 eligible 138 allocated to random cohort 135 allocated to choice cohort	R = 187 P = 40	Rücker design 40 (30%) chose treatment 97 (70%) accepted randomisation and additional 90 in choice cohort were randomised	Yes	None	N/A
27	Janevic, 2003[108]	Authors	No	Unclear	P = not possible HW = not possible RA = not clear	3079 eligible 1613 randomised allocation 1466 choice allocation	R = 575 P = 553	"Modified Rücker" design. Randomisation pre-consent. 575 (36%) accepted random allocation 553 (38%) accepted choice allocation	Yes	None	N/A

continued

No.	Study	Primary outcome identified by	Power calculation	Adequate concealment of randomisation[a]	Blinded to treatment[b]	Participation rate	Numbers in each arm (R vs P)	Percentage accepting randomisation	All participants accounted for	% with missing data on primary outcome	All patients included in the analyses of primary outcome
28	Noel, 1998[131]	Authors	No	OK	P = not possible HW = not possible RA = not clear	596 eligible 291 allocated to random cohort 305 allocated to choice cohort	R = 291 P = 305	Wennberg design Unclear. Implies all accepted allocation	Yes	*Attendance* None lost *Diabetes knowledge* R and P (overall) = 28%	Attendance: N/A Diabetes knowledge: Yes (replaced missing with baseline values) Dropouts and completers compared
29	Rokke, 1999[111]	Authors	No	Unclear	P = not possible HW = not possible RA = yes	64 eligible 29 allocated to waiting list 15 allocated to choice 20 allocated to randomisation	R = 20 P = 15 Control = 29	Wennberg design (modified) Unclear. Implies all accepted allocation	Yes	R1 = 73 R2 = 78 R3 = 21 P1 = 17 P2 = 33	No: drop-outs replaced in randomised arm and drop-outs compared with completers
30	Hardy, 1995[96] Shapiro, 1994[112]	Reviewer	Yes	Unclear	P = not possible HW = not possible RA = no	Not clear	R = 120 P = N/A	Unclear	Yes	*At 16 weeks* R1 = 3 R2 = 0 R3 = 0 R4 = 3 *At 28 weeks* R1 = 0 R2 = 3 R3 = 7 R4 = 0	Not clear

continued

No.	Study	Primary outcome identified by	Power calculation	Adequate concealment of randomisation[a]	Blinded to treatment[b]	Participation rate	Numbers in each arm (R vs P)	Percentage accepting randomisation	All participants accounted for	% with missing data on primary outcome	All patients included in the analyses of primary outcome
31	Moffett, 1999[120]	Authors	Yes	OK	P = not possible HW = not possible RA = yes	441 potentially eligible 187 (42%) participated	R = 187 P = N/A	Unclear	Yes	*At 6 weeks* R1 = 4 R2 = 4 *At 6 months* R1 = 13 R2 = 12 *At 1 year* R1 = 7 R2 = 10	Not clear
32	Devine, 1973[95]	Authors	No	Unclear	P = not possible HW = not possible RA = not clear	74 selected 48 eligible All (100%) participated	R =16 P = 16 NP = 16	The 16 who comprised the randomised group were selected from the study participants	Yes	None	N/A
33	Renjilian, 2001[94]	Authors	No	Unclear	P = not possible HW = not possible RA = no	135 screened 96 eligible and willing to participate 75 (78%) participated	R = 75	Unclear	Yes	$R1_P$ = 32 $R1_{NP}$ = 6 $R2_P$ = 20 $R2_{NP}$ = 30	No
34	Van Dyck, 1997[93]	Authors	No	Unclear	P = not possible HW = not possible RA = yes	Not reported	R = 64	Not reported	Yes	Overall = 19	Drop-outs replaced in both arms

GMFM, Gross Motor Function Measure; NP, non-preference group; VAB, Vaple Assessment Battery.
[a] OK: explicitly reported measures that demonstrate valid randomisation; Unclear: insufficient information; Vulnerable: explicitly stated measures known to be vulnerable.
[b] HW, health worker; P, participant; RA, research assessor.

Appendix 3

Baseline comparisons between randomised and preference cohorts

No.	Study	Number of baseline comparisons made and significant differences found (R_{total} vs P_{total} only)	Numbers in each arm (R vs P)	Age (years)	Gender	Education	Social class
1	Ashok, 2002[107]	**6 comparisons made** No significant differences reported	R = 368 P = 77	Mean (SD) R = 25.2 (6.8) P = 26.6 (6.5)	Not applicable	Not reported	Not reported
2	Bain, 2001[106]	**10 comparisons made** No significant differences reported	R = 36 P = 62	Mean (SD) R = 42.5 (3.9) P = 43.5 (5.3)	Not applicable	Not reported	Not reported
3	Bakker, 2000[117]	**19 comparisons made** No significant differences reported	R = 35 P = 31	Mean (SD) not reported for P group Difference reported as NS in paper	Men/women not reported for P group Difference reported as NS in paper	Not reported	Not reported
4	Bedi, 2000[109] Chilvers, 2001[110]	**13 comparisons made** No significant differences reported	R = 103 P = 220	Mean (SD) R = 37.8 (11.5) P = 37.3 (11.5)	Male R = 23% P = 19%	Not reported	Class I or II R = 31% P = 37%
5	Detre, 1999[121]	**36 comparisons made** 7 significant differences found: 1. Ethnicity – P group higher % white 2. Education – P group higher (more educated) 3. Smoking – P group lower (smoke less) 4. Activity level – P group higher (more active) 5. Congestive heart failure – P group report this less frequently 6. Fair/poor QoL – P group lower (better QoL) 7. Ejection fraction – better left ventricular fraction in P	R = 353 P = 339	Mean (no SD given) R = 62.3 P = 62.5	Male R = 57% P = 55%	High school or less R = 78% P = 67%	Not reported
6	Gossop, 1986[128]	**No comparisons made** No data given for separate cohorts	R = 20 P = 40	Not reported	Not reported	Not reported	Not reported
7	Helsing, 1998[113]	**7 comparisons made** No significant differences reported	R = 49 P = 102	Not reported	Not reported	Not reported	Not reported
8	Henshaw, 1993, 1994,[102,103] Howie 1997[126]	**10 comparisons made** No significant differences reported	R = 195 P = 168	Mean (SD) R = 24.8 (8.1) P = 24.7 (8.6)	Not applicable	Not applicable	Not reported
9	Kendrick, 2001[119]	**17 comparisons made** 2 significant differences: 1. Gender – P group higher % male 2. Roland score – P group higher (more disabled)	R = 421 P = 55	Median R = 39.0 P = 38.5	Male R = 41.3% P = 61.8%	Above 'O' level R: 72.2% P: 63.6%	Not reported

continued

No.	Study	Number of baseline comparisons made and significant differences found (R_{total} vs P_{total} only)	Numbers in each arm (R vs P)	Age (years)	Gender	Education	Social class
10	Kerry, 2000[118]	**31 comparisons made** 2 significant differences found: 1. Bodily pain score – P group higher 2. General health score – P group higher (better health) (Both SF-36 items)	R = 153 P = 427	Mean (SD) R = 43.6 (17.3) P = 42.8 (15.5)	Male R = 50.0% P = 45.9%	Not reported	Class I and II R = 24.3% P = 26.6%
11	SB King, 1994, 1995, 1997[122–124]	**29 comparisons made** 7 significant differences found: 1. No of narrowings ≥ 70 in diameter – P group had higher occurrence 2. Totally occluded vessel – P group had higher occurrence 3. CCS class III or IV angina – P group had lower occurrence 4. Intravenous heparin – P group lower 5. Calcium antagonists – P group lower 6. Topical nitrates – P group lower 7. College education – more patients had college education in P group	R = 392 P = 450	Mean (SD) R = 61.6 (10.0) P = 61.7 (10.2)	Male R = 73.7% P = 78.9%	College R = 19.1% P = 26.0%	Not reported
12	M King, 2000[26] Ward 2000[27]	**16 comparisons made** 5 significant differences found: 1. Marital status – P group less likely to be married 2. EuroQol score – P group had lower score (worse QoL) 3. SAS score – P group had higher score 4. CIS-R score – P group had higher score 5. No. of patient problems – P group reported more problems	R = 327 P = 137	Mean (SD) R = 34 (11.3) P = 40.3 (10.8)	Male R = 26.5% P = 22.6%	Degree or higher R = 23.4% P = 27.8%	Class I–III R = 61.1% P = 67.8%
13	McKay, 1995[130]	**20 comparisons made** 3 significant differences found: 1. White – P group higher % white 2. On welfare – less in the P group 3. No. of days of cocaine use – less for the P group	R = 48 P =96	Mean (SD) R = 41.8 (8.1) P = 40.7 (7.8)	Not applicable	Years of education: mean (SD) R = 12.1 (1.6) P = 12.3 (2.3)	Not reported

continued

No.	Study	Number of baseline comparisons made and significant differences found (R_{total} vs P_{total} only)	Numbers in each arm (R vs P)	Age (years)	Gender	Education	Social class
14	McKay, 1998[129]	**19 comparisons made** 4 significant findings: 1. Education – higher in P group 2. Drug composite score – higher in R group 3. Days of cocaine use – higher in R group 4. Psychiatric score – higher in R group	R = 115 P = 56	Not reported	Not applicable	Not reported	Not reported
15	Melchart, 2002[127]	**8 comparisons made** No significant differences reported	R = 28 P = 80	Mean (SD) R = 71 (24.0) P = 64.5 (22.6)	Male R = 50% P = 59.7%	Not reported	Not reported
16	Nicolaides, 1994[104]	**15 comparisons made** No significant differences reported	R = 488 P = 813	Median R = 38 P = 38	Not applicable	Not reported	Employed R = 68.2% P = 64.3%
17	Olschewski, 1992[125]	**25 comparisons made** 6 significant differences found: 1. Smoking – fewer smokers in P group 2. Non-exertional angina – more common in P group 3. Left main coronary disease – more common in P group 4. Use of beta blockers – more common in P group 5. MI on ECG – more common in R group 6. Ischaemia on ECG – more common in R group	R = 780 P = 1319	Mean (SD) R = 51.2 (7.4) P = 50.9 (8.0)	Male R = 90.3% P = 90.6%	Not reported	Employed R = 72.2% P = 72.9%
18	Paradise, 1984[99]	**11 comparisons made** 2 significant differences found: 1. Unemployment – lower in P group 2. Siblings – less likely to have siblings in P group	R = 91 P = 96	No. in each age group, by condition R: 3–4, 12 (13.1%); 5–6, 21 (23.0%); 7–15, 58 (63.7%) P: 3–4, 11 (11.4%); 5–6, 32 (33.3%); 7–15, 53 (55.2%)	Male R = 44% P = 50%	Not reported	Parents executive or professional R = 17.6% P = 25%
19	Paradise, 1990[99]	**No comparisons made**	R = 99 P = 114	Not reported	Not reported	Not reported	Not reported
20	Reddihough, 1998[100]	**No comparisons made**	R = 34 P = 32	Not reported	Not reported	Not reported	Not reported

continued

No.	Study	Number of baseline comparisons made and significant differences found (R_{total} vs P_{total} only)	Numbers in each arm (R vs P)	Age (years)	Gender	Education	Social class
21	Riedl, 2001[114]	**No comparison made**	R = 28 P = 42	Not reported	Not reported	Not reported	Not reported
22	Rovers, 2000, 2001[97,101]	**7 comparisons made** 3 significant differences found: 1. Older siblings – more children in R group had older siblings 2. Attended day care – less children in R group attended day care 3. Low educational level of mother – more mothers with low educational level in R group	R = 187 P = 133	Age of infant No significant differences	Male infant R = 58.9% P = 56.8%	For low educational level of mother, % in each cohort: R = 26.4% P = 14.8%	Not reported
23	Schumacher, 1994[113]	**7 comparisons made** No significant difference reported	R = 473 P = 247	Not reported	Not reported	Not reported	Not reported
24	Schmoor, 1996[116]	**7 comparisons made** No significant difference reported	R = 189 P = 129	Not reported	Not reported	Not reported	Not reported
25	Williams, 1999[132]	**30 comparisons made** 2 significant differences: 1. Gender – more men in R group 2. Employment – higher in P group	R = 121 P = 128	Mean (SD) R = 49.8 (11.3) P = 49.3 (18.9)	Male R = 47.8% P = 34.4%	Left school before age 16 years: R = 43.7% P = 45.3%	Employed R = 12% P = 28%
26	Cooper, 1997[105]	**Comparisons not applicable** Rücker design					
27	Janevic, 2003[108]	**Comparison not applicable** Rücker design					
28	Noel, 1998[131]	**Comparisons not applicable** Wennberg design					
29	Rokke, 1999[111]	**Comparisons not applicable** Wennberg-type design					
30	Hardy, 1995[96] Shapiro, 1994[112]	**Comparisons not applicable** Standard RCT					
31	Moffett, 1999[120]	**Comparisons not applicable** Standard RCT					

continued

No.	Study	Number of baseline comparisons made and significant differences found (R_{total} vs P_{total} only)	Numbers in each arm (R vs P)	Age (years)	Gender	Education	Social class
32	Devine, 1973[95]	**Comparisons not applicable**					
33	Renjilian, 2001[94]	**Comparisons not applicable**					
34	Van Dyck, 1997[93]	**Comparisons not applicable**					

ANOVA, analysis of variance; NP, non-preference group.

Appendix 4

Summary baseline demographic comparisons between randomised and preference cohorts

Sample size	*N*	Trial	No. of demographic	No. of significant	Direction
Small	40	Rokke, 1999[111]	N/A	N/A	N/A. Wennberg-type design
(*N* = 40–150)	48	Devine, 1973[95]	N/A	N/A	N/A. Unusual design
	60	Gossop, 1986[128]	0	–	–
	64	Van Dyck, 1997[93]	N/A	N/A	N/A. Unusual design
	66	Reddihough, 1998[100]	0	–	–
	66	Bakker, 2000[117]	3	0	–
	70	Riedl, 2001[114]	0	–	
	75	Renjilian, 2001[94]	N/A	N/A	N/A. Unusual design
	98	Bain, 2001[106]	1	0	–
	108	Melchart, 2002[127a]	2	0	–
	120	Hardy, 1994, 1995[96,112]	N/A	N/A	N/A. No preference cohort
	144	McKay, 1995[130]	8	2	1. On welfare P < R 2. White P > R
Medium	151	Helsing, 1998[113]	2	0	–
(*N* = 151–300)	171	McKay, 1998[129]	6	1	1. Education P > R
	187	Paradise, 1984[98a]	5	2	1. Unemployment P < R 2. Siblings P < R
	187	Moffet, 1999[120]	N/A	N/A	N/A. No preference cohort
	213	Paradise, 1990[99]	0	–	–
	227	Cooper, 1997[105]	N/A	N/A	N/A. Wennberg-type design
	249	Williams, 1999[132a]	8	2	1. Female: P > R 2. Employment: P > R
Large	320	Rovers, 2000, 2001[97,101]	5	3	1. Older siblings P < R 2. Attended day care P > R 3. Lower education level of mother) (i.e. P more educated) P < R
(*N* = 301+)	323	Bedi, 2000[109,110a]	5	0	–
	328	Schmoor, 1996[116]	0	–	–
	363	Henshaw, 1993, 1994[102,103,126a]	6	0	–
	445	Ashok, 2002[107]	1	0	–
	464	M King, 2000[26,27]	8	1	1. Married P < R
	476	Kendrick, 2001[119]	7	1	1. Male P > R
	580	Kerry, 2000[118]	4	0	–
	596	Noel, 1998[131]	N/A	N/A	N/A. Wennberg-type design
	692	Detre, 2000[121]	5	2	1. White P > R 2. Education P > R
	720	Schumacher, 1994[115]	0	–	–
	842	SB King, 1995[122–124b]	4	1	1. College education P > R
	1071	Janevic, 2003[108]	N/A	N/A	N/A. Wennberg-type design
	1301	Nikolaides, 1994[104]	2	0	–
	2099	Olschewski, 1992[125]	4	0	–

[a] Trials with R1 vs R2 or P1 vs P2: type comparisons or findings not explicitly reported.
Paradise, 1984:
1. Siblings P1 distribution differed from P2.
2. Parents' socio-economic status R1 distribution differed from R2.
Williams, 1999:
Unskilled worker P2 > P1.
Henshaw, 1993, 1994:
P2 lived further from treatment site (hospital stay required).

[b] This trial had entire R vs entire P comparison in addition to R1 vs P1 and R2 vs P2. In such cases only the entire cohorts' comparison is reported in the table.

Appendix 5

Summary of baseline health-related measure comparisons between the randomised and preference cohorts

Sample size	*N*	Trial	No. of health-related comparison	No. significant	Direction
Small	40	Rokke, 1999[111]	N/A	N/A	N/A. Wennberg-type design
(*N* = 40–150)	48	Devine, 1973[95]	N/A	N/A	N/A. Unusual design
	60	Gossop, 1986[128]	0	–	–
	64	Van Dyck, 1997[93]	N/A	N/A	N/A. Unusual design
	66	Reddihough, 1998[100]	0	–	–
	66	Bakker, 2000[117]	16	0	–
	70	Riedl, 2001[114]	0	–	–
	75	Renjilian, 2001[94]	N/A	N/A	N/A. Unusual design
	98	Bain, 2001[106]	9	0	–
	108	Melchart, 2002[127a]	6	0	–
	120	Hardy, 1994, 1995[96,112]	N/A	N/A	N/A. No preference cohort
	144	McKay, 1995[130]	12	1	No of days of cocaine use (mean) P < R
Medium	151	Helsing, 1998[113]	5	0	–
(*n* = 151–300)	171	McKay, 1998[129]	13	3	1. Drug composite score P < R 2. Days of cocaine use (mean) P < R 3. Psychiatric score P < R
	187	Paradise, 1984[98a]	6	0	–
	187	Moffett, 1999[120]	N/A	N/A	N/A. No preference cohort
	213	Paradise, 1990[99]	0	–	–
	227	Cooper, 1997[105]	N/A	N/A	N/A. Wennberg-type design
	249	Williams, 1999[132a]	22	0	–
Large	320	Rovers, 2000, 2001[97,101]	2	0	–
(*N* = 301+)	323	Bedi, 2000[109,110a]	8	0	–
	328	Schmoor, 1996[116]	7	0	–
	363	Henshaw, 1993, 1994[102,103,126a]	4	0	–
	445	Ashok, 2002[107]	5	0	–
	464	M King, 2000[26,27]	8	4	1. EuroQuoL P < R 2. SAS score P > R 3. CIS-R score P > R 4. No. of patient problems P > R
	476	Kendrick, 2001[119]	10	1	1. Roland disability score P > R
	580	Kerry, 2000[118]	27	2	1. Bodily pain score P > R 2. General health score P > R (R poorer health)
	596	Noel, 1998[131]	N/A	N/A	N/A. Wennberg-type design
	692	Detre, 2000[121]	31	5	1. Smoking P < R 2. Activity level P > R 3. Congestive heart failure P < R 4. Fair/poor QoL P < R 5. Ejection fraction P > R

continued

Sample size	*N*	Trial	No. of health-related comparison	No. significant	Direction
	720	Schumacher, 1994[115]	7	0	–
	842	SB King, 1995[122–124b]	25	6	1. No of narrowings ≥ 70 in diameter P > R 2. Totally occluded vessel P > R 3. CCS class III or IV angina P < R 4. Intravenous heparin P < R 5. Calcium antagonists P < R 6. Topical nitrates P < R
	1071	Janevic, 2003[108]	N/A	N/A	N/A. Wennberg-type design
	1301	Nikolaides, 1994[104]	13	0	–
	2099	Olschewski, 1992[125]	21	6	1. Non-exertional angina P > R 2. Cigarette use P < R 3. Medication (β-blockers) P > R 4. ECG Q-wave P < R 5. ECG ST depression P < R 6. Left main >50% blocked coronary artery P > R

[a] Trials with R1 vs R2 or P1 vs P2: type comparisons or findings not explicitly reported.
Paradise, 1984
History of throat infection R1 distribution differed from R2.
Bedi, 2000
1. RDC score
2. GP notes for depression.
In both P2 more likely to have severe depression.
Melchart, 2002
Previous gastroscopy R2 > R1 and P.

[b] This trial had entire R vs entire P comparison in addition to R1 vs P1 and R2 vs P2. In such cases only the entire cohorts' comparison is reported in the table.

Appendix 6

Trial summary data and description of analysis methods and results

No.	Study	Primary outcome	Randomised (R)	Preference (P) (non-randomised)	Analysis presented and results (primary outcome only)
1	Ashok, 2002[107]	Total number of days bleeding	R1 = medical termination, *N* = 188 R2 = surgical termination, *N* = 180 Baseline data not applicable 2–3 weeks: R1: *N* = 118, M = 14.2, SD = 4.8 R2: *N* = 111, M = 11.2, SD = 5.9	P1 = medical termination, *N* = 15 P2 = surgical termination, *N* = 62 Baseline data not applicable 2–3 weeks: P1: *N* = 9, M = 13.0, SD = 4.1 P2: *N* = 36, M = 10.8, SD = 4.7	*Analysis* R1 vs R2 P1 vs P2 Unadjusted, using unpaired *t*-test *Results* R1 vs R2: Difference in means (95% CI): 3.0 (1.6 to 4.4), $p < 0.0001$ P1 vs P2: Difference in means (95% CI): 2.2 (–1.5 to 5.6) NS Effect of preference not ascertained

continued

No.	Study	Primary outcome	Randomised (R)	Preference (P) (non-randomised)	Analysis presented and results (primary outcome only)
2	Bain, 2001[106]	Acceptability of treatment Measured by 12-point semantic differential scale: –3 = best +3 = worst	R1 = local anaesthetic, *N* = 20 R2 = general anaesthetic, *N* = 16 Baseline data not applicable Mean (SD) R1: [*N* = 20] Pain–painful: 1.5 (1.24) Happy–sad: 0.2 (1.06) Good–bad: 0.25 (1.02) Pleasant–unpleasant: 0.85 (1.14) Positive–negative: –0.5 (1.05) Safe–dangerous: –1.0 (1.26) Attractive–unattractive: 0.55 (1.15) Mild–harsh: 0.15 (1.18) Agreeable–disagreeable: 0.00 (1.25) Active–passive: 0.25 (0.72) Easy–hard: 0.10 (1.17) Fast–slow: –1.20 (1.32) R2: [*N* = 16] Pain–painful: 0.94 (1.53) Happy–sad: –0.13 (0.89) Good–bad: –0.5 (1.55) Pleasant–unpleasant: 0.19 (0.98) Positive–negative: –1.06 (1.29) Safe–dangerous: –1.0 (1.37) Attractive–unattractive: 0.00 (1.06) Mild–harsh: 0.00 (1.15) Agreeable–disagreeable: –0.56 (1.26) Active–passive: –0.25 (1.29) Easy–hard: –0.38 (1.15) Fast–slow: –0.69 (1.78)	P1 = local anaesthetic, *N* = 32 P2 = general anaesthetic, *N* = 30 Baseline data not applicable Mean (SD) P1: [*N* = 32] Pain–painful: 1.06 (1.11) Happy–sad: –0.25 (1.02) Good–bad: –0.34 (1.18) Pleasant–unpleasant: 0.78 (1.01) Positive–negative: –0.47 (1.32) Safe–dangerous: –1.0 (1.50) Attractive–unattractive: 0.13 (0.87) Mild–harsh: –0.10 (0.93) Agreeable–disagreeable: –0.13 (1.18) Active–passive: –0.19 (0.74) Easy–hard: –0.44 (1.11) Fast–slow: –1.16 (1.59) P2: [*N* = 30] Pain–painful: 1.00 (1.58) Happy–sad: 0.00 (1.24) Good–bad: 0.00 (0.96) Pleasant–unpleasant: 0.53 (1.20) Positive–negative: –0.50 (1.55) Safe–dangerous: –0.77 (1.43) Attractive–unattractive: 0.00 (1.05) Mild–harsh: –0.17 (0.87) Agreeable–disagreeable: –0.40 (1.13) Active–passive: 0.00 (0.78) Easy–hard: –0.27 (0.98) Fast–slow: –0.63	*Analysis* R1 vs R2 P1 vs P2 Unadjusted using *t*-test *Results* R1 vs R2: 95% CI for difference in means Pain–painful: (–0.37 to 1.50) Happy–sad: (–0.35 to 1.00) Good–bad: (–0.12 to 1.60) Pleasant–unpleasant: (–0.06 to 1.39) Positive–negative: (–0.23 to 1.35) Safe–dangerous: (–0.89 to 0.89) Attractive–unattractive: (0.14 to 1.37) Mild–harsh: (–0.65 to 0.95) Agreeable–disagreeable: (–0.19 to 1.52) Active–passive: (–0.19 to 1.19) Easy–hard: (0.31 to 1.26) Fast–slow: (–1.56 to 0.54) P1 vs P2: 95% CI for difference Pain–painful: (–0.63 to 0.75) Happy–sad: (–0.72 to 0.42) Good–bad: (–0.99 to 0.11) Pleasant–unpleasant: (–0.31 to 0.81) Positive–negative: (–0.70 to 0.76) Safe–dangerous: (–0.98 to 0.51) Attractive–unattractive: (–0.30 to 0.68) Mild–harsh: (–0.48 to 0.44) Agreeable–disagreeable: (–0.31 to 0.86) Active–passive: (–0.51 to 0.27) Easy–hard: (–0.70 to 0.36) Fast–slow: (–1.29 to 0.24) Effect of preference not ascertained

continued

No.	Study	Primary outcome	Randomised (R)	Preference (P) (non-randomised)	Analysis presented and results (primary outcome only)
3	Bakker, 2000[117]	Panic frequency	R1 = CT, *N* = 35 R2 = paroxetine, *N* = 32 R3 = clomipramine, *N* = 32 R4 = placebo, *N* = 32 Baseline R1: M = 7.0, SD = 9.0 R2: M = 5.8, SD = 8.0 R3: M = 6.0, SD = 4.9 R4: M = 4.4, SD = 6.1 Week 12 R1: M = 2.8, SD = 8.4 R2: M = 1.3, SD = 3.6 R3: M = 2.0, SD = 4.2 R4: M = 4.4, SD = 4.7	P1 = cognitive therapy (CT), *N* = 31 Baseline P1: M = 5.2, SD = 5.7 Week 12 P1: M = 0.9, SD = 1.4	*Analysis* R1 vs R2 vs R3 vs R4 Using general linear model (logged data) with the following factors: treatment, time (baseline and post-treatment), treatment by time interaction. Effect of including baseline agoraphobia score as covariate also assessed. Analysis using last observation carried forward for dropouts performed R1 vs P1 Using general linear model with the following factors: type of assignment (random/preference), time (baseline and post-treatment), time by assignment interaction. Analysis using last observation carried forward for dropouts performed *Results* R1 vs R2 vs R3 vs R4 (using last observation carried forward) Significant effect of time (panic decreased over time), treatment, and time by treatment interaction (panic decreased more with paroxetine than placebo, $p = 0.011$; clomipramine vs placebo, $p = 0.057$; CT versus placebo, $p = 0.11$) R1 vs P1 (using last observation carried forward) Significant effect of time (panic decreased over time), no significant effect of assignment type ($p = 0.45$), or time by assignment interaction

continued

No.	Study	Primary outcome	Randomised (R)	Preference (P) (non-randomised)	Analysis presented and results (primary outcome only)
4	Bedi, 2000[109] Chilvers, 2001[110]	BDI	R1 = counselling, *N* = 52 R2 = antidepressants, *N* = 51 Baseline R1: *N* = 50, M = 27.1, SD = 8.0 R2: *N* = 49, M = 27.0, SD = 8.0 8 weeks R1: *N*= 39, M = 15.2, SD = 11.6 R2: *N* = 44, M = 14.8, SD = 10.1 1 year R1: *N* = 31, M = 16.7, SD = 11.5 R2: *N* = 34, M = 14.6, SD = 13.1	P1 = counselling, *N* = 140 P2 = antidepressants, *N* = 80 Baseline P1: *N* = 140, M = 25.7, SD = 7.7 P2: *N* = 80, M = 25.4, SD = 9.4 8 weeks P1: *N* = 108, M = 14.4, SD = 9.8 P2: *N* = 56, M = 14.0, SD = 9.3 1 year P1: *N* = 96 P2: *N* = 46 1-year scores not reported	*Analysis* R1 vs R2 R vs P R1 vs P1 R2 vs P2 (R1 + P1) vs (R2 + P2) Using unpaired *t*-test and some comparison also adjusted for baseline values. Sensitivity analysis to investigate bias due to missing data. Treatment by assignment interaction tested *Results* 8 weeks R1 vs R2, adjusted Difference in means (95% CI): 0.1 (–4.2 to 4.3), $p = 0.97$ R vs P, unadjusted Difference in means (95% CI): 0.7 (–1.0 to 3.4), $p = 0.59$ R2 vs P2, unadjusted Difference in means (95% CI): 0.9 (–3.0 to 4.7), $p = 0.66$ R1 vs P1, unadjusted Difference in means (95% CI): 0.8 (–3.0 to 4.6), $p = 0.69$ (R1 + P1) vs (R2 + P2), unadjusted Difference in means (95% CI): 0.3 (–2.3 to 2.8), $p = 0.84$ 1 year R1 vs R2, unadjusted Difference in means (95% CI): 2.1 (–4.0 to 8.2), $p = 0.49$ R2 vs P2, unadjusted Difference in means (95% CI): 3.1 (–1.8 to 7.8) R1 vs P1, unadjusted Difference in means (95% CI): 4.6 (0.0 to 9.2) (patients choosing counselling did better than those randomised to it) (R1 + P1) vs (R2 + P2) Difference in means (95% CI): 0.4 (–2.7 to 3.5), $p = 0.81$; adjusted $p = 0.34$ No significant treatment by assignment interaction ($p = 0.6$)

continued

No.	Study	Primary outcome	Randomised (R)	Preference (P) (non-randomised)	Analysis presented and results (primary outcome only)
5	Detre, 1999[121]	1. All-cause mortality 2. Cardiac mortality	R1 = PTCA, *N* = 170 R2 = CABG, *N* = 173 Baseline data not applicable 5-year all-cause mortality (%) R1: 34.5 R2: 19.4 5-year cardiac mortality (%) R1: 23.4 R2: 8.2	P1 = PTCA, *N* = 182 P2 = CABG, *N* = 117 Baseline data not applicable 5-year all-cause mortality (%) P1: 14.4 P2: 14.9 5-year cardiac mortality (%) P1: 7.5 P2: 6.0	*Analysis* R1 vs R2 unadjusted P1 vs P2 unadjusted and adjusted (for clinical, angiographic and QoL baseline factors) Using log-rank test and Cox regression *Results* All-cause mortality R1 vs R2 Unadjusted RR (95% CI): 1.87 (1.24 to 2.82), $p = 0.0024$ P1 vs P2 Unadjusted RR (95% CI): 1.10 (0.64 to 1.87), $p = 0.73$ Adjusted RR (95% CI): 1.29 (0.73 to 2.28) Cardiac mortality R1 vs R2 Unadjusted RR (95% CI): 3.10 (1.64 to 5.85), $p = 0.002$ P1 vs P2 Unadjusted RR (95% CI): 1.07 (0.49 to 2.37), $p = 0.86$ Adjusted RR (95% CI): 1.41 (0.60 to 3.29) Effect of preference not ascertained
6	Gossop, 1986[128]	Abstinence (withdrawal) from opiates	R1 = outpatient R2 = inpatient Total *N* = 20 Baseline data not applicable Complete withdrawal: % (*n*) R: 35% (7)	P1 = outpatient P2 = inpatient Total *N* = 40 Baseline data not applicable Complete withdrawal: % (*n*) P: 58% (23)	*Analysis* (R1 + P1) vs (R2 + P2) R vs P R1 vs P1 R2 vs P2 Unadjusted, using χ^2 test *Results* (R1 + P1) vs (R2 + P2), unadjusted, $p < 0.001$ (opiate withdrawal more likely in inpatient group compared to outpatient group, 81% versus 17%) R vs P, unadjusted, $p = 0.10$ R1 vs P1 and R2 vs P2, not significant

continued

No.	Study	Primary outcome	Randomised (R)	Preference (P) (non-randomised)	Analysis presented and results (primary outcome only)
7	Helsing, 1998[113]	Global QoL	R1 = CT, *N* = 22 R2 = BSC, *N* = 26 Baseline R1: M = 51.6, SD = 27.9, *N* = 20 R2: M = 53.0, SD = 20.5, *N* = 26 Score change from baseline: 4 weeks R1: M change = 2.4, *N* = 18 R2: M change = –8, *N* = 21 8 weeks R1: M change = 7.3, *N* = 15 R2: M change = –6.9, *N* = 15 12 weeks: R1: M change = 3.9, *N* = 15 R2: M change = –0.7, *N* = 10 16 weeks R1: M change = 4.5, *N* = 13 R2: M change = –14.8, *N* = 9 20 weeks: R1: M change = 1.9, *N* = 13 R2: M change = –25.9, *N* = 9 24 weeks R1: M change = -2.4, *N* = 10 R2: M change = –8.3, *N* = 6 SDs not given	P1 = CT, *N* = 97 P2 = BSC, *N* = 5 No data for non-randomised cohort	*Analysis* R1 vs R2 Using Mann–Whitney *U*-test on change from baseline *Results* R1 vs R2, change from baseline 4 weeks: *p* = 0.099 8 weeks: *p* = 0.11 12 weeks: *p* = 0.84 16 weeks: *p* = 0.075 20 weeks: *p* = 0.020 24 weeks: *p* = 0.44 Outcome for preference groups not presented

continued

No.	Study	Primary outcome	Randomised (R)	Preference (P) (non-randomised)	Analysis presented and results (primary outcome only)
8	Henshaw, 1993, 1994,[102,103] Howie 1997[126]	1. Proportions opting for same treatment (after procedure) 2. The 'good–bad' item on the semantic differential scale 3. Anxiety subscale of HADS 4. Fall in self-esteem measure 5. Number opting for same procedure (2 years later) NOSP	R1 = medical abortion, *N* = 99 R2 = surgical abortion, *N* = 96 Baseline data not applicable Choice of future method: % (*n*) R1: *N* = 94 Opt for medical: 74 (70) Opt for surgical: 22 (21) Undecided: 3 (3) R2: *N* = 95 Opt for medical: 2 (2) Opt for surgical: 87 (83) Undecided: 11 (10) 'Good–bad' rating R1: *N* = 94, M = –0.3, SD = 1.6 R2: *N* = 95, M = 0.3, SD = 1.7 Baseline Anxiety subscale R1: *N* = 97, M = 9.6, SD = 4.5 R2: *N* = 93, M = 10.2, SD = 4.5 Post-treatment Anxiety subscale R1: *N* = 93, M = 5.1, SD = 4.3 R2: *N* = 95, M = 4.9, SD = 3.6 % (*n*) with fall in self-esteem R1: 40.9 (38) R2: 35.8 (34) NOSP (at 2 years): % (*n*) R1: 64 (25) (*N* = 39) R2: 87 (32) (*N* = 38)	P1 = medical abortion, *N* = 73 P2 = surgical abortion, *N* = 95 Baseline data not applicable Choice of future method: % (*n*) P1: *N* = 72 Opt for medical: 94 (68) Opt for surgical: 4 (3) Undecided: 1 (1) P2: *N* = 84 Opt for medical: 4 (3) Opt for surgical: 90 (76) Undecided: 6 (5) 'Good–bad' rating P1: *N* = 73, M = –0.1, SD = 1.7 P2: *N* = 85, M = –0.01, SD = 1.7 Baseline Anxiety subscale P1: *N* = 72, M = 10.6, SD = 4.3 P2: *N* = 93, M = 10.2, SD = 4.3 Post-treatment Anxiety subscale P1: *N* = 72, M = 4.9, SD = 4.2 P2: *N* = 84, M = 5.1, SD = 4.1 % (*n*) with fall in self-esteem P1: 41.7 (30) P2: 32.9 (28) NOSP (at 2 years): % (*n*) P1: 89 (24) (*N* = 27) P2: 89 (32) (*N* = 36)	*Analysis for outcomes 1 and 2* R1 vs R2 P1 vs P2 R vs P (for NOSP) Unadjusted using χ^2 test and unpaired *t*-test *Results* *Choice of future method* R1 vs R2: $p < 0.001$ (patients randomised to medical less likely to choose same method compared to those randomised to surgical) P1 vs P2: no significant difference in % choosing same method *Semantic differential scale* R1 vs R2: Difference in means (95% CI): 0.53 (0.06 to 1.00), $p < 0.05$ (worse rating for medical compared with surgical abortion) P1 vs P2: Difference in means (95% CI): 0.11 (–0.43 to 0.66) Effect of preference not ascertained *Analysis for outcomes 3–5* R1 vs R2 vs P1 vs P2 Unadjusted using ANOVA and χ^2 test For NOSP: R vs P, R1 vs R2, P1 vs P2, unadjusted using χ^2 test *Anxiety subscale* R1 vs R2 vs P1 vs P2 No significant differences between the four groups *Fall in self-esteem* R1 vs R2 vs P1 vs P2 No significant differences between the four groups *NOSP* R vs P: difference in proportions (95% CI): –15 (–27.4 to –2.4), $p < 0.05$ (random group less likely than preference group to opt for same procedure) R1 vs R2: difference in proportions (95% CI): –23 (–39 to –1), $p < 0.05$ (random medical less likely than random surgical group to opt for same procedure) P1 vs P2: difference in proportions (95% CI): 0 (–16 to 16)

continued

No.	Study	Primary outcome	Randomised (R)	Preference (P) (non-randomised)	Analysis presented and results (primary outcome only)
9	Kendrick, 2001[119]	1. Roland Disability Score 2. Pain score	R1 = X-ray, *N* = 210 R2 = control (no X-ray), *N* = 211 Baseline Median (IQR) RS R1: 7 (4, 11) (*N* = 210) R2: 8 (4, 12) (*N* = 211) PS R1: 2 (1, 2) (*N* = 210) R2: 2 (1, 2) (*N* = 211) Median (IQR) RS 3 months R1: 4 (1, 8) (*N* = 199) R2: 3 (1, 7) (*N* = 203) 9 months R1: 3 (0, 7) (*N* = 195) R2: 2 (0, 6) (*N* = 199) PS 3 months R1: 1 (1, 2) (*N* = 199) R2: 1 (0, 2) (*N* = 203) 9 months R1: 1 (0, 2) (*N* = 195) R2: 1 (0, 2) (*N* = 199)	P1 = X-ray, *N* = 32 P2 = Control (no X-ray), *N* = 23 Baseline Median (IQR) RS P1: 13 (7, 16) (*N* = 32) P2: 7 (4, 11) (*N* = 23) PS P1: 2 (1, 3) (*N* = 32) P2: 2 (2, 2) (*N* = 23) Median (IQR) RS 3 months P1: 7 (3, 15) (*N* = 30) P2: 3 (2, 7) (*N* = 22) 9 months P1: 3 (1, 7) (*N* = 29) P2: 1 (0, 4) (*N* = 21) PS 3 months P1: 1 (0, 2) (*N* = 30) P2: 3 (2, 7) (*N* = 22) 9 months P1: 1 (0, 2) (*N* = 29) P2: 0 (0, 1) (*N* = 21)	*Analysis* R1 vs R2 R1 vs P1 R2 vs P2 Unadjusted using Mann–Whitney *U*-test *Results* R1 vs R2 *RS* 3 months, $p = 0.05$ (RS higher for X-ray compared with control group) 9 months, $p = 0.06$ *PS* 3 months, $p = 0.06$ 9 months, $p = 0.17$ R1 vs P1 *RS* 3 months, $p = 0.02$ (RS higher for preference X-ray compared with randomised X-ray group) 9 months, $p = 0.77$ *PS* 3 months, $p = 0.73$ 9 months, $p = 0.28$ R2 vs P2 *RS* 3 months, $p = 0.50$ 9 months, $p = 0.21$ *PS* 3 months, $p = 0.26$ 9 months, $p = 0.16$

continued

No.	Study	Primary outcome	Randomised (R)	Preference (P) (non-randomised)	Analysis presented and results (primary outcome only)
10	Kerry, 2000[118]	1. Roland Disability Score 2. Consultation for back pain (a) within 6 weeks (b) from 6 weeks to 1 year	R1 = referred for X-ray, *N* = 73 R2 = not referred for X-ray, *N* = 80 Baseline RS R1: *N* = 65, M = 10.2, SD = 5.5 R2: *N* = 76, M = 10.9, SD = 5.3 Consultation in past year: % (*n*) R1: 28 (19) (*N* = 69) R2: 24 (17) (*N* = 71) RS 6 weeks post-treatment R1: *N* = 59, M = 5.9, SE = 0.7, SD = 5.4 R2: *N* = 67, M = 6.9, SE = 0.8, SD = 6.5 1 year post-treatment R1: *N* = 46, M = 4.5, SE = 0.8, SD = 5.4 R2: *N* = 57, M = 4.3, SE = 0.7, SD = 5.3 Consultation in past year Within 6 weeks post-treatment: % (*n*) R1: 33 (23) (*N* = 69) R2: 37 (26) (*N* = 71) 6 weeks – 1 year post-treatment: % (*n*) R1: 32 (22) (*N* = 69) R2: 39 (28) (*N* = 71)	P1 = referred for X-ray, *N* = 95 P2 = not referred for X-ray, *N* = 332 Baseline RS P1: *N* = 95, M = 10.9, SD = 5.5 P2: *N* = 332, M = 10.8, SD = 5.4 Consultation in past year: % (*n*) P1: 35 (32) (*N* = 91) P2: 23 (73) (*N* = 316) RS 6 weeks post-treatment P1: *N* = 76, M = 6.7, SE = 0.6, SD = 5.2 P2: *N* = 276, M = 5.4, SE = 0.3, SD = 5.0 1 year post-treatment P1: *N* = 63, M = 5.6, SE = 0.6, SD = 4.8 P2: *N* = 254, M = 4.2, SE = 0.3, SD = 4.8 Consultation in past year Within 6 weeks post-treatment: % (*n*) P1: 42 (38) (*N* = 91) P2: 29 (92) (*N* = 313) 6 weeks – 1 year post-treatment: % (*n*) P1: 44 (40) (*N* = 91) P2: 28 (89) (*N* = 313)	*Analysis* R1 vs R2; P1 vs P2 Unadjusted using *t*-tests and χ^2 tests and adjusted for age, sex and length of time with back pain at recruitment using regression and logistic regression R vs P for RS, adjusted analysis *Results* R1 vs R2 *RS* 6 weeks Unadjusted difference in means (95% CI): 1.0 (–1.1 to 3.0) Adjusted difference in means (95% CI): 0.7 (–1.4 to 2.7) 1 year Unadjusted difference in means (95% CI): –0.2 (–2.2 to 1.8) Adjusted difference in means (95% CI): 0.3 (–1.6 to 2.2) *Consultation* Within 6 weeks Unadjusted OR (95% CI): 0.9 (0.4 to 1.7) Adjusted OR (95% CI): 0.8 (0.4 to 1.8) 6 weeks – 1 year Unadjusted OR (95% CI): 0.7 (0.4 to 1.4) Adjusted OR (95% CI): 0.7 (0.3 to 1.4) P1 vs P2 *RS* 6 weeks Unadjusted difference in means (95% CI): –1.3 (–2.7 to 0) Adjusted difference in means (95% CI): –0.2 (–1.7 to 1.2) 1 year Unadjusted difference in means (95% CI): –1.4 (2.8 to 0.0), $p < 0.05$ (greater disability in the X-ray compared with non-X-ray group) Adjusted difference in means (95% CI): –0.3 (–1.8 to 1.0)

continued

No.	Study	Primary outcome	Randomised (R)	Preference (P) (non-randomised)	Analysis presented and results (primary outcome only)
					Consultation Within 6 weeks Unadjusted OR (95% CI): 1.7 (1.1 to 2.8) Adjusted OR (95% CI): 2.1 (1.2 to 3.5) 6 weeks – 1 year Unadjusted OR (95% CI): 2.0 (1.2 to 3.2) Adjusted OR (95% CI): 1.6 (1.0 to 2.7) (X-ray group consulted more than non-X-ray group) R vs P *RS* 6 weeks Adjusted difference in means (95% CI): 0.1 (–1.0 to 1.3) 1 year Adjusted difference in means (95% CI): –0.1 (–1.3 to 0.9) Effect of preference not ascertained for consultations
11	SB King, 1994, 1995, 1997[122–124]	1. All-cause mortality 2. Angina at 3 years	R1 = PTCA, *N* = 198 R2 = CABG, *N* = 194 Baseline data not applicable All-cause mortality (%) 1 year R1: 3.5 R2: 2.1 2 years R1: 5.1 R2: 5.2 3 years R1: 7.1 R2: 6.2 Angina at 3 years (%) R1: 19.6 R2: 11.7	P1 = PTCA, *N* = 168 P2 = CABG, *N* = 270 Baseline data not applicable All-cause mortality (%) 1 year P1: 0.6 P2: 1.5 2 years P1: 3.0 P2: 1.9 3 years P1: 4.9 P2: 3.0 Angina at 3 years (%) P1: 12.4 P2: 10.5	*Analysis* P1 vs P2 R vs P R1 vs P1 R2 vs P2 Unadjusted and adjusted, using log-rank test and Cox regression (adjusted for clinical and angiographic variables and stratified by 'propensity score') For angina outcome: P1 vs P2 and R1 vs P1 only, unadjusted (R1 vs R2 presented in previous publication) *Results* 3-year all-cause mortality P1 vs P2: $p = 0.32$ unadjusted R vs P: 6.6% vs 3.6%, $p = 0.044$ unadjusted R2 vs P2: $p = 0.095$ unadjusted; $p = 0.036$ adjusted R1 vs P1: $p = 0.37$ unadjusted; $p = 0.68$ adjusted Angina at 3 years P1 vs P2: $p = 0.55$ unadjusted R1 vs P1: $p = 0.079$ unadjusted

continued

No.	Study	Primary outcome	Randomised (R)	Preference (P) (non-randomised)	Analysis presented and results (primary outcome only)
12	M King, 2000[26] Ward, 2000[27]	BDI	Randomised 3-way: R1 = CBT, *N* = 63 R2 = NDC, *N* = 67 R3 = GP, *N* = 67 Baseline R1: *N* = 63, M = 27.6, SD = 8.4 R2: *N* = 67, M = 25.4, SD = 8.6 R3: *N* = 67, M = 26.5, SD = 8.9 4 months R1: *N* = 56, M = 12.7, SD = 9.5 R2: *N* = 62, M = 11.5, SD = 7.7 R3: *N* = 62, M = 17.2, SD = 11.9 12 months R1: *N* = 50, M = 9.3, SD = 8.8 R2: *N* = 58, M = 11.1, SD = 9.3 R3: *N* = 57, M = 10.2, SD = 8.5 Randomised 2-way: R1 = CBT, *N* = 134 R2 = NDC, *N* = 126 Baseline: R1: *N* = 134, M = 27.6, SD = 7.9 R2: *N* = 126, M = 27.6, SD = 9.0 4 months R1: *N* = 117, M = 12.5, SD = 10.0 R2: *N* = 112, M =12.3, SD = 8.5 12 months R1: *N* = 107, M = 9.9, SD = 10.2 R2: *N* = 102, M = 11.2, SD = 9.1	P1 = CBT, *N* = 81 P2 = NDC, *N* = 54 (GP arm not included in analyses) Baseline P1: *N* = 81, M = 26.9, SD = 9.2 P2: *N* = 54, M = 27.4, SD = 7.4 4 months P1: *N* = 68, M = 13.0, SD = 10.2 P2: *N* = 52, M = 14.0, SD = 9.1 12 months P1: *N* = 66, M = 10.7, SD = 8.1 P2: *N* = 40, M = 12.3, SD = 9.6	*Analysis* R1 vs R2 vs R3 R1 vs R2 P1 vs P2 Using general linear model with following factors: time (including baseline), treatment arm, clinical site and interactions. Last observation carried forward for missing data. Randomised and preference groups analysed separately *Results* R1 vs R2 vs R3, adjusted $p = 0.25$ Significant time by treatment interaction (depression improved more with psychological therapies in initial 4-month period but GP group 'caught up' by 12 months) R1 vs R2, adjusted No significant differences P1 vs P2, adjusted No significant differences Effect of preference not ascertained

continued

No.	Study	Primary outcome	Randomised (R)	Preference (P) (non-randomised)	Analysis presented and results (primary outcome only)
13	McKay, 1995[130]	Number of drinking days	R1 = day hospital, *N* = 24 R2 = inpatient, *N* = 24 Baseline R1: M =16.8, SD = 7.3 R2: M = 13.0, SD = 7.6 3 months R1: M = 1.4, SD = 4.6 R2: M = 3.4, SD = 7.2 6 months R1: M = 1.6, SD = 3.1 R2: M = 7.4, SD = 10.2 12 months R1: M = 2.9, SD = 5.2 R2: M = 6.7, SD = 9.1	P1 = day hospital, *N* = 65 P2 = inpatient, *N* = 31 Baseline P1: M = 14.5, SD = 8.2 P2: M = 14.8, SD = 11.8 3 months P1: M = 3.9, SD = 7.6 P2: M = 5.5, SD = 9.9 6 months P1: M = 4.2, SD = 7.2 P2: M = 4.5, SD = 8.3 12 months P1: M = 4.5, SD = 7.7 P2: M = 7.0, SD = 9.0	*Analysis* (R1 + P1) vs (R2 + P2) R vs P Using ANCOVA including the following factors: treatment, type of assignment (random/preference), time, baseline covariates and treatment by assignment interaction. *Results* (R1 + P1) vs (R2 + P2), adjusted No significant differences or interactions R vs P, adjusted No significant differences or interactions
14	McKay, 1998[129]	Days of cocaine use	R1 = day hospital, *N* = 58 R2 = inpatient, *N* = 57 Baseline R1: *N* = 58, M = 12.6, SD = 8.4 R2: *N* = 57, M = 13.1, SD = 8.8 3 months R1: *N* = 56, M = 1.7, SD = 4.6 R2: *N* = 51, M = 2.7, SD = 4.9 6 months R1: *N* = 53, M = 1.9, SD = 5.4 R2: *N* = 53, M = 3.4, SD = 6.9 12 months R1: *N* = 52, M = 1.5, SD = 4.2 R2: *N* = 49, M = 2.0, SD = 4.0	P1= day hospital, *N* = 37 P2 = inpatient, *N* = 19 Baseline P1: *N* = 37, M = 7.3, SD = 7.7 P2: *N* = 19, M = 10.1, SD = 8.7 3 months P1: *N* = 34, M = 3.8, SD = 7.9 P2: *N* = 18, M = 1.4, SD = 3.6 6 months P1: *N* = 37, M = 2.4, SD = 4.8 P2: *N* = 17, M = 2.8, SD = 6.8 12 months P1: *N* = 33, M = 1.8, SD = 3.8 P2: *N* = 18, M = 1.3, SD = 3.1	*Analysis* (R1 + P1) vs (R2 + P2) R vs P Using mixed effects regression model with the following factors: treatment, type of assignment (random/preference), time and interactions (treatment by assignment; treatment by time, assignment by time; treatment by assignment by time). Confounding effect of baseline factors investigated separately. Subjects with missing data at >1 time point included in model *Results* (R1 + P1) vs (R2 + P2) adjusted, not significant R vs P adjusted, not significant Significant effect of time ($p < 0.001$) (reduction in cocaine use over time) and assignment by time interaction ($p < 0.01$) (improvement over time greater in randomised than preference group) Other interactions not significant

continued

No.	Study	Primary outcome	Randomised (R)	Preference (P) (non-randomised)	Analysis presented and results (primary outcome only)
15	Melchart, 2002[127]	Patients' perception of the examination (Visual Analogue Scale)	R2 = acupuncture, *N* = 14 R1 = midazolam, *N* = 14 Baseline data not applicable Outcome data for primary measure read from Forest plot No SDs given R1: M = 31 mm R2: M = 19 mm	P1 = acupuncture, *N* = 21 P2 = midazolam, *N* = 51 P3 = no treatment, *N* = 8 Baseline data not applicable Outcome data for primary measure read from Forest plot No SDs given P1: M = 32 mm P2: M = 23 mm P3: M = 28 mm	*Analysis* R1 vs R2 R1 vs P1 R2 vs P2 (R1 + P1) vs (R2 + P2) Unadjusted using Mann–Whitney *U*-tests. Treatment by assignment interaction using ANOVA *Results* (R1 + P1) vs (R2 + P2) unadjusted Difference in means (95% CI): 9 mm (–2 to 20), $p = 0.12$ No significant treatment by assignment interaction Results of other tests not reported
16	Nicolaides, 1994[104]	Foetal loss 1. Total loss 2. Spontaneous loss 3. Termination	R1 = EA, *N* = 238 R2 = CVS, *N* = 250 Baseline data not applicable % (*n*) Total loss R1: 8.4 (20) R2: 4.8 (12) Spontaneous loss R1: 5.9 (14) R2: 1.2 (3) Termination R1: 2.5 (6) R2: 3.6 (9)	P1 = EA, *N* = 493 P2 = CVS, *N* = 320 Baseline data not applicable % (*n*) Total loss P1: 7.1 (35) P2: 6.9 (22) Spontaneous loss P1: 5.1 (25) P2: 3.1 (10) Termination P1: 2.0 (10) P2: 3.8 (12)	*Analysis* R1 vs R2 P1 vs P2 Unadjusted, using χ^2 test *Results* R1 vs R2 Spontaneous loss Difference in proportions (95% CI) 4.7 (1.4 to 8.0), $p < 0.05$ (loss greater for EA than CVS) Total loss and termination, no significant differences P1 vs P2 No significant differences Effect of preference not ascertained

continued

No.	Study	Primary outcome	Randomised (R)	Preference (P) (non-randomised)	Analysis presented and results (primary outcome only)
17	Olschewski, 1992[125]	1. Total mortality	R1 = medical, *N* = 390 R2 = surgery, *N* = 390 Baseline data not applicable for primary outcome 5-year mortality % (*n*) R1 = 9.2% (36) R2 = 7.4% (29)	P1 = medical, *N* = 745 P2 = surgery, *N* = 570 Baseline data not applicable for primary outcome 5-year mortality % (*n*) P1 = 8.6% (64) P2 = 7.0% (40)	*Analysis* R1 vs R2 R1 vs P1 R2 vs P2 Unadjusted using log rank test. Separate analysis using model with interactions also investigated on subset *Results* R1 vs R2 NS R1 vs P1, $p = 0.6$ R2 vs P2 NS, $p = 0.5$
18	Paradise, 1984[98]	No. of episodes of throat infection	R1 = surgery, *N* = 43 R2 = control, *N* = 48 No baseline data for primary outcome Year 1: R1: *N* = 38, M = 1.2, SD = 2.3 R2: *N* = 35, M = 3.1, SD = 2.8 Year 2: R1: *N* = 31, M = 1.6, SD = 2.8 R2: *N* = 29, M = 2.7, SD = 2.0 Year 3: R1: *N* = 22, M = 1.8, SD = 1.8 R2: *N* = 20, M = 2.2, SD = 1.8 SDs are approximated from range	P1 = surgery, *N* = 52 P2 = control, *N* = 44 No baseline data for primary outcome Year 1: P1: *N* = 44, M = 1.8, SD =2.3 P2: *N* = 34, M = 3.1, SD =3.3 Year 2: P1: *N* = 34, M = 1.2, SD = 1.3 P2: *N* = 28, M = 2.5, SD = 2.3 Year 3: P1: *N* = 15, M = 1.5, SD = 1.3 P2: *N* = 13, M = 3.2, SD =1.5 SDs are approximated from the range	*Analysis* R1 vs R2 P1 vs P2 Unadjusted, using χ^2 test on frequency table of episodes Effect of losses to follow-up discussed *Results* R1 vs R2: Year 1: $p = 0.001$ (surgical group had fewer episodes than control group) Year 2: $p = 0.001$ (surgical group had fewer episodes than control group) Year 3: NS P1 vs P2 Year 1: $p = 0.04$ Year 2: $p = 0.001$ Year 3: $p = 0.04$ (surgical group had fewer episodes than control group at each time point) Effect of preference not ascertained

continued

No.	Study	Primary outcome	Randomised (R)	Preference (P) (non-randomised)	Analysis presented and results (primary outcome only)
19	Paradise, 1990[99]	1. Proportion of time with otitis media	R1 = adenoidectomy, *N* = 52 R2 = control, *N* = 47	P1 = adenoidectomy, *N* = 47 P2 = control, *N* = 67	*Analysis* R1 vs R2 P1 vs P2
		2. Number of episodes of otitis media	No baseline data for primary outcome	No baseline data for primary outcome	Unadjusted using χ^2 tests on frequency tables. Effect of losses to follow-up discussed
			Proportion of days with otitis media (%) Year 1: R1: 15.0 (*N* = 48) R2: 28.5 (*N* = 38) Year 2 R1: 17.8 (*N* = 45) R2: 28.4 (*N* = 27) Year 3: R1: 15.1 (*N* = 37) R2: 16.7 (*N* = 15)	Proportion of days with otitis media (%) Year 1: P1: 17.9 (*N* = 39) P2: 23.3 (*N* = 61) Year 2 P1: 16.9 (*N* = 32) P2: 23.5 (*N* = 51) Data not given for year 3	*Results* *Days with otitis media* R1 vs R2 Year 1: $p = 0.04$ (otitis present less frequently for adenoidectomy group compared with control) Year 2: $p = 0.005$ (otitis present less frequently for adenoidectomy group compared with control) Year 3: $p = 0.80$ P1 vs P2 Year 1: $p = 0.55$ Year 2: $p = 0.30$
			No. of episodes Year 1: R1: *N* = 48, M = 1.06 R2: *N* = 38, M = 1.45 Year 2: R1: *N* = 45, M = 1.09 R2: *N* = 27, M = 1.67 Year 3: R1: *N* = 37, M = 0.89 R2: *N* = 15, M = 0.87 SD, year 1 = 2.0 for all groups, year 2 = 1.5 for all groups (approximated)	No. of episodes Year 1: P1: *N* = 39, M = 0.90 P2: *N* = 61, M = 1.39 Year 2: P1: *N* = 32, M = 0.59 P2: *N* = 51, M = 1.35 Data not given for year 3 SD not given	*No. of episodes* R1 vs R2 Year 1: $p = 0.51$ Year 2: $p = 0.01$ (fewer episodes for adenoidectomy group compared with control) Year 3: $p = 0.90$ P1 vs P2 Year 1: $p = 0.30$ Year 2: $p = 0.07$ Effect of preference not ascertained

continued

No.	Study	Primary outcome	Randomised (R)	Preference (P) (non-randomised)	Analysis presented and results (primary outcome only)
20	Reddihough, 1998[100]	1. Cognitive variable of the VAB 2. Total score of the GMFM	R1 = CE, *N* = 17 R2 = control (traditional programme), *N* = 17 Cognitive category of the VAB Baseline R1: *N* = 17, M = 8.1, SD = 2.5 R2: *N* = 17, M = 6.1, SD = 3.6 Post-treatment R1: *N* = 17, M = 7.5, SD = 2.9 R2: *N* = 17, M = 6.5, SD = 3.2 GMFM total score Baseline R1: *N* = 9, M = 26.2, SD = 12.7 R2: *N* = 13, M = 20.4, SD = 11.8 Post-treatment R1: *N* = 9, M = 33.2, SD = 13.8 R2: *N* = 9, M = 28.6, SD = 17.8	P1 = CE, *N* = 15 P2 = control (traditional programme), *N* = 17 Cognitive category of the VAB Baseline P1: *N* = 13, M = 6.1, SD = 3.1 P2: *N* = 13, M = 8.4, SD = 2.1 Post-treatment P1: *N* = 13, M = 6.4, SD = 2.5 P2: *N* = 13, M = 7.5, SD = 2.5 GMFM total score Baseline P1: *N* = 8, M = 27.2, SD = 15.8 P2: *N* = 11, M = 40.0, SD = 14.9 Post-treatment P1: *N* = 8, M = 29.0, SD = 16.7 P2: *N* = 11, M = 52.1, SD= 18.8	*Analysis* R1 vs R2 P1 vs P2 Using ANOVAs including the following factors: treatment, time (pre- and post-treatment), time by treatment interaction *Results* *VAB (cognitive category)* R1 vs R2 adjusted, $p < 0.05$ (CE higher score than control); no significant effect of time or treatment by time interaction P1 vs P2 adjusted, $p < 0.05$ (CE lower score than control); no significant effect of time or treatment by time interaction *GMFM total score* R1 vs R2 adjusted, not significant; significant effect of time ($p < 0.01$) (score improved over time); no significant treatment by time interaction P1 vs P2 adjusted, $p < 0.05$ (score lower for CE than control); significant effect of time ($p < 0.01$) (score improved over time); significant treatment by time interaction ($p < 0.05$) (score improved less for CE than for control group) Effect of preference not ascertained
21	Riedl, 2001[114]	Occurrence of pneumonia	R1 = SDD, *N* = 12 R2 = No SDD/control, *N* = 16 Baseline data not applicable % (*n*) R1: 42 (5) R2: 56 (9)	P1 = SDD, *N* = 13 P2 = no SDD/control, *N* = 29 Baseline data not applicable % (*n*) P1: 31 (4) P2: 31 (9)	*Analysis* R1 vs R2 (R1 + P1) vs (R2 + P2) Unadjusted using χ^2 test *Results* R1 vs R2 unadjusted, no significant difference (R1 + P1) vs (R2 + P2): 36% vs 40% unadjusted, no significant difference Effect of preference not ascertained

continued

No.	Study	Primary outcome	Randomised (R)	Preference (P) (non-randomised)	Analysis presented and results (primary outcome only)
22	Rovers, 2000, 2001[97,101]	1. Occurrence of bilateral OME during the trial period 2. Time spent with effusion over the 1-year period (% of days)	R1 = ventilation tubes, $N = 93$ R2 = watchful waiting, $N = 94$ Baseline data not given Bilateral OME: % 3 months R1: 15 ($N = 93$) R2: 77 ($N = 93$) 6 months R1: 29 ($N = 89$) R2: 66 ($N = 91$) 9 months R1: 27 ($N = 85$) R2: 57 ($N = 82$) 12 months R1: 27 ($N = 90$) R2: 53 ($N = 86$) (data read from graph) Time with effusion (% of days) R1: M = 36 R2: M = 70 SDs not given	P1 = ventilation tubes, $N = 36$ P2 = watchful waiting, $N = 97$ Baseline data not given Bilateral OME: % 3 months P1: 26 ($N = 35$) P2: 72 ($N = 94$) 6 months P1: 18 ($N = 31$) P2: 71 ($N = 88$) 9 months P1: 28 ($N = 28$) P2: 58 ($N = 80$) 12 months P1: 23 ($N = 28$) P2: 57 ($N = 78$) (data read from graph) Time with effusion (% of days) P1: M = 30 P2: M = 71 SDs not given	*Analysis* R1 vs P1 R2 vs P2 Unadjusted using χ^2 test (obtained from author communication) *Results* Bilateral OME R1 vs P1 3 months: $p = 0.13$ 6 months: $p = 0.17$ 9 months: $p = 0.83$ 12 months: $p = 0.66$ R2 vs P2 3 months: $p = 0.34$ 6 months: $p = 0.44$ 9 months: $p = 0.89$ 12 months: $p = 0.76$

continued

No.	Study	Primary outcome	Randomised (R)	Preference (P) (non-randomised)	Analysis presented and results (primary outcome only)
23	Schumacher, 1994[115]	1. Tumour recurrence 2. All-cause mortality	R1 = 3 cycles chemotherapy, *N* = 145 R2 = 6 cycles chemotherapy, *N* = 144 R3 = 3 cycles chemotherapy + tamoxifen, *N* = 93 R4 = 6 cycles chemotherapy + tamoxifen, *N* = 91 Baseline data not applicable Comparison of 3 vs 6 cycles Recurrence at 5 years % (*n*) R1 + R3: 82% (195) R2 + R4: 83% (196) Survival at 5 years R1 + R3: 29% (68) R2 + R4: 31% (73) Comparison of tamoxifen vs no treatment Recurrence at 5 years R1 + R2: 79% (149) R3 + R4: 79% (146) Survival at 5 years R1 + R2: 39% (74) R3 + R4: 34% (62)	P1 = 3 cycles chemotherapy, *N* = 72 P2 = 6 cycles chemotherapy, *N* = 104 P3 = 3 cycles chemotherapy + tamoxifen, *N* = 42 P4 = 6 cycles chemotherapy + tamoxifen, *N* = 29 Baseline data not applicable No summary data given	*Analysis* *Overall mortality* 1. Tamoxifen vs no tamoxifen R3 + R4 vs R1 + R2 P3 + P4 vs P1 + P2 2. 6 vs 3 cycles chemotherapy R2 + R4 vs R1 + R3 P2 + P4 vs P1 + P3 Regression analysis with adjustments made for relevant prognostic factors (menopausal status, no. of involved lymph nodes, tumour size, tumour grade and progesterone receptor status). Treatment by assignment interaction tested *Results* *Effects of Tamoxifen* R3 + R4 vs R1 + R2 Adjusted RR = 0.75 (0.54–1.04), $p = 0.085$ P3 + P4 vs P1 + P2 Adjusted RR = 0.53 (0.30–0.80) Effects of 3 vs 6 cycles chemotherapy R1 + R3 vs R2 + R4 Adjusted RR = 0.90 (0.67–1.19), $p = 0.45$ P2 + P4 vs P1 + P3 Adjusted RR = 0.90 (0.60–1.40), NS Treatment by assignment interaction, $p = 0.99$ for chemotherapy and $p = 0.22$ for tamoxifen Effect of preference not ascertained

continued

No.	Study	Primary outcome	Randomised (R)	Preference (P) (non-randomised)	Analysis presented and results (primary outcome only)
24	Schmoor, 1996[116]	All-cause mortality	R1 = 6 cycles chemotherapy, $N = 101$ R2 = 6 cycles chemotherapy + radiotherapy, $N = 88$ No summary data given	P1 = 6 cycles chemotherapy, $N = 88$ P2 = 6 cycles chemotherapy + radiotherapy, $N = 41$ No summary data given	*Analysis* For chemotherapy vs chemotherapy + radiotherapy (R2 vs R1 and P2 vs P2) Analysis adjusted for prognostic factors (menopausal status, no. of involved lymph nodes, tumour size, tumour grade and progesterone receptor status). Treatment by assignment interaction tested *Results* R2 vs R1 = adjusted RR 0.79 (0.50–1.30), $p = 0.32$ P1 vs P2 = adjusted RR 0.76 (0.40–1.50), NS Treatment by assignment interaction, $p = 0.94$ Effect of preference not ascertained
25	Williams, 1999[132]	Sickness impact profile	R1 = outpatient, $N = 41$ R2 = inpatient, $N = 42$ No summary data given	P1 = local anaesthesia, $N = 59$ P2 = general anaesthesia, $N = 63$ No summary data given	*Analysis* R1 vs R2 P1 vs P2 R1 vs P1 R2 vs P2 Unadjusted using ANOVA or χ^2 test *Results* R1 vs R2 Significant difference (inpatient group better than outpatient group) P1 vs P2 Significant difference (inpatient group better than outpatient group) R1 vs P1 Not significant R2 vs P2 Uncertain result

continued

No.	Study	Primary outcome	Randomised (R)	Preference (P) (non-randomised)	Analysis presented and results (primary outcome only)
26	Cooper, 1997[105]	1. Treatment acceptable 2. Opt for same treatment in future	R1 = TCRE, *N* = 93 R2 = medical, *N* = 94 Baseline data not applicable Treatment acceptable: % (*n*) R1: 92 (86) (*N* = 93) R2: 35 (33) (*N* = 93) Continue same treatment: % (*n*) R1: 92 (86) (*N* = 93) R2: 31 (29) (*N* = 93)	P1 = TCRE, *N* = 21 P2 = medical, *N* = 19 Baseline data not applicable Treatment acceptable: % (*n*) P1: 81 (17) (*N* = 21) P2: 63 (12) (*N* = 19) Continue same treatment: % (*n*) P1: 76 (16) (*N* = 21) P2: 63 (12) (*N* = 19)	*Analysis* R1 vs R2 R1 vs P1 R2 vs P2 Unadjusted using χ^2 test *Results* R2 vs R1 Treatment acceptable 95% CI for difference in proportions (–67 to –45), $p < 0.001$ Continue same treatment 95% CI for difference in proportions (–72 to –51), $p < 0.001$ (Treatment acceptability and opting for same treatment in future higher for TCRE than medical group) R1 vs P1 No significant differences R2 vs P2 Treatment acceptability and opting for same treatment in future higher for preference compared with randomised medical group

continued

No.	Study	Primary outcome	Randomised (R)	Preference (P) (non-randomised)	Analysis presented and results (primary outcome only)
27	Janevic, 2003[108]	1. Attendance/ completion of at least one unit 2. Attendance/ completion level	R1 = self-directed, *N* = 201 R2 = group programme, *N* = 190 R3 = control, *N* = 184 No data provided	P1 = self-directed, *N* = 175 P2 = group programme, *N* = 321 No data provided	*Analysis* R1 vs P1 R2 vs P2 Using logistic regression model including the following factors: type of assignment (random/preference), treatment, assignment by treatment interaction and adjusting for socio-demographic variables and recruitment sites. Confounding effect of baseline health characteristics also assessed *Results* P1 vs R1 *Attendance/completion of at least one unit:* Adjusted OR (95% CI): 1.13 (0.69 to 1.85), $p = 0.630$ *Attendance/completion level:* Adjusted OR (95% CI): 1.23 (0.83 to 1.83), $p = 0.298$ P2 vs R2 *Attendance/completion of at least one unit:* Adjusted OR (95% CI): 4.49 (2.58 to 7.81), $p = 0.0001$ *Attendance/completion level:* Adjusted OR (95% CI): 1.90 (1.29 to 2.78), $p = 0.001$ (attendance more likely in preference compared with randomised group programme)

continued

No.	Study	Primary outcome	Randomised (R)	Preference (P) (non-randomised)	Analysis presented and results (primary outcome only)
28	Noel, 1998[131]	1. Attendance classes 2. Diabetes knowledge	R1 = nutrition curriculum R2 = standard curriculum (Total *N* = 291, separate data not given) Baseline data not applicable Classes attended: % (*n*) 0: 31 (91) 1: 8 (23) 2: 5 (16) 3: 4 (11) 4: 9 (26) 5: 52 (158) Diabetes knowledge: Baseline: M = 11.8 6 months: M = 12.4	P1 = nutrition curriculum P2 = standard curriculum (Total *N* = 305, separate data not given) Baseline data not applicable Classes attended: % (*n*) 0: 28 (86) 1: 7 (20) 2: 4 (13) 3: 4 (11) 4: 6 (17) 5: 52 (158) Diabetes knowledge: Baseline: M = 11.8 6 months: M = 12.4	*Analysis* (R1 + P1) vs (R2 + P2) R vs P Using ANOVA with the following factors: treatment, type of assignment (random/preference), treatment by assignment interaction. For diabetes knowledge model included baseline values as covariate; missing values replaced with baseline values in analysis *Results* *Number of classes attended* (R1 + P1) vs (R2 + P2), $p = 0.09$ R vs P, $p = 0.46$ No treatment by assignment interaction *Diabetes knowledge* (R1 + P1) vs (R2 + P2), $p = 0.59$ R vs P, $p = 0.09$ No treatment by assignment interaction, $p = 0.61$

continued

No.	Study	Primary outcome	Randomised (R)	Preference (P) (non-randomised)	Analysis presented and results (primary outcome only)
29	Rokke, 1999[111]	1. HRSD rating scale for depression 2. BDI 3. GDS 4. Expectancy measures (credibility questionnaire, depression beliefs questionnaire)	R1 = cognitive, *N* = 11 R2 = behavioural, *N* = 9 R3 = control, *N* = 29 Baseline data not reported Data for non-preference group not reported	P1 = cognitive, *N* = 7 P2 = behavioural, *N* = 8 Baseline data not reported Data for preference group not reported	*Analysis* (R1 + P1 + R2 + P2) vs R3 (R1 + P1) vs (R2 + P2)[a] (R1 + R2) vs (P1 + P2)[a] Using ANCOVA adjusted for pretreatment scores *Results* (R1 + P1 + R2 + P2) vs R3 *HRSD*: adjusted means 5.1 versus 10.6, $p < 0.01$ *BDI*: adjusted means 8.4 versus 13.9, $p < 0.05$ *GDS*: adjusted means 9.1 versus 16.1, $p < 0.01$ (depression scores lower for SMT groups than for waiting list control according to each measure) (R1 + P1) vs (R2 + P2)[a] No significant differences in outcome for HRSD, BDI or GDS *HRSD*: adjusted means 6.5 versus 4.1 *BDI*: adjusted means 9.6 versus 8.1 *GDS*: adjusted means 10.2 versus 9.0 (R1 + R2) vs (P1 + P2) No significant differences in outcome for HRSD, BDI or GDS

[a] Comparison included R3 group who were later assigned to choice or no choice arms.

continued

No.	Study	Primary outcome	Randomised (R)	Preference (P) (non-randomised)	Analysis presented and results (primary outcome only)
30	Hardy, 1995[96] Shapiro, 1994[112]	BDI	R1 = 8 sessions PI, *N* = 30 R2 = 8 sessions CB, *N* = 29 R3 = 16 sessions PI, *N* = 28 R4 = 16 sessions CB, *N* = 30 (Total randomised, *N* = 150) Baseline R1: *N* = 30, M = 20.1, SD = 4.5 R2: *N* = 29, M = 20.1, SD = 3.8 R3: *N* = 28, M = 22.7, SD = 5.3 R4: *N* = 30, M = 22.2, SD = 5.6 Treatment end R1: *N* = 29, M = 12.6, SD = 8.1 R2: *N* = 29, M = 8.8, SD = 7.6 R3: *N* = 28, M = 9.3, SD = 7.2 R4: *N* = 29, M = 7.7, SD = 6.2 3 months R1: *N* = 30, M = 11.2, SD = 8.2 R2: *N* = 29, M = 9.5, SD = 9.2 R3: *N* = 26, M = 10.5, SD = 8.3 R4: *N* = 30, M = 8.7, SD = 7.9 (In report data are presented only by severity level. Summary data above were averaged across severity levels)	No preference group	*Analysis* (R1 + R3) vs (R2 + R4) (R1 + R2) vs (R3 + R4) In ANCOVA model including the following factors: treatment type, treatment duration, time (pretreatment, post-treatment, 3 months), treatment type by duration interaction, adjusted for baseline scores, and adjusted for therapist effect *Results* (R1 + R3) vs (R2 + R4): adjusted $p = 0.05$ (lower depression score for CB than PI) (R1 + R2) vs (R3 + R4): adjusted $p = 0.12$ Treatment type by duration interaction not significant *Treatment principle credibility (pretreatment)* CB: M = 2.0, SD = 0.8 PI: M = 1.5, SD = 0.7 *Initial credibility (pretreatment, alternative measure)* CB: M = 5.2, SD = 1.0 PI: M = 5.0, SD = 1.2 CB principle credibility rating negatively correlated with BDI outcome for PI therapy but not for CB therapy (a higher credibility of CB was associated with a better outcome of PI therapy, $r = -0.33$, $p < 0.001$) PI principle credibility rating not associated with BDI outcome for PI or CB therapy Initial credibility rating and emergent credibility rating negatively correlated with BDI outcome for 8-session treatment but not for 16-session treatment

continued

No.	Study	Primary outcome	Randomised (R)	Preference (P) (non-randomised)	Analysis presented and results (primary outcome only)
31	Moffett, 1999[120]	RDS Improvement by at least 3 points on the RDS	R1 = intervention (exercise), *N* = 89 R2 = control (standard GP care), *N* = 98 Baseline R1: *N* = 89, M = 6.65, SD = 4.02 R2: *N* = 98, M = 5.56, SD = 3.94 Change from baseline 6 weeks R1: *N* = 85, M = –2.86 R2: *N* = 94, M = –1.94 6 months R1: *N* = 77, M = –2.99 R2: *N* = 86, M = –1.64 1 year R1: *N* = 83, M = –3.19 R2: *N* = 88, M = –1.77	No preference group	*Analysis* R1 vs R2 Using ANCOVA with change in scores as dependent variable and adjusted for baseline scores. Effect of patient preference also investigated *Results* R2 vs R1 Difference in means for change in score (95% CI) 6 weeks: 0.92 (–0.02 to 1.87), *p* = 0.06. 6 months: 1.35 (0.13 to 2.57), *p* = 0.03 1 year: 1.42 (0.29 to 2.56), *p* = 0.02 Patient preference by treatment interaction not significant R1: change in score –3.10 for those who preferred intervention compared with –3.15 for those who were indifferent; 95% CI for difference (–1.47 to 3.08) R2: change in score –1.93 for those who preferred intervention compared with –1.18 for those who were indifferent; 95% CI for difference (–1.05 to 2.55)
32	Devine, 1973[95]	14-point behavioural rating scale indicating fear of snakes	R1 = systematic desensitisation, *N* = 4 R2 = encounter, *N* = 4 R3 = rational emotive, *N* = 4 R4 = modelling–behaviour rehearsal, *N* = 4 Outcome data for primary measure only in ranks	Preferred P1 = systematic desensitisation, *N* = 4 P2 = encounter, *N* = 4 P3 = rational emotive, *N* = 4 P4 = modelling–behaviour rehearsal, *N* = 4 Non-preferred NP1 = systematic desensitisation, *N* = 4 NP2 = encounter, *N* = 4 NP3 = rational emotive, *N* = 4 NP4 = modelling–behaviour rehearsal, *N* = 4 Outcome data for primary measure only in ranks	*Analysis* (R1 + P1 + NP1) vs (R2 + P2 + NP2) R vs P vs NP Using Kruskal–Wallis and Mann–Whitney *U*-tests *Results* (R1 + P1 + NP1) vs (R2 + P2 + NP2) no significant difference R vs P vs NP, $p < 0.01$ (preferred therapy group has less fear than randomised or non-preferred groups for encounter and rational emotive techniques)

continued

No.	Study	Primary outcome	Randomised (R)	Preference (P) (non-randomised)	Analysis presented and results (primary outcome only)
33	Renjilian, 2001[94]	1. Body weight 2. BMI	R1P = individual therapy, preferred, *N* = 19 R1NP = individual therapy, non-preferred, *N* = 16 R2P = group therapy, preferred, *N* = 20 R2NP = group therapy, non-preferred, *N* = 20 Baseline Weight R1P: *N* = 19, M = 97.0, SD = 14.7 R1NP: *N* = 16, M = 97.5, SD = 13.7 R2P: *N* = 20, M = 99.5, SD = 14.5 R2NP: *N* = 20, M = 94.5 SD = 12.2 BMI R1P: *N* = 19, M = 35.4, SD = 4.2 R1NP: *N* =16, M = 35.6, SD = 4.2 R2P: *N* = 20, M = 37.7, SD = 4.8 R2NP: *N* = 20, M =36.0, SD = 3.8 Post-treatment Weight R1P: *N* = 13, M = 90.4, SD = 15.7 R1NP: *N* = 15, M = 87.2, SD = 13.0 R2P: *N* = 16, M = 87.5, SD = 11.3 R2NP: *N* = 14, M = 83.1, SD = 12.3 BMI R1P: *N* = 13, M = 32.8, SD = 4.5 R1NP: *N* = 15, M = 32.2, SD = 4.6 R2P: *N* = 16, M = 33.1, SD = 4.1 R2NP: *N* = 14, M = 31.8, SD = 4.0 Changes from baseline Weight R1P: *N* = 13, M = –8.5, SD = 3.0 R1NP: *N* = 15, M = –9.6, SD = 4.2 R2P: *N* = 16, M = –10.9, SD = 4.1 R2NP: *N* = 14, M = –11.2, SD = 5.6 BMI R1P: *N* = 13, M = –3.1, SD = 1.1 R1NP: *N* = 15, M = –3.5, SD = 1.2 R2P: *N* = 16, M = –4.1, SD = 1.6 R2NP: *N* = 14, M = –4.3, SD = 2.1	No preference group	*Analysis* R1 vs R2 P vs NP Using MANOVA of change in outcome with the following factors: treatment, preference, treatment by preference interaction *Results* R1 vs R2, adjusted $p = 0.016$ (group therapy group lost more weight than individual therapy group) Effect of preference and treatment by preference interaction not significant

continued

No.	Study	Primary outcome	Randomised (R)	Preference (P) (non-randomised)	Analysis presented and results (primary outcome only)
34	Van Dyck, 1997[93]	Quantitative measurement of tolerance to trigger	R1 = combined therapy, $N = 64$ R2 = exposure only, $N = 64$ (all 64 patients had both therapies in crossover design) Baseline R1 (as first treatment): M = 7.2, SD = 5.0 R2 (as first treatment): M = 7.7, SD = 5.1 Post-treatment R1 (as first treatment): M = 13.9, SD = 14.0 R2 (as first treatment): M = 18.6, SD = 14.3 R1 (as second treatment): M = 25.9, SD = 20.2 R2 (as second treatment): M = 17.4, SD = 12.5	No preference group	*Analyses* R1 vs R2 Using ANOVA with the following factors: time (pretreatment, intermediate, post-treatment), treatment order, treatment preference and interactions *Results* Significant effect of time $p < 0.001$ (improvement over time from baseline to intermediate, and intermediate to post-treatment) No significant effect of order; no order by time interaction, $p = 0.42$ No significant effect of preference; no preference by order by time interaction, $p = 0.76$

ANCOVA, analysis of covariance; ANOVA, analysis of variance; BDI, Beck Depression Inventory; BMI, body mass index; BSC, best supportive care; CB, cognitive-behavioural therapy; CE, conductive education; CT, cognitive therapy; CT, chemotherapy; CVS, chorionic villus sampling; EA, early amniocentesis; GDS, Geriatric Depression Scale; GMFM, Gross Motor Function Measure; HADS, Hospital Anxiety and Depression Scale; HRSD, Hamilton Rating Scale for depression; IQR, Interquartile Range; M, mean; NDC, non-directive counselling; OME, otitis media with effusion; OR, odds ratio; PI, Psychodynamic Interpersonal Psychotherapy; PS, pain scale; RR, relative risk; RDS, Rowland Disability Score; RS, Roland Score; SD, standard deviation; SDD, selective bowel decontamination; SE, standard error; SMT, self-management therapy; TCRE, Transcervical Resection of Endometrium; VAB, Vaple Assessment Battery.

Appendix 7

Effect sizes for intervention-specific randomised versus preference comparisons of outcome

N	Trial	Condition and treatments under investigation	Numbers in each treatment group (i.e. R1 + R2 and P1 + P2)	Baseline effect size	Outcome effect size	Net effect size[a]
60	Gossop, 1986[128]	Opiate addiction 1. Outpatient (counselling and oral methadone) 2. Inpatient (oral methadone)	R + P = 60	N/A	Withdrawal from opiates R vs P: d^* = +0.51	N/A
66	Reddihough, 1998[100]	Young children with cerebral palsy 1. CE; educational programme to develop motor and cognitive skills 2. Traditional neuro-developmental programme (no motor component)	R1 + P1 = 32 R2 + P2 = 34	Cognitive variable[b] R1 vs P1: d = –0.71 R2 vs P2: d = +0.78 GMFM score R1 vs P1: d = +0.07 R2 vs P2: d = +1.46	Cognitive variable[b] R1 vs P1: d = –0.41 R2 vs P2: d = +0.35 GMFM score R1 vs P1: d = –0.27 R2 vs P2: d = +1.28	Cognitive variable[b] R1 vs P1: +0.3 R2 vs P2: –0.43 GMFM score R1 vs P1: –0.34 R2 vs P2: –0.18
66	Bakker, 2000[117]	Patients with panic disorder and agoraphobia 1. Cognitive therapy	R1 + P1 = 66	Panic frequency R vs P: d = +0.24	Panic frequency R vs P: d = +0.32	Panic frequency R vs P: +0.08
70	Riedl, 2001[114]	Patients with cancer of the oesophagus/cardia 1. Selective bowel decontamination 2. Intravenous prophylactic antibiotics only	R1 + P1 = 25 R2 + P2 = 45	N/A	Pneumonia occurrence R1 vs P1: d^* = +0.26 R2 vs P2: d^* = +0.58	
98	Bain, 2001[106]	Microwave endometrial ablation for heavy menstrual bleeding 1. Local anaesthesia 2. General anaesthesia	R1 + P1 = 52 R2 + P2 = 46	N/A	Pain–painful[c] R1 vs P1: d = +0.37 R2 vs P2: d = –0.04 Pleasant–unpleasant[c] R1 vs P1: d = +0.06 R2 vs P2: d = –0.31	N/A

continued

N	Trial	Condition and treatments under investigation	Numbers in each treatment group (i.e. R1 + R2 and P1 + P2)	Baseline effect size	Outcome effect size	Net effect size[a]
144	McKay, 1995[130]	28-day rehabilitation programme for male alcoholics 1. Day hospital outpatients 2. Inpatient rehabilitation	R1 + P1 = 89 R2 + P2 = 55	No. of drinking days R1 vs P1: *d* = +0.30 R2 vs P2: *d* = –0.18	No. of drinking days 3 months R1 vs P1: *d* = –0.40 R2 vs P2: *d* = –0.24 6 months R1 vs P1: *d* = –0.47 R2 vs P2: *d* = +0.31 12 months R1 vs P1: *d* = –0.24 R2 vs P2: *d* = –0.03	No. of drinking days 3 months R1 vs P1: –0.70 R2 vs P2: –0.06 6 months R1 vs P1: –0.77 R2 vs P2: +0.49 12 months R1 vs P1: –0.54 R2 vs P2: +0.15
171	McKay, 1998[129]	28-day, 12-step oriented rehabilitation programme for cocaine users 1. Day hospital outpatients 2. Inpatient rehabilitation	R1 + P1 = 95 R2 + P2 = 76	Days of cocaine use R1 vs P1: *d* = +0.66 R2 vs P2: *d* = +0.34	Days of cocaine use 3 months R1 vs P1: *d* = –0.32 R2 vs P2: *d* = +0.30 6 months R1 vs P1: *d* = –0.10 R2 vs P2: *d* = +0.09 12 months R1 vs P1: *d* = –0.07 R2 vs P2: *d* = +0.20	Days of cocaine use 3 months R1 vs P1: –0.98 R2 vs P2: –0.04 6 months R1 vs P1: –0.76 R2 vs P2: –0.25 12 months R1 vs P1: –0.73 R2 vs P2: –0.14
187	Paradise, 1984[98]	Children with recurrent throat infections 1. Surgery: tonsillectomy ± adenoidectomy 2. Standard care (no surgery)	R1 + P1 = 95 R2 + P2 = 92	No baseline data	No. of episodes of throat infection 12 months R1 vs P1: *d* = –0.26 R2 vs P2: *d* = 0 24 months R1vs P1: *d* = +0.18 R2 vs P2: *d* = +0.09 36 months R1vs P1: *d* =+0.19 R2 vs P2: *d* = –0.60	N/A

continued

N	Trial	Condition and treatments under investigation	Numbers in each treatment group (i.e. R1 + R2 and P1 + P2)	Baseline effect size	Outcome effect size	Net effect size[a]
213	Paradise, 1990[99]	Children with recurrent otitis media 1. Surgery: adenoidectomy in all cases 2. Control: not defined but assumed no treatment beyond initial medication	R1 + R1 = 99 R2 + P2 = 114	No baseline data	No. of episodes of otitis media 12 months R1 vs P1: $d = +0.08$ P1 vs P2; $d = +0.03$ 24 months R1vs P1: $d = +0.33$ R2 vs P2: $d = +0.21$	N/A
227	Cooper, 1997[105]	Women with heavy menstrual bleeding 1. Transcervical resection of the endometrium 2. Medical care: standard medical treatment	R1 + P1 = 114 R2 + P2 = 113	N/A	Treatment acceptability R1 vs P1: $d^* = -0.55$ R2 vs P2: $d^* = +0.64$ Continue same treatment R1 vs P1: $d^* = -0.71$ R2 vs P2: $d^* = +0.74$	N/A
320	Rovers, 2000, 2001[97,101]	Infants with persistent bilateral otitis media 1. Ventilation tubes 2. Watchful waiting	R1 + P2 = 129 R2 + P2 = 191	No baseline data	Occurrence of OME 3 months R1 vs P1: $d^* = -0.38$ R2 vs P2: $d^* = +0.15$ 6 months R1 vs P1: $d^* = +0.34$ R2 vs P2: $d^* = -0.13$ 9 months R1 vs P1: $d^* = -0.03$ R2 vs P2: $d^* = -0.02$ 12 months R1 vs P1: $d^* = +0.12$ R2 vs P2: $d^* = -0.09$	N/A

continued

N	Trial	Condition and treatments under investigation	Numbers in each treatment group (i.e. R1 + R2 and P1 + P2)	Baseline effect size	Outcome effect size	Net effect size[a]
323	Bedi, 2000[109,110]	Patients with depression 1. Counselling 2. Antidepressants	R1 + P1 = 192 R2 + P2 = 131	BDI R1 vs P1: d = +0.18 R2 vs P2: d = +0.18	BDI 8 weeks R1 vs P1: d = +0.07 R2 vs P2: d = +0.08 1 year R1 vs P1: d = +0.40 R2 vs P2: d = +0.24	BDI 8 weeks R1 vs P1: –0.11 R2 vs P2: –0.10 1 year R1 vs P1: +0.22 R2 vs P2: +0.06
363	Henshaw, 1994[102,103,126]	Women requiring abortion 1. Medical abortion with mifepristone and gemeprost 2. Surgical abortion by vacuum aspiration	R1 + P1 = 172 R2 + P2 = 191	Anxiety subscale (HADS) R1 vs P1: d = –0.23 R2 vs P2: d = 0	Anxiety subscale (HADS) R1 vs P1: d = +0.05 R2 vs P2: d = –0.05 Good–Bad[c] R1 vs P1: d = +0.12 R2 vs P2: d = –0.18 % opt for same treatment R1 vs P1: d^* = +1.05 R2 vs P2: d^* = +0.16	Anxiety subscale (HADS) R1 vs P1: +0.28 R2 vs P2: –0.05
445	Ashok, 2002[107]	Women requiring abortion at 10–13 weeks gestation 1. Medical termination with mifepristone and misoprostol 2. Surgical termination with vacuum aspiration	R1 + P1 = 203 R2 + P2 = 242	N/A	Number of days bleeding R1 vs P1: d = +0.27 R2 vs P2: d = +0.07	N/A
464	M King, 2000[26,27d]	People with depression or mixed anxiety and depression 1. Cognitive behavioural therapy 2. Non-directive counselling	R1 + P1 = 215 R2 + P2 = 180	BDI 2-way randomised R1 vs P1: d = +0.08 R2 vs P2: d = +0.02	BDI 2-way randomised R1 vs P1: d = –0.05 R2 vs P2: d = –0.19 12 months R1 vs P1: d = –0.09 R2 vs P2: d = –0.12	BDI 2-way randomised R1 vs P1: –0.13 R2 vs P2: –0.21 12 months R1 vs P1: –0.17 R2 vs P2: –0.14

continued

N	Trial	Condition and treatments under investigation	Numbers in each treatment group (i.e. R1 + R2 and P1 + P2)	Baseline effect size	Outcome effect size	Net effect size[a]
580	Kerry, 2000[118]	Patients with low back pain 1. X-ray with standard treatment 2. Standard treatment alone	R1 + P1 = 168 R2 + P2 = 412	RS R1 vs P1: $d = -0.13$ R2 vs P2: $d = +0.02$ GP consultation R1 vs P1: $d^* = -0.18$ R2 vs P2: $d^* = +0.03$	RS 6 weeks R1 vs P1: $d = -0.15$ R2 vs P2: $d = +0.26$ 12 months R1 vs P1: $d = -0.22$ R2 vs P2: $d = +0.02$ GP consultation 6 weeks R1 vs P1: $d^* = -0.21$ R2 vs P2: $d^* = +0.20$ 1 year R1 vs P1: $d^* = -0.28$ R1 vs P1: $d^* = +0.27$	RS 6 weeks R1 vs P1: –0.22 R2 vs P2: +0.02 12 months R1 vs P1: –0.09 R2 vs P2: 0 GP consultation 6 weeks R1 vs P1: –0.03 R2 vs P2: +0.17 1 year R1 vs P1: –0.10 R1 vs P1: +0.24
596	Noel, 1998[131]	Diabetic educational packages 1. Standard curriculum: 60% non-nutritional management, 40% nutritional 2. Nutritional: 60% nutritional, 40% non-nutritional	R + P = 596	N/A	Attending any classes R vs P: $d^* = +0.08$	N/A
642	Detre, 2000[121]	Coronary artery disease patients with diabetes 1. PTCA 2. CABG	R1 + P1 = 352 R2 + P2 = 290	N/A	All-cause mortality R1 vs P1: $d^* = +0.63$ R2 vs P2: $d^* = +0.18$ Cardiac mortality R1 vs P1: $d^* = +0.73$ R2 vs P2: $d^* = +0.19$	N/A

continued

N	Trial	Condition and treatments under investigation	Numbers in each treatment group (i.e. R1 + R2 and P1 + P2)	Baseline effect size	Outcome effect size	Net effect size[a]
830	SB King, 1995[122–124]	Patients with multi-vessel coronary artery disease 1. PTCA 2. CABG	R1 + P1 = 366 R2 + P2 = 464	N/A	All-cause mortality 1 year R1 vs P1: *d** = +0.99 R2 vs P2: *d** = +0.19 2 years R1 vs P1: *d** = +0.31 R2 vs P2: *d** = +0.58 3 years R1 vs P1: *d** = +0.22 R2 vs P2: *d** = +0.42	N/A
1071	Janevic, 2003[108]	Women with cardiovascular diseases Disease management programmes 1. Self-directed, home-based 2. Group programme	R1 + P1 = 376 R2 + P2 = 511	N/A	Attendance/completion of at least one unit R1 vs P1: *d** = +0.07 R2 vs P2: *d** = +0.83 Attendance/completion level R1 vs P1: *d** =+ 0.11 R2 vs P2: *d** = +0.35 (all from adjusted ORs)	N/A
1301	Nicolaides, 1994[104]	Pregnant women first trimester 1. EA 2. CVS	R1 + P1 = 731 R2 + P2 = 570	N/A	Total loss R1 vs P1: *d** = +0.10 R2 vs P2: *d** = –0.21 Spontaneous loss R1 vs P1: *d** = +0.09 R2 vs P2: *d** = –0.54 Termination R1 vs P1: *d** = +0.13 R2 vs P2: *d** = –0.03	N/A

continued

N	Trial	Condition and treatments under investigation	Numbers in each treatment group (i.e. R1 + R2 and P1 + P2)	Baseline effect size	Outcome effect size	Net effect size[a]
2095	Olschewski, 1992[125]	Coronary artery disease in people who had had a myocardial infarction 1. Medical treatment 2. Surgical treatment	R1 + P1 = 1135 R2 + P2 = 960	N/A	All-cause mortality R1 vs P1: d^* = +0.04 R2 vs P2: d^* = +0.03	N/A

Effect sizes are presented for those trials where comparisons were possible. Cohen[35] defines effect sizes as small = 0.2, medium = 0.5 and large = 0.8. Key: P1/P2 = preference groups; R1/R2 = randomised groups. d = standardised effects size (Cohen's d); d^* = effect size for dichotomous variables (natural log of odds ratio/1.81). + = outcome goes **in favour** of preference group (i.e. preference group less severely affected); – = outcome goes **against** the preference group (i.e. preference group more severely affected).
Abbreviations: BDI, Beck Depression Inventory; CE, conductive education; CVS, chorionic villus sampling; EA, early amniocentesis; GMFM, Gross Motor Function Measure; HADS, Hospital Anxiety and Depression Scales; OME, otitis media with effusion; OR, odds ratio; RS, Roland Score.
[a] Net effect size (outcome effect size minus baseline effect size).
[b] Cognitive variable of the Vaple Assessment Battery.
[c] Patients' views of treatment experience based on items from Semantic Differential Scales.
[d] This study used two randomised cohorts: one randomised three ways between three treatments on offer and one randomised two ways, with the 'standard care' option omitted.

Appendix 8

Search strategies

This Appendix specifies the MEDLINE, EMBASE, PsycINFO and CINAHL search strategies.

MEDLINE search strategy 1 (1966–2002)

1. exp consumer satisfaction/
2. preference$.tw.
3. 1 or 2
4. exp attitude
5. attitude$.tw.
6. knowledge, attitudes, practice/
7. knowledge.tw.
8. expecta$.tw.
9. placebos/
10. placebo$.tw.
11. exp "effect modifiers (epidemiology)"/
12. (hawthorne adj effect$).tw.
13. (halo ADJ effect$).tw.
14. exp decision making/
15. choice behavior/
16. choice$.tw.
17. exp risk/
18. risk adjustment/
19. exp risk-taking/
20. (perception adj2 risk$).tw.
21. (risk adj perception$).tw.
22. risk$.tw.
23. ((treatment$ or therap$ or intervention$) adj (choice$ or alternativ$ or outcome$ or options)).tw.
24. ((choice$ or alternativ$ or option$ or outcome$) adj2 (treatment$ or therap$ or intervention$))
25. exp informed consent/
26. (informed adj choice$).tw.
27. (informed adj participa$).tw.
28. framing.tw.
29. ((previous or past) adj (experience$ or event$ or episode$ or cases$ or outcome$))
30. (clinical adj experience$).tw.
31. (effective$ adj2 (treatment$ or therap$ or interven$)).tw.
32. information.tw.
33. exp patient acceptance of healthcare/
34. (treatment adj refusal).tw.
35. (refus$ adj2 (treatment$ or therap$ or intervention$ or surg$)).tw.
36. complian$.tw.
37. patient-centered care/
38. (patient$ adj cent$).tw.
39. exp socioeconomic factors/
40. (role adj2 preferences).tw.
41. understand$.tw.
42. (outcome adj effect$).tw.
43. (effect$ adj2 outcome$).tw.
44. exp professional-patient relations/
45. ((doctor$ or physician$ or nurse$ or practitioner$ or clinician$ or expert$)
46. patient care management/
47. patient selection/
48. physician's practice patterns/
49. (assessment adj2 location$).tw.
50. (normative adj belie$).tw.
51. (measurement$ adj preference$).tw.
52. utilit$.tw.
53. or/4-52
54. 3 and 53
55. Limit 54 to randomized controlled trial
56. Limit 54 to meta analysis
57. Limit 54 to controlled clinical trial
58. Limit 54 to clinical trial
59. random$.tw.
60. (meta-anal$ or metaanaly$ or meta analy$).tw.
61. exp clinical trials/
62. cross-over studies/
63. cross adj over.tw.
64. or/55-63
65. 54 and 64

MEDLINE Search Strategy 2 (1966–2002)

1. zelen.tw.
2. (single OR double) ADJ (random$ ADJ consent$).tw.
3. (non ADJ random$ ADJ assign$).tw.
4. (nonrandom$ ADJ assign$).tw.
5. (comprehensive ADJ cohort$ ADJ stud$).tw.
6. (partial$ ADJ randomi?ation).tw.
7. (nonrandom$ ADJ (patient$ OR participant$)).tw.
8. (preference ADJ arm$).tw.
9. OR/1-9

EMBASE (1980–week 13 2003)

1. exp physician attitude
2. exp nurse attitude
3. exp Patient Attitude
4. utilit$.ti.
5. preference$.ti.
6. choice$.ti.
7. 1 or 2 or 3 or 4 or 5 or 6
8. exp attitude
9. attitude$.mp.
10. knowledg$.mp.
11. expectation/
12. exp expectation/
13. exp Placebo/
14. placebo$.mp.
15. (effect$ adj modifier$).mp.
16. (hawthorne adj effect$).mp.
17. (halo adj effect$).mp.
18. exp decision making
19. medical decision making
20. decision$.mp.
21. choice$.mp.
22. exp RISK MANAGEMENT/ or exp SURGICAL RISK/ or exp RISK/ or exp RISK ASSESSMENT/
23. risk$.mp.
24. (perception adj2 risk$).mp.
25. (risk adj perception).mp.
26. ((treatment$ or therap$ or intervention$) adj (choice$ or alternativ$ or option$ or outcome$)).mp.
27. ((choice$ or alternativ$ or option$ or outcome$) adj2 (treatment$ or therap$ or intervention$)).mp.
28. exp informed consent/
29. (informed adj (consent$ or choice$)).mp.
30. (informed adj participa$).mp.
31. framing.mp.
32. ((previous or past) adj (experience$ or event$ or episode$ or case$ or outcome$)).mp.
33. (clinical adj experience$).mp.
34. (effective$ adj (treatment$ or therap$ or intervention$)).mp
35. information.mp.
36. (illness adj behavior).mp.
37. exp Illness Behavior/
38. (refus$ adj2 (treatment$ or therap$ or intervention$ or surg$)).mp.
39. exp Patient Compliance/
40. (patient$ adj (participation or centered or centred)).mp.
41. exp social class/ or exp social structure/ or exp socioeconomics/
42. (role adj preference$).mp.
43. understand$.mp.
44. (outcome adj effect$).mp.
45. (effect$ adj2 outcome$).mp.
46. exp doctor patient relation/
47. exp nurse patient relationship/
48. exp patient care/
49. clinical practice/
50. (assessment adj location$).mp.
51. exp quality of life/
52. (quality adj2 life).mp.
53. or/8-52
54. 7 and 53
55. (((meta$ analys$ or metaanalys$ or systematic) adj review$) or overview$).mp.
56. 54 and 55
57. from 56 keep 1

PsycINFO (1972–March week 4 2003)

1. exp preferences/ or preference measures
2. (preference$ or choice$).ti.
3. (preference$ or choice$).mp.
4. 1 or 2
5. exp choice behavior
6. 4 or 5
7. exp CONSUMER ATTITUDES/ or exp HEALTH ATTITUDES/ or exp COUNSELOR ATTITUDES/ or exp HEALTH PERSONNEL ATTITUDES/ or exp DEATH ATTITUDES/ or exp "PHYSICAL DISABILITIES (ATTITUDES TOWARD)"/ or exp "PHYSICAL ILLNESS (ATTITUDES TOWARD)"/ or exp CLIENT ATTITUDES/ or exp PSYCHOLOGIST ATTITUDES/ or exp THERAPIST ATTITUDES/ or exp PSYCHOTHERAPIST ATTITUDES/ or exp *ATTITUDES/
8. attitude$.mp.
9. exp HEALTH KNOWLEDGE/ or exp KNOWLEDGE LEVEL
10. knowledg$.mp.
11. exp expectations/ or exp experimenter expectations/ or exp parental expectations/ or exp halo effect/ or exp "experiences (events)
12. expect$.mp.
13. exp *PLACEBO
14. exp PLACEBO
15. placebo$.mp.
16. (effect adj modifiers).mp.
17. (effect adj modifier).mp.
18. (hawthorne adj effect$).mp.
19. exp decision making
20. exp RISK PERCEPTION/ or exp RISK TAKING
21. risk$.mp.
22. (perception adj2 risk$).mp.
23. ((treatment$ or therap$ or intervention$) adj

(choice$ or alternat$ or option$ or outcome$)).mp.
24. ((choice$ or alternat$ or option$ or outcome$) adj (treatment$ or therap$ or intervention$)).mp.
25. exp informed consent
26. (informed adj choice$).mp.
27. (informed adj participa$).mp.
28. framing.mp.
29. ((previous or prior or past) adj (experience$ or event$ or episode$ or outcome$)).mp.
30. (effective$ adj2 (treatment$ or therap$ or interven$)).mp.
31. information.mp.
32. exp Treatment Refusal
33. (refus$ adj2 (treatment$ or therap$ or intervention$)).mp.
34. (refus$ adj2 surg$).mp.
35. complian$.tw.
36. (patient adj center$).mp.
37. exp sociocultural factors/ or exp socioeconomic status/ or exp disadvantaged/ or exp "income (economic)"/ or exp poverty/ or exp socioeconomic class attitudes
38. demographic factors.mp.
39. exp roles/ or role conflicts/ or role expectations/ or role models/ or role perception/ or role satisfaction/ or role taking
40. (role adj preference$).mp.
41. exp Comprehension
42. understand$.mp.
43. exp Psychosocial Factors
44. (outcome adj effect$).mp.
45. (effect adj2 outcome$).mp.
46. exp interpersonal communication/ or exp interpersonal interaction
47. (professional adj patient$).mp.
48. (patient adj professional).mp.
49. exp patient selection/ or client characteristics/ or exp client treatment matching/ or exp therapist selection
50. (practice adj pattern$).mp.
51. (assessment adj location).mp.
52. or/7-51
53. 1 or 2 or 5
54. 1 or 3 or 5
55. 52 and 53
56. 52 and 54
57. limit 55 to human
58. from 57 keep 5-6,14,41-42,44-45,61,65,71,122,160,193
59. from 57 keep 192
60. ("3200" or "3300" or "3400").cc.
61. 57 and 60
62. exp quality of life/ or exp life satisfaction/ or exp lifestyle/ or exp lifestyle changes/ or exp well being/
63. (quality adj2 life).mp.
64. 62 or 63
65. 52 or 64
66. 53 and 65
67. limit 66 to human
68. 60 and 67
69. utilit$.ti.
70. 4 and 69
71. exp Treatment Effectiveness Evaluation
72. exp treatment outcomes
73. exp Psychotherapeutic Outcomes
74. exp PLACEBO/ or placebo.mp.
75. exp Followup Studies
76. random$.tw.
77. (comparative adj stud$).mp.
78. (randomi?ed adj controlled adj trial$).tw.
79. (clinical adj3 trial$).mp.
80. (research adj3 design).tw.
81. (evaluat$ adj3 stud$).tw.
82. (prospective adj3 stud$).mp.
83. 71 or 72 or 73 or 74 or 75 or 76 or 77 or 78 or 79 or 80 or 81 or 82
84. 57 and 83
85. 1 or 2 or 5 or 69
86. 52 and 65 and 85
87. 83 and 86
88. 60 and 87

CINAHL (1982–2001)

1. exp Decision Making, Patient/
2. (preference$ or choice$).ti.
3. 1 or 2
4. exp Attitude/
5. attitude$.mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
6. exp Knowledge/
7. knowledge.mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
8. expecta$.mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
9. placebos/
10. Placebo$.mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
11. (effect adj modifier$).mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
12. exp Hawthorne Effect/
13. (hawthorne adj effect).mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
14. exp Halo Effect/
15. halo adj effect.mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
16. exp Decision Making/ or Exp Decision Making. Clinical/ or exp Decision Making, Ethical/ or Decision Making, Patient/

17. decision$.mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
18. (choice adj behavio?r).mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
19. Risk Factors/ or Relative Risk/ or Risk Assessment/ or Attributable Risk/ or Risk Taking Behavior/ or Risk Management
20. risk$.mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
21. (risk adj perception).mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
22. (perception adj2 risk).mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
23. ((treatment$ or therap$ or intervent$) adj (choice$ or alternat$ or option$ or outcome$)).mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
24. ((choice$ or alternative$ or option$ or outcome$)) adj2 (treatment$ or therap$ or intervent$)).mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
25. exp Consent/
26. (informed adj consent).mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
27. (informed adj choice$).mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
28. (informed adj participa$).mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
29. framing.mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
30. ((previous or past) adj (experience$ or event$ or episode$ or case$ or outcome$)).mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
31. (effective$ adj2 (treatment$ or therap$ or interven$)).mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
32. Access to Information/ or Information Resources/ or Information Retrieval/ or Health Information/ or Consumer Health Information/ or Drug Information/ or Drug Information Services/ or Information Needs/
33. information.mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
34. Treatment Refusal/
35. (treatment adj refus$).mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
36. (refus$ adj2 (treatment$ or therap$ or intervention or surg$)).mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
37. Medication compliance/ or Patient Compliance/ or Professional Compliance/
38. complian$.mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
39. Patient Centered Care/
40. (patient adj cent$).mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
41. Minority Groups/ or Social Environment/ or Social Isolation/ or Social Problems/ or Social Welare/ or exp Socioeconomic Factors/
42. role adj preferences.mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
43. understand$.mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
44. (effect$ adj2 outcome$).mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
45. exp Professional-Client Relations/ or exp Professional-Family Relations/
46. (professional$ adj patient$).mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
47. exp Patient Selection
48. exp Quality of Life
49. (quality adj2 life).mp. [mp=title, Cinahl subject heading, abstract, instrumentation]
50. or/4-49
51. 3 and 50

Appendix 9

List of journals that were hand searched

1. *Lancet*
2. *BMJ*
3. *New England Journal of Medicine*
4. *JAMA*
5. *Controlled Clinical Trials*
6. *Pain*

Appendix 10

Data extraction form: systematic review

For ALL sections, report page number (in left margin) from article & colour highlight in the paper

Reviewer:

1. General Information

i. Author(s):

ii. Title:

iii. Source:

1a. Population

i. Target Population:

ii. Inclusion criteria [Brief note: report presence or absence and any issues related to preference]:

iii. Exclusion criteria [Brief note: report presence or absence and any issues related to preference]:

iv. Intervention Site [Including whether Primary/Secondary care, etc.]:

v. Recruitment procedure [Brief note i.e.: "GP referral"]

vi. Participation rate: [all details provided]

1b. Participant Characteristics

i. Age – (Mean, SD)

ii. Ethnicity (% White/Other):

iii. Gender Distribution:

iv. Other info (Social class & Education level only):

2. Study Design

i. What was the study design [Include No. & Type of conditions & No. of P's per condition. Use flow diagram]

ii. Results of power calculations (if any): YES /NOT REPORTED.
iii. Was this based on the primary measure? YES/ NO/ CAN'T TELL.
iv. Were these based on
   – The randomised cohort?
   – The preference cohort?
   – The entire cohort?
   – Don't know/unclear.
v. Was Randomisation/Preference offered simultaneously? Yes/No/Can't tell
vi. If not, when?

2a. Assessment of Study Quality

i. What Randomisation Method was used?

ii. Quality of randomisation
Allocation concealment [from researcher]
OK [Central randomisation, or explicit statement of "sealed opaque envelopes"]
UNCLEAR [statement of "sealed envelopes" only, "2 envelopes", "list" or "table", where concealment is unclear]
VULNERABLE [use of alternation, date of birth, where concealment is known to be vulnerable]
iii. Were Participants masked to treatment? Yes/No/Can't tell/Not Possible

iv. Were Health Workers masked to treatment? Yes/No/Can't tell/Not Possible

v. Were Research Assessors masked to treatment? Yes/No/Can't tell

vi. Note if, and how, the study examined quality control in the intervention:

3. Interventions

i. Focus of Intervention:

ii. Content of Intervention Package:

iii. Special Instructions given:

iv. Duration of Intervention

v. Primary Staff in Intervention [i.e.: Nurse, Psychologist, GP]

4. Baseline (including Demographic) Measures

i. List MAIN (Primary and secondary) instruments/measures used (Note if previously published or developed for the study):

4a. Baseline Comparisons (Demographic, Clinical and Other) between Randomised and Non-Randomised groups:

i. Comparisons between Randomised groups alone (R_1 vs. R_2)

Was this comparison made? Yes/No
Significant findings?

ii. Comparisons between Preference groups alone (P_1 vs. P_2)

Was this comparison made? Yes/No
Significant findings?

iii. Randomisation vs. Preference by Treatment (R_1 vs. P_1 and R_2 vs. P_2)
[This section requires data values as given in article]

iv. Randomisation vs. Preference by Allocation ($R_1 + R_2$ vs. $P_1 + P_2$)
[This section requires data values as given in article]

v. Other comparisons or Baseline Measures reported with significant differences:

vi. How was preference *elicited* [Include details of how options were presented and level of detail given to P's. Describe in as much detail as possible]

vii. How was preference *measured* (If scales were used note results)?

5. Outcome Analyses – Statistical Tests Used

i. Describe analyses used and the way attrition was dealt with in the analyses [Brief summary in note form]

Was intention to treat analysis used? Yes/No

ii. List outcome variables that have not been examined at baseline and instruments used for these. (Brief note. Also note if previously published or developed for the study):

5a. Outcome Comparisons

Follow-up point: *Primary Measure:*

i. Comparisons between Randomised groups alone (R_1 vs. R_2)
[Report Overall and Group N, Mean and SD for Primary measure and for *significant* or *main* secondary outcomes]

ii. Comparisons between Preference groups alone (P_1 vs. P_2)
[Report Overall and Group N, Mean and SD for Primary measure and for *significant* or *main* secondary outcomes]

iii. Randomisation vs. Preference by Treatment (R_1 vs. P_1 *and* R_2 vs. P_2)
[Report Overall and Group N, Mean and SD for Primary measure and for *significant* or *main* secondary outcomes]

iv. Treatment independent of Randomisation ($R_1 + P_1$ vs. $R_2 + P_2$)
[Report Overall and Group N, Mean and SD for Primary measure and for *significant* or *main* secondary outcomes]

v. Other Comparisons (*of interest*) reported [i.e.: "Regression analysis" or "Longitudinal results graphed"]

5b. Further Outcome Details

i. How many (%) were lost to follow up? (Use primary outcome measure if poss. Flow Diagram if poss.)

ii. What were the reported economic analysis results (if any)?

Was it done?	Yes/No
Did they compare Randomised vs Preference?	Yes/No

Results:

iii. Methodological weaknesses/limitations identified by authors or reviewers. [These only involve generic issues directly or somehow related to preference]

Identified by: Author(s)/Reviewer

6. **Reported Effects of Preference on Outcome, If Any.** (Brief Summary)

7. **Was There Discussion on Uptake in the Trial and How it Related to Preference?** [Brief note]

Appendix 11

Citations and publications included in and excluded from the review

Citations excluded (*N* = 101)

After review of the full text of these citations, they were excluded because they were:

1. Review or discussion papers.
2. Descriptive studies or surveys with no experimental element(s).
3. The studies did not have a preference cohort or did not have any assessment and analysis of preference within a standardised randomised trial.

Arpino VJ, Giddon DB, BeGole EA, Evans CA. Presurgical profile preferences of patients and clinicians. *Am J Orthodont Dentofacial Orthop* 1998; **114**:631–7.

Bailey JD, Leissner GT. Who likes what in health care design? A survey of preferences. *Health Facilities Manage* 1952;**8**:48–50.

Barnett B. Clients prefer method choices. *Network* 1998;**19**(1):14–18.

Barron DJ, Tolan MJ, Lea RE. A randomized controlled trial of continuous extra-pleural analgesia post-thoracotomy: efficacy and choice of local anaesthetic. *Eur J Anaesthesiol* 1999; **16**:236–45.

Bechel DL, Myers WA, Smith DG. Does patient-centered care pay off? *J Comm J Qual Improve* 2000;**26**:400–9.

Beel TL, Mitchiner JC, Frederiksen SM, McCormick J. Patient preferences regarding pain medication in the ED. *Am J Emerg Med* 2000; **18**:376–80.

Benbassat J, Pilpel D, Tidhar M. Patients' preferences for participation in clinical decision making: a review of published surveys [Review, 53 refs]. *Behav Med* 1998;**24**:81–8.

Bergstrom N. Development of guidelines: contribution of patient's preferences. In *Nursing informatics: combining clinical practice guidelines and patient preferences using health informatics. Proceedings of the Sixth International Nursing Informatics Symposium Post-Conference, Lidingo, Sweden, October 1–4 1997*. 1997; Spring 1998.

Bini EJ, Unger JS, Rieber JM, Rosenberg J, Trujillo K, Weinshel EH. Prospective, randomized, single-blind comparison of two preparations for screening flexible sigmoidoscopy. *Gastrointest Endosc* 2000;**52**:218–22.

Bloch KE, Iseli A, Zhang JN, Xie X, Kaplan V, Stoeckli PW, *et al*. A randomized, controlled crossover trial of two oral appliances for sleep apnea treatment. *Am J Respir Crit Care Med* 2000; **162**:246–51.

Boere-Boonekamp MM, Linden-Kuiper LT LT. Positional preference: prevalence in infants and follow-up after two years. *Pediatrics* 2001; **107**:339–43.

Booske BC, Sainfort F, Hundt AS. Eliciting consumer preferences for health plans. *Health Serv Res* 1999;**34**:839–54.

Borgeat A, Perschak H, Bird P, Hodler J, Gerbe C. Patient-controlled interscalene analgesia with ropivacaine 0.2% versus patient-controlled intravenous analgesia after major shoulder surgery: effects on diaphragmatic and respiratory function. *Anesthesiology* 2000;**92**:102–8.

Bradley C. Patient preferences and clinical trial design and interpretation: appreciation and critique of a paper by Feine, Awad & Lund [Letter; comment] [see comments]. *Commun Dent Oral Epidemiol* 1999;**27**:85–8.

Brennan PF, Moore SM. Health-related decision-making by patients. In Saba VK, McCormick KA. editors. *Essentials of computers for nurses: informatics for the new millennium*. New York: McGraw-Hill. 2001. pp. 89–100.

Brennan PF, Bonair A, Fernandez AB, Nordstrand K, Royle J, Ruland CM. Using patient preferences: the role of nursing informatics:

summary from group 1. In *Nursing informatics: combining clinical practice guidelines and patient preferences using health informatics. Proceedings of the Sixth International Nursing Informatics Symposium Post-Conference, Lidingo, Sweden, October 1–4 1997*. 1997; Spring 1997.

Brennan PF, Strombom I. Improving health care by understanding patient preferences: the role of computer technology. *J Am Med Inf Assoc* 1998; **5**:257–62.

Bridwell KH, Shufflebarger HL, Lenke LG, Lowe TG, Betz RR, Bassett GS. Parents' and patients' preferences and concerns in idiopathic adolescent scoliosis: a cross-sectional preoperative analysis. *Spine* 2000;**25**:2392–9.

Broumas AG, Basara LA. Potential patient preference for 3-day treatment of bacterial vaginosis: responses to new suppository form of clindamycin. *Adv Ther* 2000;**17**:159–66.

Brown MM, Brown GC, Sharma S, Garrett S. Evidence-based medicine, utilities, and quality of life [Review, 29 refs]. *Curr Opin Ophthalmol* 1999; **10**:221–6.

Bruera E, Sweeney C, Calder K, Palmer L, Benisch-Tolley S. Patient preferences versus physician perceptions of treatment decisions in cancer care. *J Clin Oncol* 2001;**19**:2883–5.

Brundage MD, Davidson JR, Mackillop WJ, Feldman-Stewart D, Groome P. Using a treatment-tradeoff method to elicit preferences for the treatment of locally advanced non-small-cell lung cancer. *Med Decis Making* 1998;**18**:256–67.

Brundage MD, Feldman-Stewart D, Cosby R, Gregg R, Dixon P, Youssef Y, *et al*. Phase I study of a decision aid for patients with locally advanced non-small-cell lung cancer. *J Clin Oncol* 2001; **19**:1326–35.

Bruzzi P, Costantini, M. Giving medicine a fair trial. Patients' perspective must be acknowledged. *BMJ* 2000;**321**:1530.

Bryan S, Gold L, Sheldon R, Buxton M. Preference measurement using conjoint methods: an empirical investigation of reliability. *Health Econ* 2000;**9**:385–95.

Butters JM, Winter PA. The effects of gender and race on practice pattern preferences of dental students. *J Am Coll Dent* 1999;**66**:39–46.

Button, F. Women's health. Making choices. *Nurs Manage* 1999;**5**:23–7.

Caress A, Luker KA, Ackrill P. Patient-sensitive treatment decision-making? Preferences and perceptions in a sample of renal patients … including commentary by Meyer J. *NT Res* 1998; **3**:364–73.

Carr JE, Nicolson AC, Higbee TS. Evaluation of a brief multiple-stimulus preference assessment in a naturalistic context. *J Appl Behav Anal* 2000; **33**:353–7.

Cassell EJ. Patient choices and medical decision-making: the ideal versus the reality. *Pride Inst J Long Term Home Health Care* 1993;**12**:31–6.

Chakraborty G, Gaeth GJ, Cunningham M. Understanding consumers' preferences for dental service [Review, 34 refs]. *J Health Care Marketing* 1993;**13**:48–8.

Chandler PJ, Chandler C, Dabbs ML. Provider gender preference in obstetrics and gynecology: a military population. *Mil Med* 2000;**165**:938–40.

Chou EL. Predictors of treatment acceptability, willingness to see a counselor, and counselor preferences for asian-americans and whites: acculturation, loss of face, self-construals, and collective self-esteem. *Diss Abstr International: Section B: The Sciences & Engineering* 2000; **60**(8-B):4209.

Christensen AJ, Ehlers SL, Raichle KA, Bertolatus JA, Lawton WJ. Predicting change in depression following renal transplantation: effect of patient coping preferences. *Health Psychology* 2000;**19**:348–53.

Churchill R, Khaira M, Gretton V, Chilvers C, Dewey M, Duggan C, *et al*. Treating depression in general practice: factors affecting patients' treatment preferences. *Br J Gen Pract* 2000; **50**:905–6.

Clement S, Sikorski J, Wilson J, Candy B. Incorporating patient preferences into clinical trials. Merits of alternative strategies for incorporating patient preferences into clinical trials must be considered carefully [Letter; comment]. *BMJ* 1998;**317**:78–9.

Coppola KM, Ditto PH, Danks JH, Smucker WD. Accuracy of primary care and hospital-based physicians' predictions of elderly outpatients' treatment preferences with and without advance directives. *Arch Intern Med* 2001;**161**:431–40.

Dahler-Eriksen K, Dahler-Eriksen BS, Lassen JF, Olesen F. Is the randomized controlled trial overvalued as a basis for clinical decision-making? A review with comments [Review, 30 refs]. *Ugeskr Laeger* 1998;**160**:7414–17 (in Danish).

Dahlof C. How to assess patient preference of migraine treatments. *Cephalalgia* 1999; **19** Suppl 24:2–5.

Darrouzet V, Duclos JY, Portmann D, Bebear JP. Preference for the closed technique in the management of cholesteatoma of the middle ear in children: a retrospective study of 215 consecutive patients treated over 10 years. *Am J Otol* 2000;**21**:474–81.

Desbiens NA, Mueller-Rizner N, Connors AF Jr, Wenger NS, Lynn J. The symptom burden of seriously ill hospitalized patients. SUPPORT Investigators. Study to Understand Prognoses and Preferences for Outcome and Risks of Treatment. *J Pain Symptom Manage* 1999;**17**:248–55.

Detmar SB, Aaronson NK, Wever LD, Muller M, Schornagel JH. How are you feeling? Who wants to know? Patients' and oncologists' preferences for discussing health-related quality-of-life issues. *J Clin Oncol* 2000;**18**:3295–301.

Donaldson C, Hundley V, Mapp T. Willingness to pay: a method for measuring preferences for maternity care? *Birth* 1998;**25**:32–9.

Dwight-Johnson M, Sherbourne CD, Liao D, Wells KB. Treatment preferences among depressed primary care patients. *J Gen Intern Med* 2000; **15**:527–34.

Eckerlund I, Eklof J, Nathorst-Boos J. Patient preferences or cost-effectiveness? Priorities are highly affected by introduction of economic dimension. *Lakartidningen* 2000;**97**:5782–6 (in Swedish).

Erlen JA. Treatment decision making: who should decide? *Orthop Nurs* 1998;**17**:60–4.

Estabrooks C, Goel V, Thiel E, Pinfold P, Sawka C, Williams I. Decision aids: are they worth it? A systematic review [Review, 47 refs]. *J Health Serv Res Policy* 2001;**6**:170–82.

Faller H, Vogel H, Bosch B. Patient expectations regarding methods and outcomes of their rehabilitation – a controlled study of back pain- and cancer patients. *Rehabilitation* 2000;**39**:205–14 (in German).

Filip J, McGillen C, Mosca L. Patient preferences for cardiac rehabilitation and desired program elements. *J Cardiopulm Rehabil* 1999;**19**:339–43.

Finkelstein BS, Singh J, Silvers JB, Marrero U, Neuhauser D, Cuttler L. Patient attitudes and preferences regarding treatment: GH therapy for childhood short stature. *Horm Res* 1999; **51** Suppl 1:67–72.

Fischhoff, B. What do patients want? Help in making effective choices. *Effect Clin Pract* 1999; **2**:198–200.

Gibb S, Donaldson C, Henshaw R. Assessing strength of preference for abortion method using 'willingness to pay': a useful research technique for measuring values. *J Adv Nurs* 1998;**27**:30–6.

Gramelspacher GP, Zhou XH, Hanna MP. Preferences of physicians and their patients for end-of-life care. *J Gen Intern Med* 1997;**12**:346–51.

Gready RM, Ditto PH, Danks JH, Coppola KM, Lockhart LK, Smucker WD. Actual and perceived stability of preferences for life-sustaining treatment. *J Clin Ethics* 2000;**11**:334–46.

Gregory S. Commentary. Whose choice is it? Supporting the parents' right to decide. *Neonatal Network – J Neonatal Nurs* 2001;**20**:53–4.

Hansdottir H, Gruman C, Curry L, Judge JO. Preferences for CPR among the elderly: the influence of attitudes and values. *Conn Med* 2000; **64**:625–30.

Harrison DE, Galloway S, Graydon JE, Palmer-Wickham S, Rich V, Bij L. Information needs and preference for information of women with breast cancer over a first course of radiation therapy. *Patient Educ Couns* 1999;**38**:217–25.

Harvey RM, Kazis L, Lee AF. Decision-making preference and opportunity in VA ambulatory care patients: association with patient satisfaction. *Res Nurs Health* 1999;**22**:39–48.

Havranek EP, McGovern KM, Weinberger J, Brocato A, Lowes BD, Abraham WT. Patient preferences for heart failure treatment: utilities are valid measures of health-related quality of life in heart failure. *J Cardiac Fail* 1999;**5**:85–91.

Hoare K, Sousa KH, Person L, De Ryk P, Piper J. Comparing three patient-controlled analgesia methods. *MEDSURG Nurs* 2000;**9**:33–9.

Howell CJ, Kidd C, Roberts W, Upton P, Lucking L, Jones PW, *et al*. A randomised controlled trial of epidural compared with non-epidural analgesia in labour [see comments]. *Bjog* 2001;**108**:27–33.

King MT, Hall JP, Harnett PR. A randomised crossover trial of chemotherapy in the home: patient preferences and cost analysis. *Med J Aust* 2001;**174**:312–13.

Kinmonth AL, Woodcock A, Griffin S, Spiegal N, Campbell MJ. Randomised controlled trial of patient centred care of diabetes in general practice: impact on current wellbeing and future disease risk. The Diabetes Care From Diagnosis Research Team. *BMJ* 1998;**317**:1202–8.

Kirigia JM. Economic evaluation in schistosomiasis: valuation of health states preferences. A research note. *Health Econ* 1998; **7**:551–6.

Lam RW. Patients' preferences and counselling for depression in primary care. *Lancet* 2001;**357**:575–6.

Lenert LA, Treadwell JA, Schwartz CE. Associations between health status and utilities: indirect evidence for response shift. In Schwartz CE. Sprangers. MAG (eds) *Adaptation to changing health; response shift in quality of life research*. Washington DC: American Psychological Association; 2000. pp. 123–36.

McNutt RA. Measuring patient preferences for health outcomes: a decision analytic approach. *Patient Educ Couns* 1989;**13**:271–9.

McPherson K, Chalmers I. Incorporating patient preferences into clinical trials. Information about patients' preference must be obtained first [Letter; comment]. *BMJ* 1998;**317**:78–9.

McPherson K, Britton A. The impact of patient treatment preferences on the interpretation of randomised controlled trials [Review, 20 refs]. *Eur J Cancer* 1999;**35**:1598–602.

Morgan H. Your choice or mine? *Ment Health Care Learn Disabil* 2000;**4**:64–7.

Mystakidou K. Quality of life as a parameter determining therapeutic choices in cancer care in a Greek sample. *Palliat Med* 1999;**13**:385–92.

Nease RF Jr. Challenges in the validation of preference-based measures of health-related quality of life [Letter; comment]. *Med Care* 2000; **38**:II155–9.

Neumann PJ, Goldie SJ, Weinstein MC. Preference-based measures in economic evaluation in health care [Review, 138 refs]. *Annu Rev Public Health* 2000;**21**:587–611.

Nield-Anderson L, Dixon JK, Lee K. Random assignment and patient choice in a study of alternative pain relief for sickle cell disease. *West J Nurs Res* 1999;**21**:266–74.

Odile-Carrère M, Moumijid-Ferdjaoui N, Charavel M, Mond A. Eliciting patients' preferences for adjuvant chemotherapy in breast cancer: development and validation of a bedside decision-making instrument in a French Regional Cancer Centre. *Health Expect* 2000;**3**:97–113.

O'Neill S, Bojke C, Wilson RG, Winter M, Jones NK, Purves IN, *et al*. Eliciting user preferences via survey using conjoint analysis. Experiences and results from the NHS Repeat Prescribing Project. In Bryant J, editor. *Current Perspectives in Healthcare Computing Conference, Harrogate, 20–22 March 2000*. Br Comput Society Health Inform Committee Conference Proceedings; 2000.

Olsen JA. A note on eliciting distributive preferences for health. *J Health Econ* 2000; **19**:541–50.

Palmer CS, Schmier JK, Snyder E, Scott B. Patient preferences and utilities for 'off-time' outcomes in the treatment of Parkinson's disease. *Qual Life Res* 2000;**9**:819–27.

Patrick DL, Pearlman RA, Starks HE, Cain KC, Cole WG, Uhlman RF. Validation of preferences for life-sustaining treatment: implications for advance care planning. *Ann Intern Med* 1997;**127**:509–17.

Pignone M, Bucholtz D, Harris R. Patient preferences for colon cancer screening. *J Gen Intern Med* 1999;**14**:432–7.

Protheroe J, Fahey T, Montgomery AA, Peters TJ. Effects of patients' preferences on the treatment of atrial fibrillation: observational study of patient-based decision analysis. [see comments]. *Western Journal of Medicine* 2001;**174**:311–15.

Ruland CM. Integrating patient preferences in nurses' care decisions: clinical research and its implications for nursing informatics. *Nursing informatics: combining clinical practice guidelines and*

patient preferences using health informatics. Proceedings of the Sixth International Nursing Informatics Symposium Post-Conference, Lidingo, Sweden, October 1–4, 1997. 2001, Spring.

Ryan M. Using conjoint analysis to take account of patient preferences and go beyond health outcomes: an application to in vitro fertilisation. *Social Science & Medicine* 1999;**48**:535–46.

Ryan M, Farrar S. Using conjoint analysis to elicit preferences for health care. *BMJ* 2000;**320**:1530–3.

Ryan M, Scott DA, Reeves C, Bate A, van Teijlingen ER, Russell EM, *et al*. Eliciting public preferences for healthcare: a systematic review of techniques. *Health Technol Assess (Rockv)* 2001;**5**:1–186.

Salonen R, Ashford EA, Gibbs M, Hassani H. Patient preference for oral sumatriptan 25 mg, 50 mg, or 100 mg in the acute treatment of migraine: a double-blind, randomized, crossover study. Sumatriptan Tablets S2CM11 Study Group. *Int J Clin Pract* 1999;**105** Suppl:16–24.

Sarasin FP. Decision analysis and the implementation of evidence-based medicine. *QJM* 1999;**92**:669–71.

Schwenk W, Neudecker J, Mall J, Bohm B, Muller JM. Prospective randomized blinded trial of pulmonary function, pain, and cosmetic results after laparoscopic vs. microlaparoscopic cholecystectomy. *Surgic Endosc* 2000;**14**:345–8.

Sculpher M, Gafni A. Recognizing diversity in public preferences: the use of preference sub-groups in cost-effectiveness analysis. *Health Econ* 2001;**10**:317–24.

Sorum PC. Measuring patient preferences by willingness to pay to avoid: the case of acute otitis media. *Medical Decis Making* 1999;**19**:27–37.

Swan JS, Langlotz CP. Patient preference for magnetic resonance versus conventional angiography. Assessment methods and implications for cost-effectiveness analysis: an overview. *Invest Radiol* 1998;**33**:553–9.

Thomas R, Daly M, Perryman B, Stockton D. Forewarned is forearmed – benefits of preparatory information on video cassette for patients receiving chemotherapy or radiotherapy – a randomised controlled trial. *Eur J Cancer* 2000; **36**:1536–43.

Thornett AM. Antidepressants and counselling for major depression in primary care. Measuring preference in primary care studies could be improved. *BMJ* 2001;**323**:282.

Tsuchiya A. Age-related preferences and age weighting health benefits. *Soc Sci Med* 1999; **48**:267–76.

Twycross R. The effect of treatment choices on the total cost of palliative care. *Eur J Palliative Care* 1999;**6**:94–7.

Unden AL, Elofsson S. Health from the patient's point of view. How does it relate to the physician's judgement? *Fam Pract* 2001;**18**:174–80.

van den Borne HW. The patient from receiver of information to informed decision-maker. *Patient Educ Couns* 1998;**34**:89–102.

van der, PM, Cairns J. Establishing patient preferences for blood transfusion support: an application of conjoint analysis. *J Health Serv Res Policy* 1998;**3**:70–6.

Xu G, Veloski JJ. Physician parents' influence over their children's choices of careers in generalist specialties. *Acad Med* 1998;**73**:913.

Youngblut JM. Preference vs. necessity. *J Soc Pediatr Nurs* 1998;**3**:3.

Zesk E. Informed choices. *Healthplan* 1996; **37**:15–18.

Publications retrieved for detailed examination (*N* = 155)

The following publications were retrieved for a more detailed examination.

Abdullah TI, Iddon J, Barr L, Baildam AD, Bundred NJ. Prospective randomized controlled trial of preservation of the intercostobrachial nerve during axillary node clearance for breast cancer [see comments]. *Br J Surg* 1998;**85**:1443–5.

Abrams K. Consumer choice as a predictor of job satisfaction and supervisor ratings for people with disabilities. *J Vocational Rehabil* 1997;**9**:205–15.

Ahmedzai S. Transdermal fentanyl versus sustained-release oral morphine in cancer pain: preference, efficacy, and quality of life. *J Pain Symptom Manage* 1997;**13**:254–61.

Alarcon A, Capafons A, Bayot A, Cardena E. Preference between two methods of active-alert hypnosis: not all techniques are created equal. *Am J Clin Hypn* 1999;**41**:269–76.

Albert SM, Murphy PL, Del Bene ML, Rowland LP. A prospective study of preferences and actual treatment choices in ALS [see comments]. *Neurology* 1999;**53**:278–83.

Albertsen PC, Nease RF Jr, Potosky AL. Assessment of patient preferences among men with prostate cancer [see comments]. *J Urol* 1998; **159**:158–63.

Allan L, Tooke L. Giving medicine a fair trial. Patients' preferences should be assessed. *BMJ* 2000;**321**:1529–30.

Allan L, Hays H, Jensen NH, de Waroux BL, Bolt M, Donald R, *et al*. Randomised crossover trial of transdermal fentanyl and sustained release oral morphine for treating chronic non-cancer pain [see comments]. *BMJ* 2001;**322**:1154–8.

Amari-Vaught E, Vaught W. Case study. Don't I count? *Hastings Center Rep* 1997;**27**:23–4.

Antman K, Amato D, Wood W, Corson J, Suit H, Proppe K, *et al*. Selection bias in clinical trials. *J Clin Oncol* 1985;**3**:1142–7.

Ashok PW, Kidd A, Flett GMM, Fitzmaurice A, Graham W, Templeton A. A randomised comparison of medical abortion and surgical vacuum aspiration at 10–13 weeks gestation. *Hum Reprod* 2002;**17**:92–8.

Awad MA, Shapiro SH, Lund JP, Feine JS. Determinants of patients treatment preferences in a clinical trial. *Commun Dent Oral Epidemiol* 2000; **28**:119–25.

Bain C, Cooper KG, Parkin ED. A partially randomised patient preference trial of microwave endometrial ablation using local anaesthesia and intravenous sedation or general anaesthesia: a pilot study. *Gynaecol Endosc* 2001;**10**:223–8.

Bakker A, Spinhoven P, Balkom AJLM, Vleugel L, van Dyck R. Cognitive therapy by allocation versus cognitive therapy by preference in the treatment of panic disorder. *Psychother Psychosom* 2000; **69**:240–3.

Bakker A, van Dyck R, Spinhoven P, van Balkom AJLM. Paroxetine, clomipramine, and cognitive therapy in the treatment of panic disorder. *J Clin Psychiatry* 1999;**60**:831–8.

Barrowclough C, Tarrier N, Lewis S, Sellwood W, Mainwaring J, Quinn J, *et al*. Randomised controlled trial of a needs based psychosocial intervention service for carers of people with schizophrenia. *Br J Psychiatry* 1999;**174**:505–11.

Bauchner MD, Adams W, Barnett E, Klein J. Therapy for acute otitis media. *Arch Paediatr Adolesc Med* 1996;**150**:396–9.

Bedi N, Chilvers C, Churchill R, Dewey M, Duggan C, Fielding K, *et al*. Assessing effectiveness of treatment of depression in primary care: a partially randomised preference trial. *Br J Psychiatry* 2000;**177**:312–18.

Blake P, Berry SC, Readman A, Ratcliffe M, Godley M. A comparison of epanolol and nifepidine in stable angina patients: results of a multicentre trial. *Cardiology* 1991;**79**:249–55.

Bleichrodt H, Gafni A. Time preference, the discounted utility model and health [Review, 53 refs]. *J Health Econ* 1996;**15**:49–66.

Boezaart AP, Berry RA, Laubscher JJ, Nell ML. Evaluation of anxiolysis and pain associated with combined peri and retrobular eye block for cataract surgery. *J Clin Anaesth* 1998;**10**:204–10.

Bourassa MG. Clinical trials of coronary revascularization: coronary angioplasty vs. coronary bypass grafting [Review, 12 refs]. *Curr Opin Cardiol* 2000;**15**:281–6.

Bourin M, Malinge M. A new design of trial for hypnotics comparison: a double blind crossover trial with patients preference assessment and continuation of the preferred treatment. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 1996; **20**:373–85.

Brasel KJ, Weigely JA. Blunt thoracic aortic trauma: a cost–utility approach for injury detection. *Arch Surg* 1996;**131**:619–26.

Brazier J, Deverill M. A checklist for judging preference-based measures of health related quality of life: learning from psychometrics [Review, 51 refs]. *Health Econ* 1999;**8**:41–51.

Brenneman FD, Wright JG, Kennedy ED, McLeod RS. Outcomes research in surgery [Review, 20 refs]. *World J Surg* 1999;**23**:1220–3.

Casarett D, Karlawish J, Sankar P, Hirschman KB, Asch DA. Obtaining informed consent for clinical pain research: patients concerns and information needs. *Pain* 2001;**92**:71–9.

Chen CI, Skingley P, Meyer RM. A comparison of elderly patients with aggressive histology lymphoma who were entered or not entered on to a randomised Phase II trial. *Leuk Lymphoma* 2000;**38**:327–34.

Cherkin DC, Deyo RA, Battie M, Street J, Barlow W. A comparison of physical therapy, chiropractic manipulation, and provision of an educational booklet for the treatment of patients with low back pain [see comments]. *N Engl J Med* 1998;**339**:1021–9.

Cherkin DC, Eisenberg D, Sherman KJ, Barlow W, Kaptchuk TJ, Street J, *et al*. Randomised trial comparing traditional Chinese medical acupuncture, therapeutic massage and self care education for chronic low back pain. *Archives of Internal Medicine* 2001;**161**:1081–8.

Chilvers C, Dewey M, Fielding K, Gretton V, Miller P, Palmer B, *et al*. Antidepressant drugs and generic counselling for treatment of major depression in primary care: randomised trial with patient preference arms. *BMJ* 2001;**322**:1–5 (Follow Up paper for Bedi study).

Cleland JGF, Henderson E, McLenachan J, Findlay IN, Dargie HJ. Effects of captoril an angiotensin-converting enzyme inhibitor, in patients with angina pectoris and heart failure; 1991.

Cooper KG, Grant AM, Garratt AM. The impact of using a partially randomised patient preference design when evaluating alternative managements for heavy menstrual bleeding. *Br J Obstet Gynaecol* 1997;**104**:1367–73.

Cooper KG, Parkin DE, Garratt AM, Grant AM. A randomised comparison of medical and hysteroscopic management in women consulting a gynaecologist for treatment of heavy menstrual loss. *Br J Obstet Gynaecol* 1997;**104**:1360–6.

Coyle MG, Ferguson A, Lagasse L, Oh W, Lester B. Diluted tincture of opium (DTO) and phenobarbital versus DTO alone for neonatal opiate withdrawal in term infants. *J Pediatr* 2002; **140**:561–4.

Creinin MD. Randomized comparison of efficacy, acceptability and cost of medical versus surgical abortion. *Contraception* 2000;**62**:117–24.

Creinin MD. Oral methotrexate and vaginal misoprostol for early abortion. *Contraception* 1996; **54**:15–18.

Crowley B, Hamill JJ, Lyndon S, McKellican JF, Williams P, Miller AJ. Controlled release indomethacin and sustained release diclofenac sodium in the treatment of rheumatoid arthritis: a comparative controlled trial. *Curr Med Res Opin* 1990;**12**:143–51.

Dannaeus A, Foucard T, Johansson GO. The effect of orally administered sodium cromoglycate on symptoms of food allergy. *Clinic Allergy* 1977; **7**:109–15.

Devereaux PJ, Anderson DR, Gardner MJ, Putnam W, Flowerdew GJ, Brownell BF, *et al*. Differences between perspectives of physicians and patients on anticoagulation in patients with atrial fibrillation: observation study. *BMJ* 2001;**323**:1–7.

Detre KM, Guo P, Holubkov R, Califf RM, Sopko G, Bach R, *et al*. Coronary revascularisation in diabetic patients: a comparison of the randomised and observational components of the bypass angioplasty revascularisation (BARI). *Circulation* 1999;**99**:633–40.

Di Blasi Z, Harkness E, Ernst E, Georgiou A, Kleinjen J. Influence of context effects on health outcomes: a systematic review. *Lancet* 2001; **357**:757–62.

Dolan P. Whose preferences count? *Med Decis Making* 1999;**19**:482–6.

Evans DJ, Matthews S, Pitts NB, Longbottom C, Nugent ZJ. A clinical evaluation of an erbium: YAG laser for dental cavity preparation. *Br Dent J* 2000;**188**:677–9.

Exner DV, Reiffel JA, Epstein AE, Ledingham R, Reiter MJ, Yao Q, *et al*. Beta blocker use and survival in patients with ventricular fibrillation or symptomatic ventricular tachycardia: the antiarryhthmics vs implantable defibrillators (AVID) trial. *J Am Coll Cardiol* 1999;**34**:325–33.

Franklin ME, Abramowitz JS, Kozak MJ, Levitt JT, Foa EB. Effectiveness of exposure and ritual prevention for obsessive–compulsive disorder. Randomised compared with nonrandomised samples. *J Consult Clin Psychol* 2000;**68**:594–602.

Furie RA, Cash JM, Cronin ME, Katz RS, Weisman MH, Aranow C, *et al*. Treatment of systemic lupus erythematosus with LJP 394. *J Rheumatol* 2001;**28**:257–65.

Furnham A, Yardley L, Fahmy S, Jamie A. Health beliefs and preferences for medical treatment: a comparison between medical and social science students. *Complement Ther Med* 1999;**7**:101–9.

Gale EA. A randomized, controlled trial comparing insulin lispro with human soluble insulin in patients with Type 1 diabetes on intensified insulin therapy. The UK Trial Group. *Diabet Med* 2000;17:209–14.

Gale EM. A randomised controlled trial comparing insulin lispro with human soluble insulin in patients with type 1 diabetes on intensified insulin therapy. *Diabet Med* 2000;**17**:209–14.

Glick HA, Polsky D, Willke RJ, Schulman KA. A comparison of preference assessment instruments used in a clinical trial: responses to the visual analog scale from the EuroQol EQ-5D and the Health Utilities Index [Published erratum appears in *Med Decis Making* 1999;**19**:511]. *Med Decis Making* 1999;**19**:265–75.

Gordon RM. Effects of volunteering and responsibility on the perceived values and effectiveness of a clinical treatment. *J Consult Clin Psychol* 1976;**44**:799–801.

Gossop M, Johns A, Green L. Opiate withdrawal: inpatient versus outpatient programmes and preferred versus random assignment to treatment. *BMJ* 1986;**293**:103–4.

Hall JA, Dornan MC. Patient sociodemographic characteristics as predictors of satisfaction with medical care: a meta analysis. *Soc. Sci. Med.* 1990; **30**:811–18.

Hanita M. Self report measures of patient utility: should we trust them? *J Clin Epidemiol* 2000; **53**:469–76.

Hansen PK, Smith SF, Nim J, Neldman S, Osler M. Maternal attitudes to fetal monitoring. *Eur J Obstetr Gynaecol Reprod Biol* 1985;**20**:43–51.

Hardy GE, Barkham M, Shapiro DS, Reynolds S, Rees A. Credibility and outcome of cognitive behavioural and psychodynamic–interpersonal psychotherapy. *Br J Clin Psychol* 1995;**34**:555–69.

Hedges A, Rose J, Leighton M, Turner P. A double blind comparison of meptazinol with placebo in postoperative pain. *J Clinic Pharmacol* 1977; **17**:125–7.

Hedges A, Wadsworth J, Turner P. A double blind comparison of nefopam and placebo in post-operative pain. *Curr Med Res Opin* 1978;**5**:614–17.

Helsing M, Bergman B, Thaning L, Hero U. Quality of life and survival in patients with advanced non-small cell lung cancer receiving supportive care plus chemotherapy with carboplatin and etoposide or supportive care only. A multicentre randomised phase III trial. *Eur J Cancer* 1998;**34**:1036–44.

Hemmings A. Counselling in primary care: a randomised controlled trial. *Patient Educ Couns* 1997;**32**:219–30.

Henshaw RC, Naji SA, Russell IT, Templeton AA. Comparison of medical abortion with surgical vacuum aspiration: women's preferences and acceptability of treatment. *BMJ* 1993;**307**:714–17.

Henshaw RC, Naji SA, Russell IT, Templeton AA. A comparison of medical abortion (using miferpristone and gemeprost) with surgical vacuum aspiration: efficacy and early medical sequelae. *Hum Reprod* 1994;**9**:2167–72.

Howie FL, Henshaw RC, Naji SA, Russell IT, Templeton AA. Medical abortion or vacuum aspiration? Two year follow up of a patient preference trial. *Br J Obstet Gynaecol* 1997; **104**:829–833.

Jones R, Pearson J, McGregor S, Cawsey AJ, Barrett A, Craig N, *et al*. Randomised trial of personalised computer based information for cancer patients [see comments]. *BMJ* 1999; **319**:1241–7.

Joy CB, Adams CE, Rice K. Crisis intervention for severe mental illness. *The Cochrane Library*, Issue 2. Oxford: Update Software; 2003.

Kalauokalani D, Cherkin DC, Sherman KJ, Koepsell TD, Deyo RA. Lessons from a trial of acupuncture and massage for low back pain: patient expectations and treatment effects. *Spine* 2001;**26**:1418–24.

Kendrick D, Fielding K, Bentley E, Miller P, Kerslake R, Pringle M. The role of radiography in primary care patients with low back pain of at least 6 weeks duration: a randomised (unblinded) controlled trial. *Health Technol Assess* 2001;**5**(30).

Kerr RO, Tester W. A patient preference study comparing two extended release morphine sulfate

formulations (once daily Kadian versus twice-daily MS Contin) for cancer pain. *Clin Drug Invest* 2000; **19**:25–32.

Kerry S. Dundas D, Hilton S, Rink E, Patel S, Lord J. Routine referral for radiography of patients presenting with low back pain: is patients' outcome influenced by GP's referral for plain radiography? *Health Technol Assess* 2000;**4**(20).

Kesson CM, Gray JMB, Lawson DH. Clobazam – a new hypnotic? *Br J Clin Pharmacol* 1978;**6**:243–6.

Kim SG, Hallstrom A, Love JC, Rosenberg Y, Powell J, Roth J, *et al*. Comparison of clinical characteristics and frequency of implantable defibrillator use between randomised patients in the antiarrhythmics vs implantable defibrillators (AVID) trial and nonrandomised registry patients. *Am J Cardiol* 1997;**80**:454–7.

King M, Sibbald B, Ward E. Bower P, Lloyd M, Gabbay M, *et al*. Randomised controlled trial of non-directive counselling, cognitive-behaviour therapy and usual general practitioner care in the management of depression as well as mixed anxiety and depression in primary care. *Health Technology Assess* 2000;**4**(19).

King SB, Bernhart HX, Kosinski AS, Weintraub WS, Lembo NJ, Petersen JY, *et al*. Angioplasty or surgery for multivessel coronary artery disease: comparison of eligible registry and randomised patients in the EAST trial and influence of treatment selection on outcomes. *Am J Cardiol* 1997;**79**:1453–9.

King SB, Lembo NJ, Weintraub WS, Kasinski AS, Barnhart HX, Kutner MH. Emory Angioplasty versus Surgery Trial (EAST): design, recruitment, and baseline description of patients. *Am J Cardiol* 1995;**70**:42C–59C.

Kirwan JR, Haskard DO, Higgens CS. The use of sequential analysis to assess patient preference for local skin anaesthesia during knee aspiration. *Br J Rheumatol* 1984;**23**:210–13.

Kocsis JH, Frances A, Kalman TP, Shear MK. The effect of psychobiological research on treatment outcome: a controlled study. *Arch Gen Psychiatry* 1981;**38**:511–15.

Lenert LA, Treadwell JR. Effects on preferences of violations of procedural invariance. *Med Decis Making* 1999;**19**:473–81.

Lenert L, Kaplan RM. Validity and interpretation of preference-based measures of health-related quality of life [see comments]. *Med Care* 2000; **38**:II138–50.

Lenert LA, Soetkino RM. Automated computer interviews to elicit utilities: potential applications in the treatment of deep vein thrombosis. *JAMA* 1997;**4**:49–56.

Marre M, Lievre M, Vasmant D, Gallois Y, Hadjadj S, Reglier JC, *et al*. How patients' preferences for risk information influence treatment choice in a case of high risk and high therapeutic uncertainty: asymptomatic localized prostate cancer. *Med Decis Making* 1999; **19**:394–8.

Mazur DJ, Hickam DH. Patient preferences: survival vs quality of life considerations. *J Gen Intern Med* 1993;**8**:374–7.

Mazur JE. Preferences for and against stimuli paired with food. *J Exp Anal Behav* 1999;**72**:21–32.

McKay JR, Alterman AI, McLellan T, Snider EC, O'Brien CP. Effect of random versus nonrandom assignment in a comparison of inpatient and day hospital rehabilitation for male alcoholics. *J Consult Clin Psychol* 1995;**63**:70–8.

McKay JR, Alterman AI, McLellan T, Boardman CR, Mulvaney FD, O'Brien CP. Random versus nonrandom assignment in the evaluation of treatment for cocaine abusers. *J Consult Clin Psychol* 1998;**66**:697–701.

Mehl AL. Physician autonomy, patient choice, and immunization performance measures. *Effect Clin Pract* 1999;**2**:289–93.

Mehlum D, Grasel G, Fankhauser C. Prospective crossover evaluation of four methods of clinical management of tinnitus. *Otolaryngol Head Neck Surg* 1984;**92**:448–53.

Melchart D, Steger HG, Linde K, Makarian K, Hatahet Z, Brenke R, Saller R. Integrating patient preferences in clinical trials: a pilot study of acupuncture vs midazolam for gastroscopy. *J Altern Complement Med* 2002;**8**:265–74.

Merz WA. Placebo response in panic disorder. A review. *Eur Psychiatry* 1994;**9**:123–7.

Michel MC, Goepel M. Treatment satisfaction of patients with lower urinary tract symptoms: randomised controlled trials vs. real life practice [Review, 34 refs]. *Eur Urol* 2000;**38** (Suppl 1):40–7.

Moffett JK, Torgerson D, Bell-Syer S, Jackson D, Llewlyn-Philips H, Farrin A, *et al*. Randomised controlled trial of exercise for low back pain: clinical outcomes, costs and preferences. *BMJ* 1999;**319**:279–83.

Mulrow CD. Helping an obese patient make informed choices [Review, 16 refs]. *BMJ* 1998; **317**:266–7.

Ngai SW, Chan YK, Lam SW, Lao TT. Labour characteristics and uterine activity: misoprostol compared with oxytocin in women at term with prelabour rupture of the membranes. *Br J Obstet Gynaecol* 2000;**107**:222–7.

Nikolaides K, Brizot ML, Patel F, Snijders R. Comparison of chorionic villus sampling and amniocentesis for fetal karyotyping at 10–13 weeks gestation. *Lancet* 1994;**344**:435–9.

Nieuwkerk PT, Hajenius PJ, van der Veen F, Ankum WM, Wijker W, Bossuyt PM. Systemic methotrexate therapy versus laparoscopic salpingostomy in tubal pregnancy. Part II. Patient preferences for systemic methotrexate. *Fertil Steril* 1998;**70**:518–22.

Noel PH, Larme AC, Meyer J, Marsh G, Correa A, Pugh JA. Patient choice in diabetes education curriculum. *Diabetes Care* 1998;**21**:896–901.

O'Connor A, Pennie RA, Dales RE. Framing effects on expectations decisions and side effects experienced: the case of influenza immunisation. *J Clin Epidemiol* 1996;**49**:1271–6.

O'Connor AM, Stacey D, Rovner D, Holmes-Rovner M, Tetroe J, Llewellyn-Thomas H, *et al*. Decision aids for people facing health treatment or screening decisions. *The Cochrane Library*, Issue 3. Oxford: Update Software; 2001.

O'Conor RM, Blomquist GC. Measurement of consumer–patient preferences using a hybrid contingent valuation method. *J Health Econ* 1997; **16**:667–83.

Orme MLE. Phenylbutazone: plasma concentrations and effectiveness in patients with rheumatoid arthritis. *J Int Med Res* 1977; **5** (Suppl. 2):40–7.

Panikbutra K, Lee CT, Babmberg P. Single dose oral norfloxacin or intramuscular spectinomycin to treat gonorrhoea (PPNG and non-PPNG infections): analysis of efficacy and patient preference. *Genitourinary Med* 1988;**64**:235–40.

Paradise JL, Bluestone CD, Bachman RZ, Colborn K, Bernard BS, Taylor FH, *et al*. Efficacy of tonsillectomy for recurrent throat infection in severely affected children. *N Engl J Med* 1984; **310**:674–83.

Paradise JL, Bluestone CD, Rogers KD, Taylor FH, Colborn K, Bachman RZ, *et al*. Efficacy of adenoidectomy for recurrent otitis media in children previously treated with tympanostomy-tube placement. Results of parallel randomised and nonrandomised trials. *JAMA* 1990; **263**:2066–73.

Peacock I, Tattersall RB. The difficult choice of treatment for poorly controlled maturity onset diabetes: tablets or insulin. *BMJ* 1984; **288**:1956–60.

Peck BM, Asch DA, Goold SD, Roter DL, Ubel PA, McIntyre LM, *et al*. Measuring patient expectations: does the instrument affect satisfaction or expectations? *Med Care* 2001; **39**:100–8.

Percy ME, Llewellyn-Thomas H. Assessing preferences about the DNR order: does it depend how you ask? *Med Decis Making* 1995;**15**:209–16.

Pieters WR, Stallhaert RALM, Prins J, Greefhorst APM, van Uffelen R, Schreurs AJM, *et al*. A study on the clinical equivalence and patient preference of fluticasone propionate 250 μg twice daily via the Diskus™ Accuhaler™ Inhaler or the Diskhaler™ Inhaler in adult asthmatic patients. *J Asthma* 1998;**35**:337–45.

Plotnick LH, Duchame FM. Should inhaled anticholinergics be added to B2 antagonists for treating acute childhood and adolescent asthma? A systematic review. *BMJ* 1998;**317**:971–7.

Poyner TF, Menday AP, Williams ZV. Patient attitudes to topical antipsoriatic treatment with calcipotriol and dithranol. *J Eur Acad Dermatol Venereol* 2000;**14**:153–8.

Protheroe J, Fahey T, Montgomery AA, Peters TJ. The impact of patients' preferences on the treatment of atrial fibrillation: observational study of patient based decision analysis [see comments]. *BMJ* 2000;320:1380–4.

Ramin SM, Gambling DR, Lucas MJ, Sharma SK, Sidawi JE, Leveno KJ. Randomised trial of epidural versus intravenous analgesia during labour. *Obstet Gynaecol* 1995;**86**:783–9.

Reddihough DS, King J, Coleman G, Catanese T. Efficacy of programmes based on conductive education for young children with cerebral palsy. *Dev Med Child Neurol* 1998;**40**:763–70.

Rees CE, Bath PA. The information needs and source preferences of women with breast cancer and their family members: a review of the literature published between 1988 and 1998. *J Adv Nurs* 2000;**31**:833–41.

Renjilian DA, Nezu AM, Shermer RL, Perri MG, McKelvey WF, Anton SD. Individual versus group therapy for obesity: effects of matching participants to their treatment preferences. *J Consult Clin Psychol* 2001;**69**:717–21.

Rex DK, Imperiale TF, Portish V. Patients willing to try colonoscopy without sedation: associated clinical factors and results of a randomized controlled trial. *Gastrointest Endosc* 1999;**49**:554–9.

Ricard N, Kind P, Christian S, Jensen M, Stewart J. Link between patient preferences and treatment outcomes in seasonal allergic rhinitis: an empiric investigation. *Clin Ther* 1999;**21**:268–77.

Rickels K. Gingrich RL, Morris RJ, Rosenfeld H, Perloff MM, Clark EL, *et al*. Triazolam in insomniac family practice patients. *Clin Pharmacol Ther* 1975;**18**:3115–24.

Rischin D, White MA, Matthews JP, Toner GC, Watty K, Sulkowski AJ, *et al*. A randomised crossover trial of chemotherapy in the home: patient preferences and cost analysis. *Med J Aust* 2000;**173**:125–7.

Rokke PD, Tomhave JA, Jocic Z. The role of client choice and target selection in self-management therapy for depression in older adults. *Psychol Aging* 1999;**14**:155–69.

Rosen AS, Nystedt L, Bygdeman M, Lundstrom V. Acceptability of a nonsurgical method to terminate very early pregnancy in comparison to vacuum aspiration. *Contraception* 1979;**19**:108–17.

Ross S, Grant A, Counsell C, Gillespie W, Russell I, Prescott R. Barriers to participation in randomised controlled trials: a systematic review [Review, 86 refs]. *J Clin Epidemiol* 1999;**52**:1143–56.

Rowland N, Godfrey C, Bower P, Mellor-Clark J, Heywood P, Hardy R. Counselling in primary care: a systematic review of the research evidence. *Br J Guidance Couns* 2000;**28**:215–31.

Rovers MM, Straatman H, Ingels K, van der Wilt GJ, van den Broek P, Zielhuis GA. Generalisability of trial results based on randomised versus non-randomised allocation of OME infants to ventilation tubes or watchful waiting. *J Clin Epidemiol* 2001;**54**:789–94.

Rutschmann OT, Vernazza PL, Bucher HC, Opravil M, Ledergerber B, Telenti A, *et al*. Long-term hydroxyurea in combination with didanosine and stavudine for the treatment of HIV-1 infection. Swiss HIV Cohort Study. *AIDS* 2000; **14**:2145–51.

Sanchez-Menegay C, Stalder H. Do physicians take into account patients expectations? *J Gen Intern Med* 1994;**9**:404–6.

Schulberg HC, Katon W, Simon GE, Rush J. Treating major depression in primary care practice. *Arch Gen Psychiatry* 1998;**55**:1121–7.

Schulz R. Effects of control and predictability on the physical and psychological well being of the institutionalised aged. *J Pers Soc Psychol* 1976; **33**:563–73.

Seligman MEP. The effectiveness of psychotherapy. *Am Psychol* 1995;**50**:965–74.

Senore C, Battista RN, Ponti A, Segnan N, Shapiro SH, Rosso S, *et al*. Comparing participants and non participants in a smoking cessation trial: selection factors associated with general practitioner recruitment activity. *J Clin Epidemiol* 1999;**52**:83–9.

Shapiro DA, Barkham M, Rees A, Hardy GE, Reynolds S, Startup M. Effects of treatment duration and severity of depression on the effectiveness of cognitive–behavioural and psychodynamic–interpersonal psychotherapy. *J Consult Clin Psychol* 1994;**62**:522–34.

Shepherd R. Social determinants of food choice [Review, 41 refs]. *Proc Nutr Soc* 1999;**58**:807–12.

Simpson S, Corney R, Fitzgerald P, Beecham J. A randomised trial to evaluate the effectiveness and cost effectiveness of counselling patients with chronic depression. *Health Technol Assess* 2000;**4**(36).

Sivan Y, Arce P, Eigen H, Nickerson BG, Newth CJL. A double blind randomised study of sodium cromoglycate versus placebo in patients with cystic fibrosis and bronchial hyperreactivity. *J Allergy Clin Immunol* 1990;**85**:649–54.

Slavin MB, Benrubi GI, Parker R, Griffin CR, Magee MJ. Single dose fluconazole vs intravaginal terconazole in treatment of candida vaginitis: comparison and pilot study. *J Fl Med Assoc* 1992; **79**:693–6.

Smeeth L. Commentary: patients, preferences, and evidence. *BMJ* 2000;**320**:1384.

Smeeth L. Patients, preferences, and evidence. *West J Med* 2001;**174**:316.

Smith RB, Moodie J. Comparative efficacy and tolerability of two antibacterial/anti-inflammatory formulations ('Otomize' spray and 'Otosporin' drops) in the treatment of otitis externa in general practice. *Curr Med Res Opin* 1990;**11**:661–7.

Spencer JW, Cox DN. A comparison of chorionic villi sampling and amniocentesis: acceptability of procedure and maternal attachment to pregnancy. *Obstet Gynecol* 1988;**72**:714–18.

Sproat JE, Dalcin A, Weitauer N, Roberts RS. Hypertrophic sternal scars: silicone gel sheet versus kenalog injection treatment. *Plast Reconstr Surg* 1991;**90**:988–92.

Stalmeier PFM, Bezembinder TGG, Unic AJ. Proportional heuristics in time tradeoff and conjoint measurement. *Med Decis Making* 1996; **16**:36–44.

Strohmaier K, Snyder E, Adamsons I. A multicenter study comparing dorzolamide and pilocarpine as adjunctive therapy to timolol: patient preference and impact on daily life. *J Am Optometr Assoc* 1998;**69**:441–51.

Strohmaier WL, Schubert G, Rosenkranz T, Weigl A. Comparison of extracorporeal shock wave lithotripsy and ureteroscopy in the treatment of ureteral calculi: a prospective study. *Eur Urol* 1999;36:376–9.

Stuart-Harris R, Simes RJ, Coates AS, Raghaven D, Devine R, Tattersall MHN. Patient treatment preference in advanced breast cancer: a randomised crossover study of doxorubicin and mitozantrone. *Eur J Cancer Clin Oncol* 1987; **23**:557–61.

Tallon D, Chard J, Dieppe P. Relation between agendas of the research community and the research consumer. *Lancet* 2000;**355**:2037–40.

Thijs-Boer FM, Thijs JT, van de Wiel HB. Conventional or adhesive external breast prosthesis? A prospective study of the patients' preference after mastectomy. *Cancer Nurs* 2001; **24**:227–30.

Torok PG, Dunn JR. Self-collection of antepartum anogenital group B streptococcus cultures. *J Am Board Fam Pract* 2000;**13**:107–10.

Ungerleider JT, Sarna G, Fairbanks LA, Goodnight J, Andrysiak T, Jamison K. THC or compazine for the cancer chemotherapy patient – the UCLA study. Part II: patient drug preference. *Am J Clin Oncol* 1985;**8**:142–7.

Van den Berg L, Lobatto RM, Zuurmond WWA, de Lange JJ, Wagemans MFM, Bezemert PD. Patients refusal to participate in clinical research. *Eur J Anaesthesiol* 1997;**14**:287–9.

Van der Reijden WA, van der Kwaak H, Vissink A, Veerman ECI, Amerongen AVN. Treatment of xerostomia with polymer-based saliva substitutes in patients with sjorgens syndrome. *Arthritis Rheum* 1996;**39**:57–63.

Van Dyck R, Spinhoven P. Does preference for treatment matter? A study of exposure *in vivo* with or without hypnosis in the treatment of panic disorder with agoraphobia. *Behav Modif* 1997; **21**:172–86.

Viberti GC, Passa P. Determinants of elevated urinary albumin in the 4397 Type 2 diabetic subjects for the DIABHYCAR study in Western Europe and North Africa. *Diabetes Care* 2000; **23** (Suppl 2):B40–8.

Wang RIH, Stockdale SL. A subjective and objective method assessing the efficacy of hypnotic medications in insomniacs. *J Clin Pharmacol* 1977;**45**:728–33.

Williams AC, Nicholas MK, Richardson PH, Pither CE, Fernandes J. Generalising from a controlled trial: the effects of patient preference versus randomisation on the outcome of inpatient versus outpatient chronic pain management. *Pain* 1999;**83**:57–65.

Wright D, Paterson C, Scott N, Hair A, O'Dwyer PJ. Five year follow up of patients undergoing laparoscopic or open groin hernia repair. *Ann Surg* 2002;**235**:333–7.

Young A, Topham C, Moore J, Turner J, Wardle J, Downes M, Evans V, Kay S. A patient preference study comparing raltitrexed ('Tomudex') and bolus

or infusional 5-fluorouracil regimens in advanced colorectal cancer: influence of side-effects and administration attributes. *Eur J Cancer Care (Engl Lang Ed)* 1999;**8**:154–61.

Zimmerer-Branum S, Nelson D. Occupationally embedded exercise versus rote exercise: a choice between occupational forms by elderly nursing home patients. *Am J Occup Ther* 1995;**49**:397–402.

Papers excluded after evaluation of full text (*N* = 120)

After review of the full text of these publications, they were excluded because they were:

1. Review or discussion papers.
2. Descriptive studies or surveys with no experimental element(s).
3. The studies did not have a preference cohort or did not have any assessment and analysis of preference within a standardised randomised trial.

Abdullah TI, Iddon J, Barr L, Baildam AD, Bundred NJ. Prospective randomized controlled trial of preservation of the intercostobrachial nerve during axillary node clearance for breast cancer [see comments]. *Br J Surg* 1998;**85**:1443–5.

Abrams K. Consumer choice as a predictor of job satisfaction and supervisor ratings for people with disabilities. *J Vocat Rehabil* 1997;**9**:205–15.

Ahmedzai S. Transdermal fentanyl versus sustained-release oral morphine in cancer pain: preference, efficacy, and quality of life. *J Pain Symptom Manage* 1997;**13**:254–61.

Alarcon A, Capafons A, Bayot A, Cardena E. Preference between two methods of active-alert hypnosis: not all techniques are created equal. *Am J Clin Hypn* 1999;**41**:269–76.

Albert SM, Murphy PL, Del Bene ML, Rowland LP. A prospective study of preferences and actual treatment choices in ALS [see comments]. *Neurology* 1999;**53**:278–83.

Albertsen PC, Nease RF Jr, Potosky AL. Assessment of patient preferences among men with prostate cancer [see comments]. *J Urol* 1998;**159**:158–63.

Allan L, Tooke L. Giving medicine a fair trial. Patients' preferences should be assessed. *BMJ* 2000;**321**:1529–30.

Allan L, Hays H, Jensen NH, de Waroux BL, Bolt M, Donald R, Kalso E. Randomised crossover trial of transdermal fentanyl and sustained release oral morphine for treating chronic non-cancer pain [see comments]. *BMJ* 2001;**322**:1154–8.

Amari-Vaught E, Vaught W. Case study. Don't I count? *Hastings Center Rep* 1997;**27**:23–4.

Antman K, Amato D, Wood W, Corson J, Suit H, Proppe K, *et al*. Selection bias in clinical trials. *J Clin Oncol* 1985;**3**:1142–7.

Awad MA, Shapiro SH, Lund JP, Feine JS. Determinants of patients treatment preferences in a clinical trial. *Commun Dent Oral Epidemiol* 2000;**28**:119–25.

Barrowclough C, Tarrier N, Lewis S, Sellwood W, Mainwaring J, Quinn J, *et al*. Randomised controlled trial of a needs based psychosocial intervention service for carers of people with schizophrenia. *Br J Psychiatry* 1999;**174**:505–11.

Bauchner MD, Adams W, Barnett E, Klein J. Therapy for acute otitis media. *Arch Paediatr Adolesc Med* 1996;**150**:396–9.

Blake P, Berry SC, Readman A, Ratcliffe M, Godley M. A comparison of epanolol and nifepidine in stable angina patients: results of a multicentre trial. *Cardiology* 1991;**79**:249–55.

Bleichrodt H, Gafni A. Time preference, the discounted utility model and health [Review, 53 refs]. *J Health Econ* 1996;**15**:49–66.

Boezaart AP, Berry RA, Laubscher JJ, Nell ML. Evaluation of anxiolysis and pain associated with combined peri and retrobular eye block for cataract surgery. *J Clin Anaesth* 1998;**10**:204–10.

Bourassa MG. Clinical trials of coronary revascularization: coronary angioplasty vs. coronary bypass grafting [Review, 12 refs]. *Curr Opin Cardiol* 2000;**15**:281–6.

Bourin M, Malinge M. A new design of trial for hypnotics comparison: a double blind crossover trial with patients preference assessment and continuation of the preferred treatment. *Prog Neuropsychopharmacol Biol Psychiatry* 1996; **20**:373–85.

Brasel KJ, Weigely JA. Blunt thoracic aortic trauma: a cost–utility approach for injury detection. *Arch Surg* 1996;**131**:619–26.

Brazier J, Deverill M. A checklist for judging preference-based measures of health related quality of life: learning from psychometrics [Review, 51 refs]. *Health Econ* 1999;**8**:41–51.

Brenneman FD, Wright JG, Kennedy ED, McLeod RS. Outcomes research in surgery [Review, 20 refs]. *World J Surg* 1999;**23**:1220–3.

Casarett D, Karlawish J, Sankar P, Hirschman KB, Asch DA. Obtaining informed consent for clinical pain research: patients concerns and information needs. *Pain* 2001;**92**:71–9.

Chen CI, Skingley P, Meyer RM. A comparison of elderly patients with aggressive histology lymphoma who were entered or not entered on to a randomised Phase II trial. *Leuk Lymphoma* 2000; **38**:327–34.

Cherkin DC, Deyo RA, Battie M, Street J, Barlow W. A comparison of physical therapy, chiropractic manipulation, and provision of an educational booklet for the treatment of patients with low back pain [see comments]. *N Engl J Med* 1998;**339**:1021–9.

Cherkin DC, Eisenberg D, Sherman KJ, Barlow W, Kaptchuk TJ, Street J, *et al*. Randomised trial comparing traditional Chinese medical acupuncture, therapeutic massage and self care education for chronic low back pain. *Arch Intern Med* 2001;**161**:1081–8.

Cleland JGF, Henderson E, McLenachan J, Findlay IN, Dargie HJ. Effects of captoril an angiotensin-converting enzyme inhibitor, in patients with angina pectoris and heart failure; *Journal of American College of Cardiologists* 1991;**17**:733–9.

Coyle MG, Ferguson A, Lagasse L, Oh W, Lester B. Diluted tincture of opium (DTO) and phenobarbital versus DTO alone for neonatal opiate withdrawal in term infants. *J Pediatr* 2002; **140**:561–4.

Creinin MD. Randomized comparison of efficacy, acceptability and cost of medical versus surgical abortion. *Contraception* 2000;**62**:117–24.

Creinin MD. Oral methotrexate and vaginal misoprostol for early abortion. *Contraception* 1996; **54**:15–18.

Crowley B, Hamill JJ, Lyndon S, McKellican JF, Williams P, Miller AJ. Controlled release indomethacin and sustained release diclofenac sodium in the treatment of rheumatoid arthritis: a comparative controlled trial. *Curr Med Res Opin* 1990;**12**:143–51.

Dannaeus A, Foucard T, Johansson GO. The effect of orally administered sodium cromoglycate on symptoms of food allergy. *Clin Allergy* 1977; **7**:109–15.

Devereaux PJ, Anderson DR, Gardner MJ, Putnam W, Flowerdew GJ, Brownell BF, *et al*. Differences between perspectives of physicians and patients on anticoagulation in patients with atrial fibrillation: observation study. *BMJ* 2001; **323**:1–7.

Di Blasi Z, Harkness E, Ernst E, Georgiou A, Kleinjen J. Influence of context effects on health outcomes: a systematic review. *Lancet* 2001; **357**:757–62.

Dolan P. Whose preferences count? *Med Decis Making* 1999;**19**:482–6.

Evans DJ, Matthews S, Pitts NB, Longbottom C, Nugent ZJ. A clinical evaluation of an erbium: YAG laser for dental cavity preparation. *Br Dent J* 2000;**188**:677–9.

Exner DV, Reiffel JA, Epstein AE, Ledingham R, Reiter MJ, Yao Q, *et al*. Beta blocker use and survival in patients with ventricular fibrillation or symptomatic ventricular tachycardia: the antiarryhthmics vs implantable defibrillators (AVID) trial. *J Am Coll Cardiol* 1999;**34**:325–33.

Franklin ME, Abramowitz JS, Kozak MJ, Levitt JT, Foa EB. Effectiveness of exposure and ritual prevention for obsessive–compulsive disorder. Randomised compared with nonrandomised samples. *J Consult Clin Psychol* 2000;**68**:594–602.

Furie RA, Cash JM, Cronin ME, Katz RS, Weisman MH, Aranow C, *et al*. Treatment of systemic lupus erythematosus with LJP 394. *J Rheumatol* 2001;**28**:257–65.

Furnham A, Yardley L, Fahmy S, Jamie A. Health beliefs and preferences for medical treatment: a comparison between medical and social science students. *Complement Ther Med* 1999;**7**:101–9.

Gale EA. A randomized, controlled trial comparing insulin lispro with human soluble insulin in patients with Type 1 diabetes on intensified insulin therapy. The UK Trial Group. *Diabet Med* 2000;**17**:209–14.

Gale EM. A randomised controlled trial comparing insulin lispro with human soluble insulin in patients with type 1 diabetes on intensified insulin therapy. *Diabet Med* 2000; **17**:209–14.

Glick HA, Polsky D, Willke RJ, Schulman KA. A comparison of preference assessment instruments used in a clinical trial: responses to the visual analog scale from the EuroQol EQ-5D and the Health Utilities Index [Published erratum appears in *Med Decis Making* 1999;**19**:511]. *Med Decis Making* 1999;**19**:265–75.

Gordon RM. Effects of volunteering and responsibility on the perceived values and effectiveness of a clinical treatment. *J Consult Clin Psychol* 1976:**44**:799–801.

Hall JA, Dornan MC. Patient sociodemographic characteristics as predictors of satisfaction with medical care: a meta analysis. *Soc. Sci. Med.* 1990; **30**:811–18.

Hanita M. Self report measures of patient utility: should we trust them? *J Clin Epidemiol* 2000; **53**:469–76.

Hansen PK, Smith SF, Nim J, Neldman S, Osler M. Maternal attitudes to fetal monitoring. *Eur J Obstet Gynaecol Reprod Biol* 1985;**20**:43–51.

Hedges A, Rose J, Leighton M, Turner P. A double blind comparison of meptazinol with placebo in post operative pain. *J Clin Pharmacol* 1977;**17**:125–7.

Hedges A, Wadsworth J, Turner P. A double blind comparison of nefopam and placebo in post-operative pain. *Curr Med Res Opin* 1978; **5**:614–17.

Hemmings A. Counselling in primary care: a randomised controlled trial. *Patient Educ Couns* 1997;**32**:219–30.

Jones R, Pearson J, McGregor S, Cawsey AJ, Barrett A, Craig N, *et al*. Randomised trial of personalised computer based information for cancer patients [see comments]. *BMJ* 1999; **319**:1241–7.

Joy CB, Adams CE, Rice K. Crisis intervention for severe mental illness. *The Cochrane Library*, Issue 2. Oxford: Update Software; 2003.

Kalauokalani D, Cherkin DC, Sherman KJ, Koepsell TD, Deyo RA. Lessons from a trial of acupuncture and massage for low back pain: patient expectations and treatment effects. *Spine* 2001;**26**:1418–24.

Kerr RO, Tester W. A patient preference study comparing two extended release morphine sulfate formulations (once daily Kadian versus twice-daily MS Contin) for cancer pain. *Clin Drug Invest* 2000;**19**:25–32.

Kesson CM, Gray JMB, Lawson DH. Clobazam – a new hypnotic? *Br J Clin Pharmacol* 1978;**6**:243–6.

Kim SG, Hallstrom A, Love JC, Rosenberg Y, Powell J, Roth J, *et al*. Comparison of clinical characteristics and frequency of implantable defibrillator use between randomised patients in the antiarrhythmics vs implantable defibrillators (AVID) trial and nonrandomised registry patients. *Am J Cardiol* 1997;**80**:454–7.

Kirwan JR, Haskard DO, Higgens CS. The use of sequential analysis to assess patient preference for local skin anaesthesia during knee aspiration. *Br J Rheumatol* 1984;**23**:210–13.

Kocsis JH, Frances A, Kalman TP, Shear MK. The effect of psychobiological research on treatment outcome: a controlled study. *Arch Gen Psychiatry* 1981;**38**:511–15.

Lenert LA, Treadwell JR. Effects on preferences of violations of procedural invariance. *Med Decis Making* 1999;**19**:473–81.

Lenert L, Kaplan RM. Validity and interpretation of preference-based measures of health-related quality of life [see comments]. *Med Care* 2000; **38**:II138–50.

Lenert LA, Soetkino RM. Automated computer interviews to elicit utilities: potential applications in the treatment of deep vein thrombosis. *JAMA* 1997;**4**:49–56.

Marre M, Lievre M, Vasmant D, Gallois Y, Hadjadj S, Reglier JC, *et al*. How patients' preferences for risk information influence treatment choice in a case of high risk and high therapeutic uncertainty: asymptomatic localized prostate cancer. *Med Decis Making* 1999;**19**:394–8.

Mazur DJ, Hickam DH. Patient preferences: survival vs quality of life considerations. *J Gen Intern Med* 1993;**8**:374–7.

Mazur JE. Preferences for and against stimuli paired with food. *J Exp Anal Behav* 1999;**72**:21–32.

Mehl AL. Physician autonomy, patient choice, and immunization performance measures. *Effect Clin Pract* 1999;**2**:289–93.

Mehlum D, Grasel G, Fankhauser C. Prospective crossover evaluation of four methods of clinical management of tinnitus. *Otolaryngol Head Neck Surg* 1984;**92**:448–53.

Merz WA. Placebo response in panic disorder. A review. *Eur Psychiatry* 1994;**9**:123–7.

Michel MC, Goepel M. Treatment satisfaction of patients with lower urinary tract symptoms: randomised controlled trials vs. real life practice [Review, 34 refs]. *Eur Urol* 2000;**38** Suppl 1:40–7.

Mulrow CD. Helping an obese patient make informed choices [Review, 16 refs]. *BMJ* 1998; **317**:266–7.

Ngai SW, Chan YK, Lam SW, Lao TT. Labour characteristics and uterine activity: misoprostol compared with oxytocin in women at term with prelabour rupture of the membranes. *Br J Obstet Gynaecol* 2000;**107**:222–7.

Nieuwkerk PT, Hajenius PJ, van der Veen F, Ankum WM, Wijker W, Bossuyt PM. Systemic methotrexate therapy versus laparoscopic salpingostomy in tubal pregnancy. Part II. Patient preferences for systemic methotrexate. *Fertil Steril* 1998;**70**:518–22.

O'Connor A, Pennie RA, Dales RE. Framing effects on expectations decisions and side effects experienced: the case of influenza immunisation. *J Clin Epidemiol* 1996;**49**:1271–6.

O'Connor AM, Stacey D, Rovner D, Holmes-Rovner M, Tetroe J, Llewellyn-Thomas H, *et al*. Decision aids for people facing health treatment or screening decisions. *The Cochrane Library*, Issue 3. Oxford: Update Software; 2001.

O'Conor RM, Blomquist GC. Measurement of consumer–patient preferences using a hybrid contingent valuation method. *J Health Econ* 1997; **16**:667–83.

Orme ML'E. Phenylbutazone: plasma concentrations and effectiveness in patients with rheumatoid arthritis. *J Int Med Res* 1977; **5** (Suppl. 2):40–7.

Panikbutra K, Lee CT, Babmberg P. Single dose oral norfloxacin or intramuscular spectinomycin to treat gonorrhoea (PPNG and non-PPNG infections): analysis of efficacy and patient preference. *Genitourinary Med* 1988;**64**:235–40.

Peacock I, Tattersall RB. The difficult choice of treatment for poorly controlled maturity onset diabetes: tablets or insulin. *BMJ* 1984; **288**:1956–60.

Peck BM, Asch DA, Goold SD, Roter DL, Ubel PA, McIntyre LM, *et al*. Measuring patient expectations: does the instrument affect satisfaction or expectations? *Med Care* 2001;**39**:100–8.

Percy ME, Llewellyn-Thomas H. Assessing preferences about the DNR order: does it depend how you ask? *Med Decis Making* 1995;**15**:209–16.

Pieters WR, Stallhaert RALM, Prins J, Greefhorst APM, van Uffelen R, Schreurs AJM, *et al*. A study on the clinical equivalence and patient preference of fluticasone propionate 250 μg twice daily via the Diskus™ Accuhaler™ Inhaler or the Diskhaler™ Inhaler in adult asthmatic patients. *J Asthma* 1998;**35**:337–45.

Plotnick LH, Duchame FM. Should inhaled anticholinergics be added to B2 antagonists for treating acute childhood and adolescent asthma? A systematic review. *BMJ* 1998;**317**:971–7.

Poyner TF, Menday AP, Williams ZV. Patient attitudes to topical antipsoriatic treatment with calcipotriol and dithranol. *J Eur Acad Dermatol Venereol* 2000;**14**:153–8.

Protheroe J, Fahey T, Montgomery AA, Peters TJ. The impact of patients' preferences on the treatment of atrial fibrillation: observational study of patient based decision analysis [see comments]. *BMJ* 2000;**320**:1380–4.

Ramin SM, Gambling DR, Lucas MJ, Sharma SK, Sidawi JE, Leveno KJ. Randomised trial of epidural versus intravenous analgesia during labour. *Obstet Gynaecol* 1995;**86**:783–9.

Rees CE, Bath PA. The information needs and source preferences of women with breast cancer and their family members: a review of the literature published between 1988 and 1998. *J Adv Nurs* 2000;**31**:833–41.

Rex DK, Imperiale TF, Portish V. Patients willing to try colonoscopy without sedation: associated clinical factors and results of a randomized controlled trial. *Gastrointest Endosc* 1999;**49**:554–9.

Ricard N, Kind P, Christian S, Jensen M, Stewart J. Link between patient preferences and treatment outcomes in seasonal allergic rhinitis: an empiric investigation. *Clin Ther* 1999;**21**:268–77.

Rickels K, Gingrich RL, Morris RJ, Rosenfeld H, Perloff MM, Clark EL, *et al*. Triazolam in insomniac family practice patients. *Clin Pharmacol Ther* 1975;**18**:3115–24.

Rischin D, White MA, Matthews JP, Toner GC, Watty K, Sulkowski AJ, *et al*. A randomised crossover trial of chemotherapy in the home: patient preferences and cost analysis. *Med J Aust* 2000;**173**:125–7.

Rosen AS, Nystedt L, Bygdeman M, Lundstrom V. Acceptability of a nonsurgical method to terminate very early pregnancy in comparison to vacuum aspiration. *Contraception* 1979;**19**:108–17.

Ross S, Grant A, Counsell C, Gillespie W, Russell I, Prescott R. Barriers to participation in randomised controlled trials: a systematic review [Review, 86 refs]. *J Clin Epidemiol* 1999; **52**:1143–56.

Rowland N, Godfrey C, Bower P, Mellor-Clark J, Heywood P, Hardy R. Counselling in primary care: a systematic review of the research evidence. *Br J Guidance Couns* 2000;**28**:215–31.

Rutschmann OT, Vernazza PL, Bucher HC, Opravil M, Ledergerber B, Telenti A, *et al*. Long-term hydroxyurea in combination with didanosine and stavudine for the treatment of HIV-1 infection. Swiss HIV Cohort Study. *AIDS* 2000; **14**:2145–51.

Sanchez-Menegay C, Stalder H. Do physicians take into account patients expectations? *J Gen Intern Med* 1994;**9**:404–6.

Schulberg HC, Katon W, Simon GE, Rush J. Treating major depression in primary care practice. *Arch Gen Psychiatry* 1998;**55**:1121–7.

Schulz R. Effects of control and predictability on the physical and psychological well being of the institutionalised aged. *J Pers Soc Psychol* 1976; **33**:563–73.

Seligman MEP. The effectiveness of psychotherapy. *Am Psychol* 1995;**50**:965–74.

Senore C, Battista RN, Ponti A, Segnan N, Shapiro SH, Rosso S, *et al*. Comparing participants and non participants in a smoking cessation trial: selection factors associated with general practitioner recruitment activity. *J Clin Epidemiol* 1999;**52**:83–9.

Shepherd R. Social determinants of food choice [Review, 41 refs]. *Proc Nutr Soc* 1999;**58**:807–12.

Simpson S, Corney R, Fitzgerald P, Beecham J. A randomised trial to evaluate the effectiveness and cost effectiveness of counselling patients with chronic depression. *Health Technology Assess* 2000; **4**(36).

Sivan Y, Arce P, Eigen H, Nickerson BG, Newth CJL. A double blind randomised study of sodium cromoglycate versus placebo in patients with cystic fibrosis and bronchial hyperreactivity. *J Allergy Clin Immunol* 1990;**85**:649–54.

Slavin MB, Benrubi GI, Parker R, Griffin CR, Magee MJ. Single dose fluconazole vs intravaginal terconazole in treatment of candida vaginitis: comparison and pilot study. *J Fl Med Assoc* 1992; **79**:693–6.

Smeeth L. Commentary: patients, preferences, and evidence. *BMJ* 2000;**320**:1384.

Smeeth L. Patients, preferences, and evidence. *West J Med* 2001;**174**:316.

Smith RB, Moodie J. Comparative efficacy and tolerability of two antibacterial/anti-inflammatory formulations ('Otomize' spray and 'Otosporin' drops) in the treatment of otitis externa in general practice. *Curr Med Res Opin* 1990;**11**:661–7.

Spencer JW, Cox DN. A comparison of chorionic villi sampling and amniocentesis: acceptability of procedure and maternal attachment to pregnancy. *Obstet Gynecol* 1988;**72**:714–18.

Sproat JE, Dalcin A, Weitauer N, Roberts RS. Hypertrophic sternal scars: silicone gel sheet versus kenalog injection treatment. *Plast Reconstr Surg* 1991;**90**:988–92.

Stalmeier PFM, Bezembinder TGG, Unic AJ. Proportional heuristics in time tradeoff and conjoint measurement. *Med Decis Making* 1996; **16**:36–44.

Strohmaier K, Snyder E, Adamsons I. A multicenter study comparing dorzolamide and pilocarpine as adjunctive therapy to timolol: patient preference and impact on daily life. *J Am Optometr Assoc* 1998;**69**:441–51.

Strohmaier WL, Schubert G, Rosenkranz T, Weigl A. Comparison of extracorporeal shock wave lithotripsy and ureteroscopy in the treatment of ureteral calculi: a prospective study. *Eur Urol* 1999; **36**:376–9.

Stuart-Harris R, Simes RJ, Coates AS, Raghaven D, Devine R, Tattersall MHN. Patient treatment preference in advanced breast cancer: a randomised crossover study of doxorubicin and mitozantrone. *Eur J Cancer Clin Oncol* 1987; **23**:557–61.

Tallon D, Chard J, Dieppe P. Relation between agendas of the research community and the research consumer. *Lancet* 2000;**355**:2037–40.

Thijs-Boer FM, Thijs JT, van de Wiel HB. Conventional or adhesive external breast prosthesis? A prospective study of the patients' preference after mastectomy. *Cancer Nurs* 2001; **24**:227–30.

Torok PG, Dunn JR. Self-collection of antepartum anogenital group B streptococcus cultures. *J Am Board Fam Pract* 2000;**13**:107–10.

Ungerleider JT, Sarna G, Fairbanks LA, Goodnight J, Andrysiak T, Jamison K. THC or compazine for the cancer chemotherapy patient – the UCLA study. Part II: patient drug preference. *Am J Clin Oncol* 1985;**8**:142–7.

Van den Berg L, Lobatto RM, Zuurmond WWA, de Lange JJ, Wagemans MFM, Bezemert PD. Patients refusal to participate in clinical research. *Eur J Anaesthesiol* 1997;**14**:287–9.

Van der Reijden WA, van der Kwaak H, Vissink A, Veerman ECI, Amerongen AVN. Treatment of xerostomia with polymer-based saliva substitutes in patients with sjorgens syndrome. *Arthritis Rheum* 1996;**39**:57–63.

Viberti GC, Passa P. Determinants of elevated urinary albumin in the 4397 Type 2 diabetic subjects for the DIABHYCAR study in Western Europe and North Africa. *Diabetes Care* 2000; **23** (Suppl 2):B40–8.

Wang RIH, Stockdale SL. A subjective and objective method assessing the efficacy of hypnotic medications in insomniacs. *J Clin Pharmacol* 1977; **45**:728–33.

Wright D, Paterson C, Scott N, Hair A, O'Dwyer PJ. Five year follow up of patients undergoing laparoscopic or open groin hernia repair. *Ann Surg* 2002;**235**:333–7.

Young A, Topham C, Moore J, Turner J, Wardle J, Downes M, *et al*. A patient preference study comparing raltitrexed ('Tomudex') and bolus or infusional 5-fluorouracil regimens in advanced colorectal cancer: influence of side-effects and administration attributes. *Eur J Cancer Care (Engl Lang Ed)* 1999;**8**:154–61.

Zimmerer-Branum S, Nelson D. Occupationally embedded exercise versus rote exercise: a choice between occupational forms by elderly nursing home patients. *Am J Occup Ther* 1995;**49**:397–402.

Papers retrieved for inclusion (*N* = 37)

Ashok PW, Kidd A, Flett GMM, Fitzmaurice A, Graham W, Templeton A. A randomised comparison of medical abortion and surgical vacuum aspiration at 10–13 weeks gestation. *Hum Reprod* 2002;**17**:92–8.

Bain C, Cooper KG, Parkin ED. A partially randomised patient preference trial of microwave endometrial ablation using local anaesthesia and intravenous sedation or general anaesthesia: a pilot study. *Gynaecol Endosc* 2001;**10**:223–8.

Bakker A, Spinhoven P, Balkom AJLM, Vleugel L, van Dyck, R. Cognitive therapy by allocation versus cognitive therapy by preference in the treatment of panic disorder. Psychother Psychosom 2000;**69**:240–3.

[a] Bakker A, van Dyck R, Spinhoven P, van Balkom AJLM. Paroxetine, clomipramine and cognitive therapy in the treatment of panic disorder. *J Clin Psychiatry* 1999;**60**:831–8.

Bedi N, Chilvers C, Churchill R, Dewey M, Duggan C, Fielding K, *et al*. Assessing effectiveness of treatment of depression in primary care: a partially randomised preference trial. *Br J Psychiatry* 2000;**177**:312–18.

Chilvers C, Dewey M, Fielding K, Gretton V, Miller P, Palmer B, *et al*. Antidepressant drugs and generic counselling for treatment of major depression in primary care: randomised trial with patient preference arms. *BMJ* 2001;**322**:1–5.

Cooper KG, Grant AM, Garratt AM. The impact of using a partially randomised patient preference design when evaluating alternative managements

for heavy menstrual bleeding. *Br J Obstet Gynaecol* 1997;**104**:1367–73.

[a] Cooper KG, Parkin DE, Garratt AM, Grant AM. A randomised comparison of medical and hysteroscopic management in women consulting a gynaecologist for treatment of heavy menstrual loss. *Br J Obstet Gynaecol* 1997;**104**:1360–6.

Detre KM, Guo P, Holubkov R, Califf RM, Sopko G, Bach R, *et al*. Coronary revascularisation in diabetic patients: a comparison of the randomised and observational components of the bypass angioplasty revascularisation (BARI). *Circulation* 1999;**99**:633–40.

Gossop M, Johns A, Green L. Opiate withdrawal: inpatient versus outpatient programmes and preferred versus random assignment to treatment. *BMJ* 1986;**293**:103–4.

Hardy GE, Barkham M, Shapiro DS, Reynolds S, Rees A. Credibility and outcome of cognitive–behavioural and psychodynamic–interpersonal psychotherapy. *Br J Clin Psychol* 1995;**34**:555–69.

[a] Shapiro DA, Barkham M, Rees A, Hardy GE, Reynolds S, Startup M. Effects of treatment duration and severity of depression on the effectiveness of cognitive–behavioural and psychodynamic–interpersonal psychotherapy. *J Consult Clin Psychol* 1994;**62**:522–34.

[a] Shapiro D, Barkham M, Hardy G, Morrison L. The second Sheffield psychotherapy project: rationale, design and preliminary outcome data. *Br J Med Psychol* 1990:**63**:97–108.

Helsing M, Bergman B, Thaning L, Hero U. Quality of life and survival in patients with advanced non-small cell lung cancer receiving supportive care plus chemotherapy with carboplatin and etoposide or supportive care only. A multicentre randomised phase III trial. *Eur J Cancer* 1998;**34**:1036–44.

Henshaw RC, Naji SA, Russell IT, Templeton AA. Comparison of medical abortion with surgical vacuum aspiration: women's preferences and acceptability of treatment. *BMJ* 1993; **307**:714–17.

Henshaw RC, Naji SA, Russell IT, Templeton AA. A comparison of medical abortion (using miferpristone and gemeprost) with surgical vacuum aspiration: efficacy and early medical sequelae. *Hum Reprod* 1994;**9**:2167–72.

[a] Howie FL, Henshaw RC, Naji SA, Russell IT, Templeton AA. Medical abortion or vacuum aspiration? Two year follow up of a patient preference trial. *Br J Obstet Gynaecol* 1997; **104**:829–33.

Kendrick D, Fielding K, Bentley E, Miller P, Kerslake R, Pringle M. The role of radiography in primary care patients with low back pain of at least 6 weeks duration: a randomised (unblinded) controlled trial. *Health Technol Assess* 2001;**5**:(30).

Kerry S. Dundas D, Hilton S, Rink E, Patel S, Lord J. Routine referral for radiography of patients presenting with low back pain: is patients outcome influenced by GP's referral for plain radiography? *Health Technology Assess* 2000;**4**(20).

King M, Sibbald B, Ward E. Bower P. Lloyd M, Gabbay M, Byford S. Randomised controlled trial of non-directive counselling, cognitive–behaviour therapy and usual general practitioner care in the management of depression as well as mixed anxiety and depression in primary care. *Health Technol Assess* 2000;**4**(19).

King SB, Barnhart HX, Kosinski AS. Angioplasty or surgery for multivessel coronary artery disease: comparison of eligible registry and randomised patients in the EAST trial and influence of treatment selection on outcomes. *Am J Cardiol* 1997;**79**:1453–9.

[a] King SB, Lembo NJ, Weintraub WS, Kasinski AS, Barnhart HX, Kutner MH. Emory Angioplasty Versus Surgery Trial (EAST): design, recruitment, and baseline description of patients. *Am J Cardiol* 1995;**70**:42C–59C.

King SB, Lembo NJ, Weintraub WS, Kosinski AS, Barnhart HX, Kutner MH, *et al*. A randomized trial comparing coronary angioplasty with coronary bypass surgery. Emory Angioplasty v/s Surgery Trial (EAST). *N Engl J Med* 1994; **331**:1044–50.

McKay JR, Alterman AI, McLellan T, Snider EC, O'Brien CP. Effect of random versus nonrandom assignment in a comparison of inpatient and day hospital rehabilitation for male alcoholics. *J Consult Clin Psychol* 1995;**63**:70–8.

McKay JR, Alterman AI, McLellan T, Boardman CR, Mulvaney FD, O'Brien CP. Random versus nonrandom assignment in the evaluation of treatment for cocaine abusers. *J Consult Clin Psychol* 1998;**66**:697–701.

Melchart D, Steger HG, Linde K, Makarian K, Hatahet Z, Brenke R, *et al*. (2002) Integrating patient preferences in clinical trials: a pilot study of acupuncture vs midazolam for gastroscopy. *J Altern Complement Med* 1998;**8**:265–74.

Moffett JK, Torgerson D, Bell-Syer S, Jackson D, Llewlyn-Philips H, Farrin A, *et al*. Randomised controlled trial of exercise for low back pain: clinical outcomes, costs and preferences. *BMJ* 1999;**319**:279–83.

Nikolaides K, Brizot ML, Patel F, Snijders R. Comparison of chorionic villus sampling and amniocentesis for fetal karyotyping at 10–13 weeks gestation. *Lancet* 1994;**344**:435–9.

Noel PH, Larme AC, Meyer J. Marsh G, Correa A, Pugh JA. Patient choice in diabetes education curriculum. *Diabetes Care* 1998;**21**:896–901.

Paradise JL, Bluestone CD, Bachman RZ, Colborn K, Bernard BS, Taylor FH, *et al*. Efficacy of tonsillectomy for recurrent throat infection in severely affected children. *N Engl J Med* 1984; **310**:674–83.

Paradise JL, Bluestone CD, Rogers KD, Taylor FH, Colborn K, Bachman RZ, *et al*. Efficacy of adenoidectomy for recurrent otitis media in children previously treated with tympanostomy-tube placement. Results of parallel randomised and nonrandomised Trials. *JAMA* 1990; **263**:2066–73.

Reddihough DS, King J, Coleman G, Catanese T. Efficacy of programmes based on conductive education for young children with cerebral palsy. *Dev Med Child Neurol* 1998;**40**:763–70.

Renjilian DA, Nezu AM, Shermer RL, Perri MG, McKelvey WF, Anton SD. Individual versus group therapy for obesity: effects of matching participants to their treatment preferences. *J Consult Clin Psychol* 2001;**69**:717–21.

Rokke PD, Tomhave JA, Jocic Z. The role of client choice and target selection in self-management therapy for depression in older adults. *Psychol Aging* 1999;**14**:155–69.

Rovers MM, Straatman H, Ingels K, van der Wilt GJ, van den Broek P, Zielhuis GA. Generalisability of trial results based on randomised versus non-randomised allocation of OME infants to ventilation tubes or watchful waiting. *J Clin Epidemiol* 2001;**54**:789–94.

Van Dyck R, Spinhoven P. Does preference for treatment matter? A study of exposure *in vivo* with or without hypnosis in the treatment of panic disorder with agoraphobia. *Behav Modif* 1997; **21**:172–86.

Williams AC, Nicholas MK, Richardson PH, Pither CE, Fernandes J. Generalising from a controlled trial: the effects of patient preference versus randomisation on the outcome of inpatient versus outpatient chronic pain management. *Pain* 1999;**83**:57–65.

[a] These are secondary or follow-up papers used for additional information on particular trials.

Papers retrieved from search of references (*N* = 14)

Brockelhurst P. Partially randomised patient preference trials. *Br J Obstet Gynaecol* 1997; **104**:1332–5.

[b] CASS Principal Investigators and Their Associates. Coronary artery surgery study (CASS): a randomised trial of coronary artery bypass surgery. *J Am Coll Cardiol* 1984;**3**:114–28.

[b] Devine DA, Fernald PS. Outcome effects of receiving a preferred, randomly assigned or nonpreferred therapy. *J Consult Clin Psychol* 1973; **41**:104–7.

Halpern SD. Prospective preference assessment: a method to enhance the ethics and efficiency of randomised controlled trials. *Control Clin Trials* 2002;**23**:274–88.

Holub Z. Hormonal substitution treatment in women operated on account of endometriosis and adenomyosis: prospective follow-up study. *Ceska Gynekol* 2001;**66**:405–8.

Katon W, Von Korff M, Lin E, Walker E, Simon G, Robinson P, *et al*. Methodological issues in randomised trials of liaison psychiatry in primary care. *Psychosom Med* 1994;**56**:97–103.

[c] Juster H, Heimberg RG, Engelberg B. Self selection and sample selection in a treatment study of social phobia. *Behav Res Ther* 1995;**33**:321–4.

Moynihan C, Bliss JM, Davidson J, Burchell L, Horwich A. Evaluation of adjuvant psychological therapies in patients with testicular cancer: randomised controlled trial. *BMJ* 1998; **316**:429–35.

Olschweski M, Scheurlen H. Comprehensive cohort study: an alternative to randomised consent design in a breast preservation trial. *Methods Inf Med* 1985;**24**:131–4.

[b] Olschewski M, Schumacher M, Davis K. Analysis of randomised and nonrandomized patients in clinical trials using the comprehensive cohort follow-up study design. *Control Clin Trials* 1992; **13**:226–39.

Riley S, Morris B, Walker L. Reese P, White L. Comparison of transdermal nitroglycerin systems: Transderm-nitro© and Nitro-dur©. *Clin Ther* 1992; **14**:438–44.

Rodin J, Langer EJ. Long-term effects of a control relevant intervention with the institutionalised aged. *J Pers Soc Psychol* 1977;**35**:897–902.

Rokke PD, Lall R. The role of therapy in enhancing tolerance to acute pain. *Cogn Ther Res* 1992;**16**:53–65.

Rokke PD, Scogin F. Depression treatment preferences in younger and older adults. *J Clin Geropsychol* 1995;**1**:243–57.

[b] Schmoor C, Olschewski M, Schumacher M. Randomized and non-randomized patients in clinical trials: experience with comprehensive cohort studies. *Stat Med* 1996;**15**:263–71.

[b] Schumacher M, Bastert G, Bojar H, Hubner K, Olschewski M, Sauerbrei W, *et al*. Randomised "X" trial evaluating hormonal treatment and the duration of chemotherapy in node positive breast cancer patients. *J Clin Oncol* 1994:**12**:2086–93.

Sikorski J, Wilson J, Clement S, Das S, Smeeton N. A randomised controlled trial comparing two schedules of antenatal visits: the antenatal care project. *BMJ* 1996;**312**:546–53.

Snowdon C, Elbourne D, Garcia J. Zelen randomisation: attitudes of parents participating in a neonatal clinical trial. *Control Clin Trials* 1998; **20**:149–71.

The German Breast Cancer Study Group. Therapy of small breast cancer – four year results of a prospective non-randomized study. *Breast Cancer Res Treat* 1995;**34**:1–13.

[b] These papers were included in the final review ($N = 5$).

[c] Although this study met the inclusion criteria, we were unable to obtain sufficient outcome data to include it in the review.

Papers retrieved from further searches ($N = 2$)

Janevic MR, Janz NK, Dodge JA, Lin X, Pan W, Sinco B, *et al*. The role of choice in health education intervention trials: a review and case study. *Soc Sci Med* 2003;**56**:1581–94.

Riedl S, Peter B, Geiss HK, Aulmann M, Bach A, Lehnert T. Microbiological and clinical effects of selective bowel decontamination in transthoracic resection of carcinoma of the oesophagus and cardia. *Chirug* 2001;**72**:1160–70.

Health Technology Assessment Programme

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.ncchta.org)

Diagnostic Technologies & Screening Panel

Members

Chair,
Dr Ron Zimmern, Director of the Public Health Genetics Unit, Strangeways Research Laboratories, Cambridge

Ms Norma Armston, Lay Member, Bolton

Professor Max Bachmann Professor of Health Care Interfaces, Department of Health Policy and Practice, University of East Anglia

Professor Rudy Bilous Professor of Clinical Medicine & Consultant Physician, The Academic Centre, South Tees Hospitals NHS Trust

Dr Paul Cockcroft, Consultant Medical Microbiologist and Clinical Director of Pathology, Department of Clinical Microbiology, St Mary's Hospital, Portsmouth

Professor Adrian K Dixon, Professor of Radiology, University Department of Radiology, University of Cambridge Clinical School

Dr David Elliman, Consultant Paediatrician/ Hon. Senior Lecturer, Population Health Unit, Great Ormond St. Hospital, London

Professor Glyn Elwyn, Primary Medical Care Research Group, Swansea Clinical School, University of Wales Swansea

Mr Tam Fry, Honorary Chairman, Child Growth Foundation, London

Dr Jennifer J Kurinczuk, Consultant Clinical Epidemiologist, National Perinatal Epidemiology Unit, Oxford

Dr Susanne M Ludgate, Medical Director, Medicines & Healthcare Products Regulatory Agency, London

Professor William Rosenberg, Professor of Hepatology, Liver Research Group, University of Southampton

Dr Susan Schonfield, Consultant in Public Health, Specialised Services Commissioning North West London, Hillingdon Primary Care Trust

Dr Phil Shackley, Senior Lecturer in Health Economics, School of Population and Health Sciences, University of Newcastle upon Tyne

Dr Margaret Somerville, PMS Public Health Lead, Peninsula Medical School, University of Plymouth

Dr Graham Taylor, Scientific Director & Senior Lecturer, Regional DNA Laboratory, The Leeds Teaching Hospitals

Professor Lindsay Wilson Turnbull, Scientific Director, Centre for MR Investigations & YCR Professor of Radiology, University of Hull

Professor Martin J Whittle, Associate Dean for Education, Head of Department of Obstetrics and Gynaecology, University of Birmingham

Dr Dennis Wright, Consultant Biochemist & Clinical Director, Pathology & The Kennedy Galton Centre, Northwick Park & St Mark's Hospitals, Harrow

Pharmaceuticals Panel

Members

Chair,
Dr John Reynolds, Chair Division A, The John Radcliffe Hospital, Oxford Radcliffe Hospitals NHS Trust

Professor Tony Avery, Head of Division of Primary Care, School of Community Health Services, Division of General Practice, University of Nottingham

Ms Anne Baileff, Consultant Nurse in First Contact Care, Southampton City Primary Care Trust, University of Southampton

Professor Stirling Bryan, Professor of Health Economics, Health Services Management Centre, University of Birmingham

Mr Peter Cardy, Chief Executive, Macmillan Cancer Relief, London

Professor Imti Choonara, Professor in Child Health, Academic Division of Child Health, University of Nottingham

Dr Robin Ferner, Consultant Physician and Director, West Midlands Centre for Adverse Drug Reactions, City Hospital NHS Trust, Birmingham

Dr Karen A Fitzgerald, Consultant in Pharmaceutical Public Health, National Public Health Service for Wales, Cardiff

Mrs Sharon Hart, Head of DTB Publications, *Drug & Therapeutics Bulletin*, London

Dr Christine Hine, Consultant in Public Health Medicine, South Gloucestershire Primary Care Trust

Professor Stan Kaye, Cancer Research UK Professor of Medical Oncology, Section of Medicine, The Royal Marsden Hospital, Sutton

Ms Barbara Meredith, Lay Member, Epsom

Dr Andrew Prentice, Senior Lecturer and Consultant Obstetrician & Gynaecologist, Department of Obstetrics & Gynaecology, University of Cambridge

Dr Frances Rotblat, CPMP Delegate, Medicines & Healthcare Products Regulatory Agency, London

Professor Jan Scott, Professor of Psychological Treatments, Institute of Psychiatry, University of London

Mrs Katrina Simister, Assistant Director New Medicines, National Prescribing Centre, Liverpool

Dr Richard Tiner, Medical Director, Medical Department, Association of the British Pharmaceutical Industry, London

Dr Helen Williams, Consultant Microbiologist, Norfolk & Norwich University Hospital NHS Trust

Therapeutic Procedures Panel

Members

Chair,
Professor Bruce Campbell,
Consultant Vascular and General Surgeon, Department of Surgery, Royal Devon & Exeter Hospital

Dr Aileen Clarke,
Reader in Health Services Research, Public Health & Policy Research Unit, Barts & the London School of Medicine & Dentistry, London

Dr Matthew Cooke, Reader in A&E/Department of Health Advisor in A&E, Warwick Emergency Care and Rehabilitation, University of Warwick

Dr Carl E Counsell, Clinical Senior Lecturer in Neurology, Department of Medicine and Therapeutics, University of Aberdeen

Ms Amelia Curwen, Executive Director of Policy, Services and Research, Asthma UK, London

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

Professor Paul Gregg, Professor of Orthopaedic Surgical Science, Department of General Practice and Primary Care, South Tees Hospital NHS Trust, Middlesbrough

Ms Bec Hanley, Co-Director, TwoCan Associates, Hurstpierpoint

Ms Maryann L Hardy, Lecturer, Division of Radiography, University of Bradford

Professor Alan Horwich, Director of Clinical R&D, Academic Department of Radiology, The Institute of Cancer Research, London

Dr Simon de Lusignan, Senior Lecturer, Primary Care Informatics, Department of Community Health Sciences, St George's Hospital Medical School, London

Professor Neil McIntosh, Edward Clark Professor of Child Life & Health, Department of Child Life & Health, University of Edinburgh

Professor James Neilson, Professor of Obstetrics and Gynaecology, Department of Obstetrics and Gynaecology, University of Liverpool

Dr John C Pounsford, Consultant Physician, Directorate of Medical Services, North Bristol NHS Trust

Karen Roberts, Nurse Consultant, Queen Elizabeth Hospital, Gateshead

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

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 & Ptnrs), 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.ncchta.org) 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.

The National Coordinating Centre for Health Technology Assessment,
Mailpoint 728, Boldrewood,
University of Southampton,
Southampton, SO16 7PX, UK.
Fax: +44 (0) 23 8059 5639 Email: hta@soton.ac.uk
http://www.ncchta.org

ISSN 1366-5278