Health Technology Assessment 2005; Vol. 9: No. 2

Do the findings of case series studies vary significantly according to methodological characteristics?

K Dalziel, A Round, K Stein, R Garside,
E Castelnuovo and L Payne

January 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.

Do the findings of case series studies vary significantly according to methodological characteristics?

K Dalziel,[1] A Round,[1] K Stein,[1*] R Garside,[1] E Castelnuovo[1] and L Payne[2]

[1] Peninsula Technology Assessment Group, Peninsula Medical School, Universities of Exeter and Plymouth, Exeter, UK

[2] Wessex Institute for Health, University of Southampton, UK

* Corresponding author

Declared competing interests of authors: none. Ken Stein is a member of the editorial board for *Health Technology Assessment*, although he was not involved in the editorial process for this report.

Published January 2005

This report should be referenced as follows:

Dalziel K, Round A, Stein K, Garside R, Castelnuovo E, Payne L. Do the findings of case series studies vary significantly according to methodological characteristics? *Health Technol Assess* 2005;**9**(2).

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 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 02/33/01. As 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

Do the findings of case series studies vary significantly according to methodological characteristics?

K Dalziel,[1] A Round,[1] K Stein,[1*] R Garside,[1] E Castelnuovo[1] and L Payne[2]

[1] Peninsula Technology Assessment Group, Peninsula Medical School, Universities of Exeter and Plymouth, Exeter, UK

[2] Wessex Institute for Health, University of Southampton, UK

* Corresponding author

Objectives: To review the use of case series in National Institute for Clinical Excellence (NICE) Health Technology Assessment (HTA) reports, to review systematically the methodological literature for papers relating to the validity of aspects of case series design, and to investigate characteristics and findings of case series using examples from the UK's Health Technology Assessment programme.

Data sources: Electronic databases. NICE website. Reports produced as part of the UK's HTA programme.

Review methods: NICE HTAs that used information from case series studies were obtained from the NICE website and a range of quality criteria applied. Searches of electronic databases, handsearched journals and the bibliographies of papers were made in order to find studies that assessed aspects of case series design, analysis or quality in relation to study validity. Hypotheses relating to the design of case series studies were developed and empirically investigated using four case examples from existing reports produced as part of the UK's HTA programme (functional endoscopic sinus surgery for nasal polyps, spinal cord stimulation for chronic back pain, percutaneous transluminal coronary angioplasty and coronary artery bypass grafting for chronic angina). Analysis was undertaken comparing studies within each review.

Results: There was no consensus on which case series to include in HTAs, how to use them or how to assess their quality, despite them being used in 30% of NICE HTAs. No previous studies empirically investigating methodological characteristics of case series were found. However, it is possible that the search strategy failed to find relevant studies. Poor reporting of case series characteristics severely constrained analysis and there were insufficient data to investigate all the hypotheses. Findings were not consistent across the different topics and were subject to considerable uncertainty. All the examples in our analysis were surgical interventions, which are prone to additional confounding factors due to difficulties of standardisation compared with drug treatment. Our findings may not be generalisable outside the interventions studied. The case series reports included generally exhibited poor reporting of methodological characteristics. This constrained our analysis. The use of several methods of analysis has led to apparently discrepant results. Given the number of analysis performed, the usual level of significance ($p = 0.05$) should be viewed with caution. The most important limitation of this study is the small number of cases on which the findings are based. The results are therefore tentative and should be viewed with caution.

Conclusions: Case series are incorporated in a significant proportion of health technology assessments. Quality criteria have been used to appraise the quality of case series and decide on their inclusion in reviews of studies using this design. In this small series of case studies drawn from HTAs carried out for the NHS HTA programme, little evidence was found to support the use of many of the factors included in quality assessment tools. Importantly, no relationship was found between study size and outcome across the four examples studied. Isolated examples of a potentially important relationship between other methodological factors and outcome were shown, such as blinding of outcome measurement, but these were not shown consistently across the small number of examples studied. This study is based on a very small sample of studies and should therefore be considered as exploratory. Further investigation of the relationship between methodological features and outcome is

justified given the frequency of use of case series in health technology assessments. Further research into the methodological features of case series and their outcome is justified in a wider sample of technologies and larger sets of case series. Value of information analyses including case series could be explored. Further exploration of the differences between case series and randomised controlled trial results, preferably using registry or comprehensive case series data, would be valuable.

Contents

Glossary and list of abbreviations

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

Glossary

Analysis of variance A technique that isolates and assesses the contribution of categorical independent variables to variation in the mean of a continuous dependent variable.

Bias Processes leading to the deviation of results or inferences form the truth. Any trend in the collection, analysis, interpretation, publication or review of data can lead to conclusions that are systematically different from the truth.

Case series Uncontrolled observational studies involving an intervention and outcome for more than one person.

Confounding Distortion of the estimated effect of an exposure on an outcome caused by some extraneous factor (such as age, social class) associated with both.

Heterogeneity Differing aspects of study details across trials. These may be clinical (such as different doses, patient selection, points of disease). These can lead to statistical heterogeneity – where the results of trials are different from one other.

Observational studies An inclusive term for non-experimental studies, including surveys, case control and cohort studies. Although they may be comparative, they are often essentially descriptive.

Meta-analysis Techniques for combining the results of two or more independent studies to provide a test with more power than the originals to answer a question of interest.

Robust regression A form of regression, more conservative than ordinary least-squares regression, that resists the influence of outliers. It performs better with non-ideal data but results in slightly larger standard errors.

List of abbreviations

AIDS	acquired immunodeficiency syndrome
ANCOVA	analysis of covariance
ANOVA	analysis of variance
CABG	coronary artery bypass grafting
CAST	Cardiac Arrhythmia Suppression Trial
CI	confidence interval
CRD	Centre for Reviews and Dissemination
FESS	functional endoscopic sinus surgery
HRT	hormone replacement therapy
i.i.d.	independently and identically distributed
ITT	intention-to-treat
NICE	National Institute for Clinical Excellence
NIH	National Institutes of Health
NYHA	New York Heart Association
OLS	ordinary least-squares
PTCA	percutaneous transluminal coronary angioplasty
QEOs	quasi-experimental and observational studies
RCT	randomised controlled trial
REACT	Recruitment and Enrolment Assessment in Clinical Trials
RR	relative risk
SCS	spinal cord stimulation
SE	standard error
TAR	Technology Assessment Report
VAS	visual analogue scale

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

Objectives

This study had three aims:

1. To review the use of case series in National Institute for Clinical Excellence (NICE) Health Technology Assessment (HTA) reports.
2. To review systematically the methodological literature for papers relating to the validity of aspects of case series design.
3. To investigate characteristics and findings of case series using examples from the UK's Health Technology Assessment programme.

Background

Although randomised controlled trials (RCTs) offer the most robust evidence for effectiveness, this level of data is not always available for health technology assessments. Given that policy decisions still need to be made even in the absence of RCT evidence, it is important to try to understand the elements of case series design that determine their quality. Although a simple hierarchy of evidence will place case series as a weak form as evidence, individual studies, just like individual RCTs, may vary widely in quality and different studies of the same intervention may produce widely different estimates of outcome frequency. The validity of any study, whatever its form, will depend on the quality of its design, execution and interpretation. Nevertheless, case series studies are the most vulnerable to bias and confounding. RCTs attempt to minimise challenges to internal validity through minimising selection, performance, detection and attrition biases. However, this may lead to problems of external validity if strict exclusion criteria lead to a population being assessed, which is very different to that treated in clinical practice.

The aspects of quality that influence the validity of RCTs have been empirically studied, and it is generally agreed that adequate blinding, concealment and randomisation methods are crucial. A number of different scales and checklists for quality exist, but not all of them are empirically based or rigorously developed. As the authors were not aware of agreed aspects of quality for case series that were important, this study aimed to look at what types of quality measure had been used in NICE HTAs and to search the literature systematically to see if empirical studies of case series had been published.

While comparisons of the results from RCTs and other study designs have been undertaken, they have been restricted to observational studies with control groups. These yield conflicting results, with non-randomised studies variously showing greater treatment effects, similar treatment effects and lower treatment effects in different subject areas investigated. The evidence suggests that non-randomised controlled evidence shows more variance than RCTs and the direction of effect is unpredictable. As we were not aware of such investigations of case series and RCT results, we aimed to investigate this.

Review of the use of case series in NICE HTAs

Currently completed NICE HTAs were obtained from the NICE website. Of the 47 completed HTAs, 14 (30%) had included information from case series studies.

In two cases no RCTs were identified and the other 12 reports also included data from between two and 70 RCTs. The number of case series included ranged from two to 159. Inclusion criteria for case series included study size and length of follow-up. Various quality criteria were applied (n = 9), with the CRD Report criteria being used in three cases. Data from case series were used to confirm RCT results, to inform an economic model, to explore variation and for meta-analysis.

We found that there was no consensus on which case series to include in HTAs, how to use them or how to assess their quality, despite them being used in 30% of NICE HTAs.

Systematic review of methodological literature

We carried out searches in electronic databases, handsearched journals and examined the bibliographies of papers in order to find studies that assessed aspects of case series design, analysis or quality in relation to study validity. No empirical studies were found. However, it is known that searches that are sensitive enough to identify case series are difficult to design with appropriate specificity and it is possible that we failed to locate such studies.

Investigation of characteristics and findings of case series studies

A number of hypotheses relating to the design of case series studies were developed *a priori*. These were empirically investigated using four case examples from existing reports produced as part of the UK's HTA programme.

We included HTAs that had at least 40 case series studies available, included at least one good-quality RCT and contained information on the age of participants as a minimum description of the included population. We identified three reports on four topics – functional endoscopic sinus surgery for nasal polyps, spinal cord stimulation for chronic back pain, percutaneous transluminal coronary angioplasty (PTCA) and coronary artery bypass grafting (CABG) for chronic angina.

Data were extracted on outcome measures and study population characteristics.

Analysis was undertaken on a between-study level within each review. For each hypothesis, continuous variable data were explored through scatter plots and robust regression. Regression analysis weighted by sample size were also performed. Binary data were explored through *t*-tests and Mann–Whitney tests. Analysis of variance (ANOVA) was also performed, weighted for sample size. Multivariate analysis using disease severity, age and male sex was performed using multivariate robust regression or ANOVA as appropriate.

Comparisons between cases series and RCTs were performed using the intervention arms of RCTs as a comparator. There were only enough data to do this for PTCA and CABG. Meta-analysis of RCT data was compared with weighted robust regressions using the intervention as the confounding factor and estimating the coefficient size.

Poor reporting of case series characteristics severely constrained analysis and there were insufficient data to investigate all the hypotheses. Findings were not consistent across the different topics and were subject to considerable uncertainty.

No relationship was found between sample size and outcome frequency. No relationship was found between prospective data collection and outcome frequency. One analysis each (in different topic areas) found a significant association between multi-centre studies and outcome, between independent outcome measurement and outcome frequency and between earlier publication and outcome frequency. Length of follow-up was found to be significantly associated with outcome frequency in three analyses. One topic area had scored case series for quality and this was found to be associated with outcome. However, this quality score contained items which we investigated separately in this review, without evidence of impact.

Compared with RCT evidence, which showed no difference between PTCA and CABG, case series estimates of mortality showed a 1–2% increase in mortality for CABG. For angina recurrence, neither case series nor RCT data showed any difference between the two interventions.

Limitations

We found no previous studies empirically investigating methodological characteristics of case series. However, it is possible that the search strategy failed to find relevant studies.

All the examples in our analysis were surgical interventions, which are prone to additional confounding factors owing to difficulties of standardisation compared with drug treatment. Our findings may not be generalisable outside the interventions studied.

The case series reports included generally exhibited poor reporting of methodological characteristics. This constrained our analysis.

The use of several methods of analysis has led to apparently discrepant results. Given the number of analyses performed, the usual level of significance ($p = 0.05$) should be viewed with caution.

The most important limitation of our study is the small number of cases on which our findings are based. The results are therefore tentative and should be viewed with caution.

Conclusions

Case series are incorporated in a significant proportion of health technology assessments.

A wide range of quality criteria have been used to appraise the quality of case series and decide on their inclusion in reviews of studies using this design. In this small series of case studies drawn from HTAs carried out for the NHS HTA programme, we found little evidence to support the use of many of the factors included in quality assessment tools. Importantly, we found no relationship between study size and outcome across the four examples studied.

Isolated examples of a potentially important relationship between other methodological factors and outcome were shown, such as blinding of outcome measurement, but these were not shown consistently across the small number of examples studied.

Comparison of case series and RCT data was possible in only two examples studied but demonstrated a greater range in outcomes reported in case series, reflecting the likelihood that this design includes different populations. However, outcomes were not better in case series, contrary to expectations.

Estimates of comparative efficacy of alternative techniques by comparing case series studies were shown to be different from analyses based on RCTs. However, it is not clear from this whether this is an effect of confounding or indicates different efficacy in different populations.

This study is based on a very small sample of studies and should therefore be considered as exploratory. Further investigation of the relationship between methodological features and outcome is justified given the frequency of use of case series in health technology assessments.

Need for further research

Further research into the methodological features of case series and their outcome is justified in a wider sample of technologies and larger sets of case series.

Value of information analyses including case series could be explored.

Further exploration of the differences between case series and RCT results, preferably using registry or comprehensive case series data, would be valuable.

Chapter 1

Aim

The aims were to examine the existing use of case series evidence in Natural Institute for Clinical Excellence (NICE) Technology Assessment Reports (TARS), to review systematically the literature regarding the methodological characteristics of case series, to investigate characteristics of case series studies, and to determine whether the findings of case series studies vary significantly according to methodological characteristics

This report is composed of three main sections:

1. a review of case series use in NICE Health Technology Assessments (HTAs)
2. a systematic review of methodological literature for papers relating to the validity of aspects of case series design
3. an investigation of characteristics and findings of case series (using four examples from the HTA programme).

Chapter 2
Background

Definition of case series

Case series are uncontrolled observational studies involving an intervention and outcome for more than one person.[1]

Case series in health technology assessment

Health technology assessment is the evaluation of the effectiveness, costs and wider impact of health technologies, defined as all methods used by health professionals to promote health, prevent and treat disease and improve rehabilitation and care. Central to the conduct of health technology assessment is establishing effectiveness. A range of study types may be relevant to this aspect of health technology assessment and these can be arranged in a hierarchy according to the extent to which the design minimises bias (see *Box 1*).

Although a hierarchy of the strength of evidence for effectiveness based on different study designs in healthcare research is now well established, the need for early policy decisions on adoption or reimbursement within healthcare systems means that new technologies are frequently subject to health technology assessment. For a variety of reasons it may be necessary to include non-randomised controlled trial (RCT) data, including case series, in health technology assessments. Indeed, uncontrolled case series are sometimes the only type of evidence available.[2] It may also be necessary to review case series data as part of a health technology assessment in the following circumstances:

- Where there is strong political or social pressure surrounding the funding of a new technology and a lack of comparative evidence.
- Where the apparent effectiveness from uncontrolled studies is so promising that waiting for comparative evidence is considered unacceptable.
- Where comparative evidence may not be available for all important outcomes or all populations, or comparative evidence is available in only a restricted population. Case series data may therefore be sought to supplement more rigorous evidence or provide information on possible effects in a less selected population.
- Where case series data provide more detailed evidence regarding the safety of a technology (such as the detection of rarer adverse events through large series).
- Where length of follow-up is a critical issue for an assessment and case series provide more data than comparative studies.

This report is predicated on the view that if decisions are to be made in the absence of more rigorous evidence from well-designed RCTs, then it is important to try and understand whether there are elements of case series design that determine their quality. The National Institutes of Health (NIH) in the USA estimated that only about 20% of health technologies currently in use have been analysed by means of an RCT,[3] although higher proportions have been estimated for the most commonly used interventions.[4,5] Indeed, there are some fields (for example, cancer treatments) where ethical constraints or the attitudes of physicians and their patients mean that the use of randomised controls to evaluate efficacy is not universally regarded as essential.[6] Clinicians working in the field of AIDS developed a set of criteria to define the circumstances under which they believe trials can be carried out without randomised controls:

- No appropriate control exists.
- Untreated patients have universally poor prognosis.
- Expected side-effects of treatment do not compromise benefits.
- Expected benefits must be large and unambiguous.
- Strong scientific rationale for treatment will lead to widespread acceptance.[6]

As health technology assessment is policy driven, and decisions need to be made on the basis of available evidence, it is likely that case series will continue to form some or all of the evidence base for effectiveness for some technologies. Given this, and the large volume of case series evidence that may need to be addressed in assessments, it is important to evaluate how case series studies can best be utilised.

BOX 1 *Hierarchy of study designs for studies of effectiveness [from CRD Report No. 4 (2nd edition)]*

Level description

1. Experimental studies (e.g. RCT with concealed allocation).
2. Quasi-experimental studies (e.g. experimental study without randomisation).
3. Controlled observational studies.
   (a) Cohort studies.
   (b) Case–control studies.
4. Observational studies without controls.
5. Expert opinion based on pathophysiology, bench research or consensus.

Description of selected study designs

Experimental

A study in which some conditions, particularly decisions concerning the allocation of participants to different intervention groups, are under the control of the investigator.

Randomised controlled trial

Follow-up of participants randomly allocated to intervention or control groups, with a comparison of outcome rates during the time covered. Randomisation (with concealment of allocation sequence) avoids bias because both known and unknown determinants of outcomes are on average evenly distributed between intervention and control groups.

Quasi-experimental

A study on which the allocation of participants to different intervention groups is controlled by the investigator but the method falls short of genuine randomisation and allocation concealment.

Observational

A study in which natural variation in interventions or exposure among study participants is investigated to explore the effect of the interventions or exposure on health outcomes.

Cohort study

Comparison of outcomes between participants who have received an intervention and a group that has not (i.e. not allocated by the investigator) in a follow-up study.

Case–control study

Comparison of exposure to interventions between participants with the outcome (cases) and those without the outcome (controls).

Cross-sectional study

Examination of the relationship between disease and other variables of interest as they exist in a defined population at one particular time.

Before and after study

Comparison of findings in study participants before and after an intervention.

Case series

Description of a number of cases of an intervention and outcome (without comparison with a control group).

RCTs versus non-RCTs

Research has already been undertaken to examine the differences in estimated treatment effects estimated through different study designs. However, such studies have concentrated on study designs that contain a control group. It has been assumed that non-randomised studies were likely to overestimate treatment effects compared with RCTs. However, this has not been consistently found in empirical studies. Some recent papers have suggested that there is little difference between the results produced by RCTs and well-conducted observational studies.[7–10] However, it is noteworthy that these papers excluded case series and observational studies with historical controls. Further, it has been noted that these studies have looked at only a very few topic areas and, given the hundreds of thousands of trials conducted, they are likely to be subject to strong selection bias.[11]

In fact, where the results from RCTs and non-randomised observational studies of the same treatment have been compared, various findings have been noted. Kunz and Oxman[12] reviewed eight empirical studies from 1977 to 1996 that compared the results of RCTs and non-

randomised studies of the same interventions. They found that in five the non-randomised studies overestimated the effects compared with RCTs, in two they underestimated the effect and one study found similar effects in both randomised and non-randomised studies.

Linde and colleagues[13] examined the results of 24 RCTs and 25 studies of other design (non-randomised cohorts, cross-sectional surveys and case series) of using acupuncture for chronic headache. They found that non-randomised studies of good quality yielded similar results to RCTs. Interestingly, they devised a quality score which concentrated on specific clinical aspects such as a clear headache diagnosis, use of a headache diary, at least two clinical headache outcomes and follow-up at both early and late dates, in addition to methodological elements. They argued that greater attention should be paid to the rigour of clinical elements of trials rather than the current focus on methodological characteristics.

RCTs of the same interventions and controls can produce very different estimates of treatment effect. Overall the evidence suggests that non-RCTs demonstrate more variance than RCTs and the direction of estimated effect is unpredictable.

Types of potential bias

Case series are considered to be the weakest study design from which to obtain evidence on effectiveness and, as shown in *Box 1*, results derived from them are ranked as low quality. However, like RCTs, case series studies may vary enormously in quality. Whatever broad study design is used, the validity of its results will depend on the quality of the study that produced it and, where a particular type of study design is to be included in an assessment, it is necessary to appraise quality. By high-quality trial we mean one in which bias is minimised and the effect demonstrated likely to be a 'true' reflection of reality.

Case series vary in size and may be retrospective or prospective. Some may follow a predefined protocol whereas others collect data opportunistically. Inclusion processes vary, in particular with respect to whether consecutive cases are enrolled in prospective studies and whether inclusion and exclusion criteria are applied. The absence of a control group severely limits the utility of case series in healthcare decision-making as it is not clear if the outcome can be attributed to the intervention. Case series studies have no way of controlling for a placebo effect, which has been shown to have a significant effect compared with no treatment for many outcomes such as pain.[14,15] Furthermore, case series are prone to a range of biases, including sampling and selection, detection and reporting, observer and measurement, response and publication. Subjects who are aware of the purpose of a study are further prone to the effects of 'social desirability bias' where they may behave according to socially accepted norms or the expectations of the researcher, leading to possible inaccuracies in the reporting of behaviour or symptoms.[16]

In their comparison of effect sizes derived from non-randomised and randomised studies, MacLehose and colleagues[10] present the likely effects of different kinds of bias on effect sizes. This is shown in *Table 1*.

TABLE 1 *Most likely effects of different kinds of bias on effect size estimates*[10]

	More extreme	Less extreme	Either
Information bias			
Outcome: non-differential		✓	
Outcome: differential	✓		
Intervention: non-differential		✓	
Intervention: differential	✓		
Confounder: non-differential			✓
Confounder: differential	✓		
Selection bias or confounding			
RCTs and quasi-experimental cohort studies	✓		
Observational cohort studies			✓
Case–control studies			✓
Differential care bias	✓		

Information bias (misclassification or error in measuring, outcomes or confounding variables) is classed as non-differential where it affects intervention and control groups equally or differential if it affects them unequally.[10]

MacLehose and colleagues[10] note that selection bias arises in different ways in experimental and non-experimental study designs. For RCTs and quasi-experimental studies, selection bias arises as a result of biased allocation of patients to groups, and is likely to result in a more extreme beneficial effect. For observational studies, the effect may be in either direction, as clinicians and researchers may elect to treat those patients who they believe will benefit, or treat those who are judged to be unlikely to benefit from standard treatment but are sicker and at higher risk of a poor outcome.[10]

Using two clinical examples, this review compared the effect size estimates obtained from RCTs and quasi-experimental and observational studies (QEOs). They assessed the quality of studies based on criteria modified from Downs and Black[17] (quality of reporting, external validity, internal validity – bias, confounding or selection bias, power of the study). Within the topic areas they examined (mammographic screening of women to reduce breast cancer mortality and folic acid to prevent neural tube defects), they found that high-quality QEOs showed no tendency for effect sizes to be more extreme than for RCTs. A more extreme effect was seen in poor-quality QEOs. They concluded that within these topic areas, QEOs may give valid results if important confounders are controlled for. However, they also found that the reporting of QEOs was poor and that it was difficult adequately to distinguish and measure variations in different aspects of quality between studies. No association was found between study quality and effect size, suggesting that relative risk (RR) is not associated with quality in a predictable way. Alternatively the instrument used failed to distinguish methodological quality adequately.

Internal and external validity trade-off

RCTs are designed to maximise internal validity.[18] Internal validity is addressed by trying to minimise selection bias, performance bias, detection bias and attrition bias through controlling the characteristics of study participants, achieving truly random treatment allocation, blinding of participants and researchers and applying a consistent approach to deviations from protocol and loss to follow-up through intention-to-treat analysis. External validity relates to the extent to which a study's results can be applied to other circumstances – its generalisability to other groups of patients or to usual clinical care. It encompasses aspects such as the study population, treatment regimen, levels of other care, type of outcomes and length of follow-up.

RCTs typically have relatively strict selection criteria which often exclude the old and the young, those with co-morbidity or those thought likely not to comply.[18] In addition, patients choosing to enrol in RCTs may be different from those who do not wish to participate. For example, the Recruitment and Enrolment Assessment in Clinical Trials (REACT) project assessed 140 patients who had declined to participate in the Cardiac Arrhythmia Suppression Trial (CAST) compared with 260 who did enrol.[19] Enrollees were more likely to be male ($p < 0.001$) and to be younger ($p = 0.002$) than non-participators, and also less likely to have medical insurance ($p < 0.001$) (US study). Enrolled patients were more likely to have had at least one episode of ventricular tachycardia (25% vs 14%; $p = 0.025$). Other differences relating to psychological factors, health beliefs and understanding of informed consent were also found.

The systematic review of clozapine-treated subjects with treatment-resistant schizophrenia by Brambilla and colleagues[20] included non-RCT data because of the highly selected nature of RCT samples and the definition of 'treatment-resistant schizophrenia', neither of which reflected clinical practice. The authors included 50 studies of which 18 were clinical trials, 23 were prospective observational studies and nine were retrospective observational studies. They found that RCTs on this topic included small samples of highly selected patients whereas observational studies used entry criteria that were similar to everyday clinical practice and had longer follow-up than RCTs. RCTs indicated a smaller treatment effect and higher drop-out rate than observational studies, and retrospective studies showed the greatest effect and the smallest drop-out rate. Study quality was not specifically examined but the results suggest that larger sample sizes, longer follow-up and prospective data collection may be quality markers.

In comparing elderly patients with aggressive histology lymphoma who were entered or not entered into a randomised trial, Chen and colleagues[21] found that non-randomised patients were older, had a poorer performance status, were less likely to be given a treatment with a curative

rather than palliative intent and were less likely to complete six cycles of treatment. Five-year survival was found to be superior for those patients entered into the RCT.

McKee and colleagues[22] concluded from their study of 18 papers evaluating interventions by both randomised and non-randomised study designs that those excluded from RCTs tend to have a worse prognosis than those included. They also noted that those in RCTs for treatment tended to be less affluent, educated and healthy than those who were not, whereas the reverse case was seen in RCTs of preventative interventions.

RCTs which are designed to test the efficacy of interventions are more likely to maximise internal validity. In contrast, large pragmatic trials (e.g. the AD2000 trial of donepezil for dementia[23]) aim to evaluate the effectiveness of interventions in conditions that reflect, as closely as possible, routine clinical practice in relatively undifferentiated populations. However, in order to promote recruitment and retention, such trials typically use a limited number of simple outcome measures, thereby limiting the extent to which the study can demonstrate the mechanisms by which effectiveness is achieved. However, such trials typically follow smaller efficacy studies in more exclusive populations.

As with RCTs, the external validity of case series studies may be variable and may be difficult to judge where reporting of study populations is unclear. Large, prospective and comprehensive case series might be argued to have high external validity. However, as we show later, reporting of patient characteristics is frequently poor in case series studies and it may be difficult for readers to judge the generalisability of such studies. Whether there is a trade-off between external and internal validity in case series is open to question. Given that the design has inherent and serious weaknesses, increased patient selection is unlikely to yield the same benefits to internal validity in an uncontrolled study as in an RCT.

Assessment of quality: RCTs

In order to judge the reliability of study results which will inform decision-making, it is important for the quality of those studies to be assessed. Given the increased reliability of effect estimates gleaned through well-designed RCTs, it is understandable that most examinations of study quality have concentrated on assessing studies of this design. Assessment of RCT quality using standardised checklists and scales according to various indicators is well established. The CONSORT statement was produced in 1996 and aimed to ensure that important aspects of studies design (such blinding, concealment and adequate randomisation) were described in trial reports.[24] However, the CONSORT statement contains elements that relate to the reporting (such as use of a flow chart to show participant pathways) and conduct (such as a description of protocol deviations) of trials in addition to their methods. Although some aspects (such as blinding and proper allocation) have been shown to quantitatively affect the results of a study,[25–28] others have not. Concealment of treatment allocation appears to be the most important aspect for preventing bias in RCTs.[25,26]

Many different scoring systems for rating the quality of trials have been used, some of which produce a single summary score to indicate quality. Summary scores may be particularly problematic if they are used to score the quality of trials for meta-analysis.[29] An annotated bibliography of scales developed to assess the quality of RCTs was published by Moher and colleagues in 1995.[30] This found 25 scales (23 published) and nine checklists used in the literature which aimed to assess the quality of RCTs. The authors stated that there are a number of shortcomings with these scales. Fifteen (60%) were designed to assess the quality of a trial of any type of intervention, whereas 10 (40%) were aimed specifically at certain types of intervention or outcome (e.g. contrast media, pain). The number of items used to assess a trial ranged from three to 34 (median 15). The scales could be scored in between <10 and 45 minutes (median 10 minutes) and yield total possible scores of 1 to 170.

Only one scale (that by Jadad and co-workers[31]) met the authors' criteria of 'rigorous development', by which they meant that the paper documented how items were initially selected, how and why the final items were included, how the scale differentiated between trials of different quality and on the range of scores obtained through its development.

The authors noted that scales that yield higher scores for double blinding will automatically discriminate against surgical trials in which blinding is not possible. It does not mention blinding for outcome measurement. Other aspects of the development and use of the different scales are shown in *Table 2*.

TABLE 2 *Aspects of the development and use of quality scores assessed by Moher and colleagues*[30]

Item in the development of quality scales	Number (%)
Defined the construct quality used in their scale development	6 (24)
Designed to assess quality of trial report	3 (12)
Designed to assess methodological quality	8 (32)
Designed to assess both methodological and trial report quality	14 (56)
Used 'accepted criteria' to select items for inclusion from textbooks of clinical trials	24 (96)
Includes at least one item regarding patient assignment	22 (88)
Includes at least one item regarding masking	20 (80)
Includes at least one item about follow-up	11 (44)
Includes at least one item about statistical analysis	21 (84)
Describes how the items should be scored when assessing quality	18 (72)
Reported inter-rater reliability	12 (48)
Provides detailed instructions on how scores should be assigned to each item	17 (68)
Uses a weighting system to score quality	8 (32)
Recommends steps to minimise bias for those completing quality assessments	4 (16)

In a paper referred to in the article by Moher and co-workers, the same authors used six different scales to assess the quality of 12 trials used in a meta-analysis. They found that overall quality scores varied considerably across the scales and that differences in trial quality varied from 23% to 74% across scales. Similar results were obtained using rank scores of individual trials. Other authors have come to similar conclusions.[29]

With scoring systems for quality, the reasons for the weight that each item is given is often unclear. It is also uncertain whether a combination of small imperfections of a study could be as detrimental to a study's validity as the failure of a single, crucial element such as blinding.[18]

Some controversy clearly remains about the value of some of the scales used to evaluate quality for RCTs. It must be useful to distinguish those trials that have such serious flaws that they may call the validity of the results into question. However, whereas some aspects of quality have been empirically shown to affect results, others have not. The way in which authors have chosen to weight quality elements in those checklists that produce a single quality score is not usually evidence based. Further, using different scales, the same trial may receive inconsistent quality scores whereas different reviewers using the same scale may find little inter-rater reliability. With such uncertainty, the incorporation of quality data by weighting trials in the pooling of RCT results remains controversial.

As stated by Fletcher[18] in a discussion about evaluating interventions, "grading of trials according to specific aspects of their design, conduct and reporting, assumes that we know the degree to which each element of the trials, if improperly done, biases the results and how much. Unfortunately, we do not." In a stronger caution against using quality scores as a way of influencing trials used in meta-analyses, Shapiro[32] argues that "quality cannot be scored, measured, taken into account" but that "quality is best evaluated qualitatively … the author should give good reasons for judging the quality of any given study as good or bad in transparent and easily understood language."

Existing assessments of study quality: relevance to case series

The discussion above outlines the current debates on the manner and value of different methods of assessing the quality of RCTs and how the information from quality assessment should be used in systematic reviews and meta-analyses. There is even less certainty and understanding as to which aspects of observational data, especially those obtained through case series study design, are crucial to quality. It is not known if these influence estimated treatment effects and, if they do, in what way. Case series studies have traditionally been excluded from most systematic reviews as they represent the lowest level of study evidence in most simple hierarchies of study design.

Quality scales and checklists discussed in this section were identified in three ways: through the search strategy described in Chapter 3, through the studies identified in this chapter and from an unpublished review by J. J. Deeks (Centre for Statistics in Medicine, Institute of Health Sciences, Oxford) and J. Dinnes (Southampton Health Technology Assessments Centre, University of Southampton), 2003.

The review of HTAs carried out for NICE using cases series is described fully in Chapter 3. This found eight types of checklists used including the Cochrane framework, the Centre for Reviews and Dissemination (CRD) Report (No. 4) and modified Spitzer checklists.[33] There are a number of additional methods of assessing the quality of case series studies, which are reproduced for information in Appendix 3.

The review by Deeks and Dinnes evaluates checklists and scales for report evaluation. They found 194 tools, mostly published. The purpose of the quality assessment tools was mostly for inclusion in systematic reviews (65%) or critical appraisal (35%). The latter are primarily designed to help clinicians assess the usefulness of an intervention in their practice, rather than comparing quality across studies. The tools included those which had been designed for RCTs, some of which can also be used, with or without modification, for other study designs. Most identified tools were designed to be used for more than one study design and none dealt solely with case series design. Only 23 (17%) included items specific to different study designs. It is likely that all case series will score poorly on this type of scale as the elements relating to control groups, randomisation and blinding of subjects will all be missing.

The review by Dinnes and Deeks is largely concerned with tools for observational studies that do contain control groups. None were identified that were designed solely for case series studies. The authors consider the most important items in a scale to be those relating to the creation of the intervention groups and those relating to the comparison of the groups at the analysis stage. The items in these domains are as follows:

Creation of intervention groups:

- generation of random sequence*
- concealment of allocation*
- how allocation occurred*
- any attempt to balance groups by design*
- description of study design
- contamination.

Analysis – comparability:

- assessment of baseline comparability*
- identification of prognostic factors
- case mix adjustment.

Clearly, by design, uncontrolled case series studies will not adhere to those items marked with an asterisk above. This is also a problem with many of the identified checklists which are designed to be used with a number of different study designs, including uncontrolled observational studies (case series). The evaluation tools presented below are a selection of those available; further examples of checklists and evaluation tools developed are shown in Appendix 4. Those presented in this section give a flavour of the types of checklists that are available. A systematic review of these scales was not an aim of the project. Of the quality checklists identified that can be used with case series design, none were validated quantitatively. Further, a large range of items were considered to be worth including in such checklists, including items relating to the presentation of the results, the way in which the paper was presented and written, the development of hypothesised conclusions and as items such as sampling, measurement and analysis. There was little apparent consensus. The number of items included in checklists and scales ranged from six to 65 (median 35) and some may take considerable time to complete. Further, the questions may be open to considerable rater interpretation and many checklists had not been checked for inter- or intra-rater reliability (see Appendix 4 for details).

The CRD at York (CRD Report No. 4, 2nd edition, 2001) suggest some quality criteria for assessment of case series:

- Is the study based on a representative sample selected from a relevant population?
- Are the criteria for inclusion explicit?
- Did all individuals enter the survey at a similar point in their disease progression?
- Was follow-up long enough for important events to occur?
- Were outcomes assessed using objective criteria or was blinding used?
- If comparisons of subseries were being made, was there sufficient description of the series and the distribution of prognostic factors?

Several aspects of this list may be difficult to assess in practice, especially in the absence of well-reported inclusion and exclusion criteria, for example, patients entering at a similar point in disease progression or sufficient follow-up period.

The Cochrane Reviewers' Handbook[1] does not provide any specific advice on how to assess the quality of case series. The Cochrane Non-randomised Study Methods Group is currently producing guidelines on the use of non-randomised studies in Cochrane reviews. Draft chapters are

available at http://www.cohrane.dk/nrsmg.htm. Currently the draft chapter expects the eventual advice on case series studies to be exclusion from Cochrane reviews of effectiveness.

Downs and Black[17] designed and assessed a checklist capable of looking at both randomised and non-randomised study designs. The study aimed to test the feasibility of creating a valid and reliable checklist that was appropriate for use with both study designs and which produced an overall quality score that included elements for external validity in addition to reporting quality, internal validity and power. The authors state that external validity has been neglected as a quality criterion in scales designed for RCTs. However, despite generally finding good inter-rater and test–retest reliability, the authors found that the items used to assess external validity were least reliable. They suggest that the items may require further study and that reviewers need to be trained to assess external validity.

A study by Boulware and co-workers systematically reviewed RCTs, observational studies with controls and observational studies without controls (case series) describing behavioural interventions for hypertension.[34] All study types were evaluated for rigor of study design and given a score (maximum 100). Case series studies were evaluated based on:

1. Study population:
   (a) description of inclusion and exclusion criteria
   (b) description of study population in terms of demographics and pertinent clinical characteristics
   (c) similarity between those enrolled and not enrolled.
2. Intervention description and outcomes assessment:
   (a) description of the intervention
   (b) handling and comparability of withdrawals
   (c) relevance and description of outcomes.
3. Statistical analysis and reporting:
   (a) mention of power calculation
   (b) appropriateness of statistical tests
   (c) presentation of statistical significance
   (d) adjustment for potential confounding.

As additional areas (such as quality of randomisation and blinding) are also included in the scoring systems for observational trials with controls and RCTs, and all totals add up to 100, different scores are given to each item across the different study designs. Weights given to different aspects do not appear to be evidence based and the influence of each item on the reliability of the results is not explored.

Cowley (1995)[35] developed a list of critical appraisal criteria to evaluate uncontrolled case series for total hip replacement and separate tools for RCTs and other comparative studies. It is not clear how the items for evaluation were selected and only one author made the judgements. Her key criteria were:

- method of selection of patients identified and appropriate
- number of patients deceased or lost to follow-up reported or included in appropriate statistical analysis
- follow-up period, range and mean given
- prosthesis models specified
- clearly defined criteria for measuring outcomes.

An additional 12 items were listed as other criteria, including blinding of radiological assessors and clinical evaluation performed independently of the operating surgeon (criteria met by 5% and 11% of papers, respectively).

A checklist from DuRant[36] provides separate design and result questions for different study designs, including retrospective chart (medical record) and survey designs and cross-sectional studies. The lists are lengthy (32 core items and then 10–31 items on design and analysis), although not designed to be exhaustive.

Coleridge Smith[37] also developed a checklist for measuring quality of case series and other designs. It was developed by a task force wishing to evaluate literature on chronic venous disorders of the leg. Items contained in the checklist are:

1. Study population
   (a) Was the study population adequately described?
   (b) Were the eligibility criteria explicit and appropriate?
   (c) Were participation rates adequate in all groups?
   (d) Were groups similar as to potentially confounding variables?
   (e) If not, were differences appropriately accounted for?
2. Exposure
   (a) Is the exposure adequately described?
   (b) Was the measure of the exposure sound?
   (c) Is the exposure variable valid (i.e. expresses the actual exposure)?

   (d) Is the exposure variable reliable (i.e. repetition gives similar results)?
   (e) Were observers blinded to outcome status?
   (f) If not was the outcome ascertained equally in all exposure groups?
3. Outcome
   (a) Is the outcome adequately described?
   (b) Was the measurement of the outcome sound?
   (c) Is the outcome variable valid (i.e. addresses the actual outcome)?
   (d) Is the outcome variable reliable (i.e. repetition gives similar results)?
   (e) Were observers blinded to exposure status?
   (f) If not, was the outcome ascertained equally in all exposure groups?
4. Statistical analysis
   (a) Was the statistical analysis adequate?
   (b) Were tests appropriately selected and used?
   (c) If not, are results reported so that the appropriate analysis can be undertaken?
5. Follow-up
   (a) Was the follow-up rate adequate?
   (b) Was the follow-up of sufficient duration?
6. Overall rating of internal validity.
7. Overall rating of clinical relevance.
8. Key headings.
9. Reason for overall ratings.

The inclusion of a question related to the control group demonstrates that this is not designed solely for case series studies.

For the first five items, studies are scored 'yes', 'no', 'partly', 'not reported', 'not applicable', 'I do not know'. For overall ratings of internal validity and clinical relevance, a four-point scale, 'strong', 'moderate', 'weak' or 'very weak' was used. Agreement between reviewers was tested using kappa and weighted kappa tests. Agreement about retaining a paper ranged from 0.18 [95% confidence interval (CI) 0.06 to 0.42) to 0.85 (95% CI 0.65 to 1) depending on whether, as in the former case, two clinicians were evaluating a paper or two methodologists, as in the latter.

For the purposes of this review, we were interested in trying to identify basic elements of study design which may have a measurable impact on the resultant effect size estimates. None of the checklists and evaluation tools found contain such a quantitative element. On a practical level, many are also very lengthy, and may be impractical for use in NICE TARs which are time constrained or where the number of case series on a topic is large. The included items for the scales identified by the review are presented in Appendix 4.

Sources of heterogeneity and bias in case series studies

Previous sections have considered general issues of study design and bias, particularly the importance of randomisation, the tools available to appraise quality and their relevance to case series. Many characteristics may give rise to bias within studies and heterogeneity between studies. The following list of factors are not unique to case series but their potential effect is considered. Some of these factors form the basis for hypothesis testing later in this report.

1. Population differences
Population differences such as age, sex, race, other co-morbidity, severity of disease (as assessed by severity score, setting of study, previous treatment). Whether age, sex, race or other co-morbidities have any relationship to outcome may depend on the condition under study.

2. Degree and nature of selection
In normal clinical practice, it is desirable that health professionals are capable of identifying individuals who will respond to treatment. However, where a particular treatment has been actively chosen over other treatments because of anticipated greater effect in that person, this may become a confounding factor. Indications for treatment may vary from centre to centre and may be associated with outcome independently of treatment. Studies that do not have consecutive enrolment into a study may be particularly prone to this bias. In addition, multi-centre case series may recruit only small numbers from each site and could be more prone to selection bias.

3. Random variation
Other things being equal, small samples may show more variation about a true population effect than larger studies. However, the effect of this can be estimated from CIs around the result.

4. Placebo effect/other treatment effect
Placebo effect has been shown to affect outcomes positively, especially continuous outcomes such as pain.[14,15] In addition, participants who know that they are enrolled in a study may do better than those who receive 'normal care' as they perceive enhanced treatment. Health professionals who know that a study is taking place may offer different treatment or care to that normally offered or project protocols may dictate higher than usual levels of care. Prospective enrolment may therefore have an influence on apparent effect.

5. Performance bias
The extent to which the intervention is clearly defined and then delivered in a standardised way is a potential source of bias in all intervention studies. In cases series where selection criteria are poorly defined or reported, it may be difficult to judge whether all cases were treated similarly. Longer term case series may be subject to important confounding effects from developments in other technologies (e.g. changes in relevant diagnostic techniques or improved after-care following surgery)

6. Loss to follow-up
This may be overt in prospective studies or covert in retrospective studies where notes may not be available for all patients undergoing treatment. If the chance of follow-up is associated with outcome, then the overall estimate of outcome will be biased. It is probable that this bias is more likely with retrospective studies. One study found that discrepancies in effect beyond chance between randomised and non-randomised studies were less common in prospective studies.[38]

7. Length of follow-up
Length of follow-up is often associated with outcome. People may have a more or less favourable outcome the longer the follow-up period, depending on the natural history of the condition and the intervention. Some adverse effects may only be identified through longer term follow-up.

8. Measurement bias
Few case series studies report using independent or blind outcome measurement, a design feature which has been shown to be of importance in controlled studies.

9. Date and place of publication
It may be the case that earlier studies were of poorer quality or less consistently reported than more recent studies. Publication bias of a new intervention means that more dramatic results are more likely to be published. Case series studies with negative results may be even less likely to be published than RCTs with negative results.

10. Differences between centres
Differences in technique or expertise or application of intervention may exist between centres. This is a true variation in outcome.

There is some consensus over which characteristics are likely to be important but also some uncertainty, for example whether a larger sample size is better. Egger and colleagues in their book *Systematic Reviews in Health Care* (2001)[39] emphasise that, when assessing observational studies, confounding and bias should be assessed and that sources of heterogeneity should be explored carefully. They contend that unlike with RCTs, where smaller studies are assumed to be more subject to chance (have larger CIs) and are weighted accordingly,[39] larger observational studies are not necessarily better and so should not be given more weight:

> "In the case of observational studies the main problem is not lack of precision but is bias or confounding, therefore smaller studies may give more accurate results. They may be able to devote more attention to characterising the exposure and confounding factors than larger studies, and then may be able to collect more details."[39]

There may also be trade-offs between criteria, such as prospective design and large sample size. Prospective studies may reduce selection bias but are likely to be smaller than retrospective studies. Underlying all these considerations is the difficulty of separating the effects of these potential biases and confounders from any actual treatment effect.

Publication bias represents a further potentially important challenge in systematic reviews which include case series, but the scale of the problem is not well understood. Given the inherent methodological limitations of the study design, small case series which suggest relatively modest effectiveness may, theoretically, be more prone to publication bias than other designs. This may take the form of exclusion from the published literature or publication in more obscure journals which are not indexed on major electronic databases and so less likely to be identified through standard literature searching.

Conclusions

Evidence from case series studies is used in HTA reports in a number of scenarios. Although, as a study design, case series is considered a weak form of evidence, it is not known what aspects of study design may contribute to one case series study being of better quality than another. Appraisal tools for case series exist but none is exclusively for case series design and none appear to be evidence based. Generic tools for quality appraisal will highlight the limits of uncontrolled case series, but they are little help beyond this in identifying crucial elements that distinguish the quality of one case series study from another.

Plan of review

The remainder of this review consists of three distinct parts. First, we have examined completed NICE TARs to see how previous reports have handled the inclusion of case series data (this chapter). Second, a search strategy was developed to identify existing methodological papers that empirically examined the effect of various design elements of case series studies on estimated treatment effect (Chapter 3). Finally, using published technology assessments that include both RCT and case series data, we will investigate the effect of selected elements of design (prospective data collection, consecutive recruitment, etc.) on estimates of treatment effect (Chapter 4).

Chapter 3

Review of the use of case series in NICE HTAs

Rationale

The review team were aware that information derived from case series studies has been included in NICE HTA reports. We wanted both to quantify this use and to examine the way in which data from this source have been used.

Methods

Search strategy

The list of current completed NICE HTAs was obtained from the NICE website on the 6 September 2002.

Inclusion and exclusion criteria

NICE TARs were assessed for the inclusion of case series evidence by one researcher (KD). Assessments that included diagnostic case series were excluded. Assessments were only included if the case series contributed to the evidence base for effectiveness or safety.

Data extraction/synthesis

One reviewer (KD) extracted the following data: title and author of assessment, publication date, assessment group, number of case series included, amount and type of other evidence included, criteria used for the inclusion of case series, methods used to assess the quality of case series, method used to synthesis results and conclusions drawn from case series evidence. Results are presented as a descriptive summary and are tabulated.

Results

The search identified 47 completed NICE HTA reports. Of the 47 assessments, 14 met the inclusion criteria by including case series evidence.

The characteristics of included studies are presented in *Table 3*. Reports were completed between June 1997 and June 2002. Nine of the included assessments were of pharmaceuticals, three of devices and two of surgical procedures. The 14 included assessment reports were produced by academic groups from the Universities of Birmingham, York, Aberdeen, Sheffield and Southampton. Two of the technology reports did not identify or include any randomised evidence for the assessment. The other 12 reports included between one and 70 RCTs as part of the evidence base. The number of case series included in the reports ranged from two to 159. Additional evidence was used in some assessment reports and consisted of systematic reviews, cohort studies, non-randomised comparative studies, case–control studies and surveillance studies.

All but one of the studies reported the reason for including case series evidence. The reasons were as follows: absence of RCTs/other evidence (five reports), requested by NICE (two reports), longer follow-up data (two reports), evidence on main outcome (two reports), safety data (two reports) or reason not stated (one report).

The reports applied a variety of case series inclusion criteria. Four studies were included purely on the basis of study size and two on the basis of length of follow-up, three reports did not state inclusion criteria, three stated that they did not apply inclusion criteria and two used a combination of criteria.

We also assessed whether reasons were given for the criteria chosen. For six studies this was not applicable as no criteria were specified. Other reasons for inclusion criteria included limiting workload, avoiding selected and unrepresentative samples, enabling the largest pool of studies to be included, as a result of reviewing survival analysis (to determine required length of follow-up), maximising generalisability and selecting the best available evidence. Three studies did not state their reason for the inclusion criteria.

Eleven of the included reports synthesised their results narratively and/or used tabulation. One study pooled results for use in the economic model sensitivity analysis, another provided medians and used regression to explore variation and another included studies in a meta-analysis. Most of the reports did not draw conclusions from the case series evidence or did so with caution (nine reports). One report used case series results

TABLE 3 *Characteristics of the included assessment reports (those including case series evidence)*

Author, completion date	Subject of review	Review group	No. of included RCTs	No. and type of other included evidence	No. of included case series	Reason for including case series	Criteria for case series inclusion	Reason for inclusion criteria	Methods used to synthesis case series results	Conclusions drawn from case series evidence
Hyde, Sept. 2001[40]	Fludarabine for B-cell chronic lymphocytic leukaemia	University of Birmingham	2	None	7	Absence of RCTs	>50 patients	To limit workload	Descriptive/ tabulation	Confirm that cautious interpretation of RCTs is appropriate
Forbes, July 2001[41]	Pegylated liposomal doxurubicin hydrochloride for ovarian cancer	University of York	2	None	6	Requested by NICE	None stated	Not applicable	Descriptive/ tabulation	None
Vale, June 2002[42]	Metal on metal hip resurfacing arthroplasty for hip disease	University of Aberdeen	1 (comparator)	Systematic review = 3 (comparator)	20 (5 intervention, 15 comparator)	Lack of RCT data	>2 year follow-up	Not stated	Descriptive. Pooled analysis for use in economic model sensitivity analysis	None
Peters, April 2002[43]	Inhaler devices for chronic asthma in older children	University of Sheffield	10	6 (other not stated)	14	Patient preference data	None stated	Not applicable	Descriptive	Provided information and recommenda-tions
Vardulaki, Dec. 2000[44]	Laparoscopic surgery for colorectal cancer	External (Royal College of Surgeons)	2	Prospective cohort = 7 Retrospective cohort = 5 Historically controlled cohort = 2	37	Long-term follow-up	>10 patients	Studies not meeting criteria likely to contain selected and unrepresenta-tive patients	Medians, descriptive, pooled estimates and regression to explore variation	Broad suggestions
Wake, March 2002[45]	Rituximab for refractory or recurrent stage III or IV follicular non-Hodgkin's lymphoma	University of Birmingham	0	None	4	Lack of comparative evidence	>10 patients	Not stated	Qualitative	None

continued

TABLE 3 *Characteristics of the included assessment reports (those including case series evidence) (cont'd)*

Author, completion date	Subject of review	Review group	No. of included RCTs	No. and type of other included evidence	No. of included case series	Reason for including case series	Criteria for case series inclusion	Reason for inclusion criteria	Methods used to synthesis case series results	Conclusions drawn from case series evidence
Ward, May 2001[46]	Gemcitabine for pancreatic cancer	University of Sheffield	7	7 (design not stated)	57 (phase II trials)	Not stated	None	Not applicable	Results tabulated	None
Dinnes, April 2001[47]	Temozolomide for recurrent malignant glioma	University of Southampton	1	None	6	Anticipated lack of data	>45 patients	To enable the pool of larger studies to be included	Narrative synthesis/ description	Speculation but not conclusions
Lewis, March 2002[48]	Trastuzumab for breast cancer	University of York	2	None	2	Requested by NICE	None stated	Not applicable	Narrative summary/ tabulation	Conclusions made while highlighting limitations
Fitzpatrick, June 1997[49]	Different prostheses for primary total hip replacement	Universities of Oxford and York	11	Non-randomised comparative = 18	159	Longer follow-up data	>5 years follow-up	Review of survival analysis confirmed that cut-off was conservative	Included in meta-analysis	Conclusions drawn with caution
Jobanputra, Dec. 2000[50]	Autologous chondrocyte transplantation for hyaline cartilage defects in knees	University of Birmingham	0	None	20	Only available evidence	None	Not applicable	Narrative summary	None
Bagnall, March 2002[51]	Atypical antipsychotics in schizophrenia	University of York	70 (new RCTs for update)	Cohort = 13 Case–control = 1	27	Safety data	>2 years follow-up or >2000 patients	Not stated	Narrative	None

continued

TABLE 3 *Characteristics of the included assessment reports (those including case series evidence) (cont'd)*

Author, completion date	Subject of review	Review group	No. of included RCTs	No. and type of other included evidence	No. of included case series	Reason for including case series	Criteria for case series inclusion	Reason for inclusion criteria	Methods used to synthesis case series results	Conclusions drawn from case series evidence
Woolacott, March 2002[52]	Bupropion SR and nicotine replacement therapy for smoking cessation	Universities of York and Birmingham	18	Systematic review = 2 Non-RCTs = 3 Uncontrolled = 19 Case–control = 1 Surveillance studies = 5	17 (case series or reports)	Safety data	None	Not applicable	Narrative	Unable to determine
Bryant, Oct. 2001[53]	Growth hormone in children	University of Southampton	21		11	Lack of data from RCTs on main outcome measure 'final height'	Excluded for Turners, renal failure and ISS. (idiopathic short stature). For GHD (growth hormone deficiency) included >300 patients. For PWS (Prader–Willi syndrome) included only case series study found	Best available evidence for final height measure. Cut-off at 300 patients for GHD to maximise generalisability	Narrative review and tabulation	Used to provide inputs for primary outcome into economic model. Conclusions drawn from model

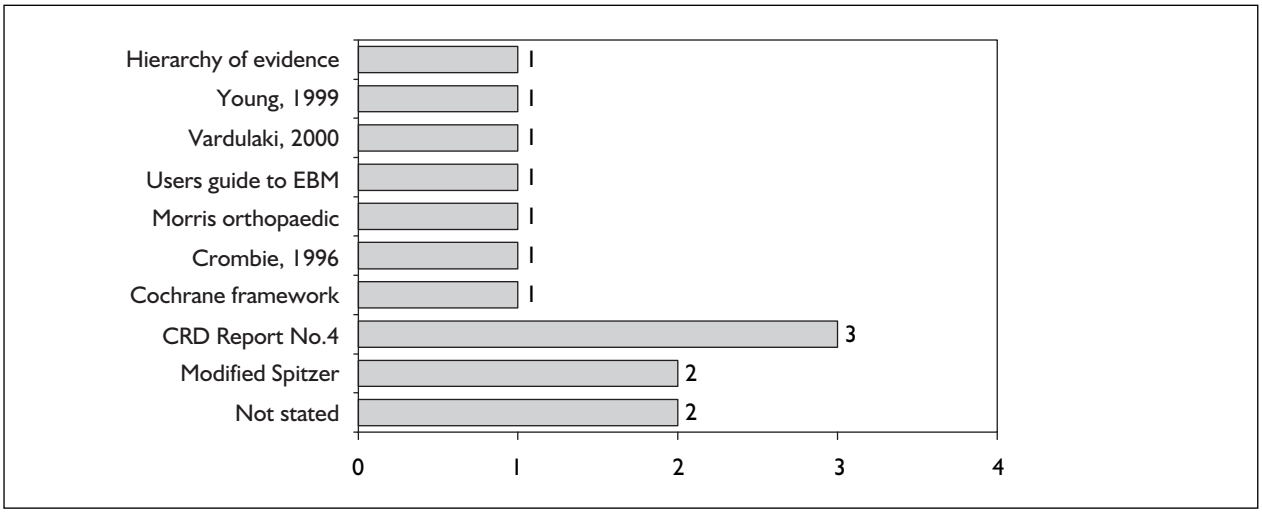

FIGURE 1 *Quality assessment criteria used*

to confirm results of RCTs, another used case series to provide information on patient preference, one made broad suggestions, another used case series evidence as inputs for the main outcome measure in the economic model and in a further report the conclusions drawn were unable to be determined.

A variety of methods were used in the reports to assess the quality of case series evidence. Ten different quality assessment tools or checklists were used in the 14 reports. Two reports did not state the source of their quality assessment items (*Figure 1*).

The only quality assessment items to be used in more than one report were the modified Spitzer checklist and the CRD Report No. 4. A total of 19 different quality assessment items were assessed across the 14 reports. The most commonly used quality items were as follows: a clear description of the included patients/cases (seven reports), description of loss to follow-up (seven reports), length of follow-up sufficient/described (eight reports) and valid, objective, masked outcome measurement (seven reports). There was a great deal of variation in the use of the other quality items – see *Table 4* for details of the items included in the quality assessment of case series. A tick (✓) indicates that an item was included in the quality checklist and a dash (–) indicates that it was not included or not stated. Some of the instruments include items such as the inclusion of a control group and the comparability of groups because the quality checklists and scales are not designed solely for case series, but can also be used with other types of study design.

Conclusions from the review of case series in HTAs considered by NICE

Despite being regarded as poor-quality evidence, case series studies have been included in almost 30% of completed NICE HTAs to date. In two cases, no RCT data were available for inclusion; the other HTAs also included RCT data from one to 70 RCTs. Case series data were used in various ways: to confirm the interpretation of limited RCT evidence, to inform an economic model, to draw conclusions and to make recommendations.

There is variation in the criteria for including case series (such as study size or length of follow-up) and also the methods used to assess the quality of the case series. It is likely that case series will continue to provide evidence in HTAs and it is therefore important to establish quality items that can be shown to impact on the reliability of case series findings.

The next chapter describes a search for methodological papers that attempt to empirically assess the effect of various case series design elements on estimated treatment effect.

TABLE 4 *Methods used to assess case series quality and a checklist of which quality criteria were assessed*

Author, completion date	**Hyde, Sept. 2001**[40]	**Forbes, July 2001**[41]	**Vale, June 2002**[42]	**Peters, April 2002**[43]	**Vardulaki, Dec. 2000**[44]	**Wake, March 2002**[45]	**Ward, May 2001**[46]	**Dinnes, April 2001**[47]	**Lewis, March 2002**[48]	**Fitzpatrick, June 1997**[49]	**Jobanputra, Dec. 2000**[50]	**Bagnall, March 2002**[51]	**Woolacott, March 2002**[52]	**Bryant Oct. 2001**[53]	**Total**
Methods used to assess quality	Cochrane framework	Crombie, 1996[54]	Morris *et al.*, 1988[55] orthopaedic checklist	Users Guide to EBM: Harm	Vardulaki instrument, 2000[44]	Young *et al.*, 1999[56]	Hierarchy of evidence	Modified Spitzer checklist	CRD Report No. 4	Not stated	Not stated	CRD Report No. 4	CRD Report No. 4 (cohort studies)	Modified Spitzer checklist	
Clear aims/question	–	✓	✓	–	✓	–	–	–	–	✓	–	–	–	–	4
Use of control group	–	✓	–	✓	–	–	–	–	–	–	–	–	–	–	2
Adequate study design	–	✓	✓	–	–	–	–	–	–	✓	–	–	–	–	3
Application of hierarchy of evidence	–	–	–	–	–	–	✓	–	–	–	–	–	–	–	1
Prospective enrolment	✓	–	–	–	✓	✓	–	–	–	–	–	–	–	–	3
Consecutive cases	✓	–	–	–	✓	✓	–	–	–	–	–	–	–	–	3
Appropriate sampling/ representative sample	–	–	–	–	–	–	–	✓	✓	–	–	✓	–	✓	4
Adequate sample size	–	✓	–	–	–	–	–	✓	–	–	–	–	–	✓	3
Explicit inclusion/ exclusion criteria	–	–	–	–	–	–	–	✓	✓	–	–	✓	–	✓	4
Patients entered study at similar point in disease	–	–	–	–	–	–	–	–	✓	–	–	✓	✓	–	3
Comparability of groups	–	–	–	✓	–	–	–	✓	–	–	–	–	✓	✓	4
Description of patients/ cases	✓	✓	✓	–	✓	✓	–	–	–	✓	–	–	✓	–	7
Description of intervention	–	–	✓	–	–	–	–	–	–	✓	–	–	✓	–	3
Treatment compliance	–	✓	–	–	–	–	–	–	–	–	–	–	–	–	1
Loss to follow-up	✓	✓	–	–	–	✓	–	✓	–	✓	–	–	✓	✓	7
Length of follow-up described/sufficient	✓	✓	✓	✓	–	–	–	–	✓	✓	–	✓	✓	–	8
Definition of outcomes	–	–	✓	–	✓	–	–	–	–	✓	–	–	–	–	3

continued

TABLE 4 *Methods used to assess case series quality and a checklist of which quality criteria were assessed (cont'd)*

Author, completion date	Hyde, Sept. 2001[40]	Forbes, July 2001[41]	Vale, June 2002[42]	Peters, April 2002[43]	Vardulaki, Dec. 2000[44]	Wake, March 2002[45]	Ward, May 2001[46]	Dinnes, April 2001[47]	Lewis, March 2002[48]	Fitzpatrick, June 1997[49]	Jobanputra, Dec. 2000[50]	Bagnall, March 2002[51]	Woolacott, March 2002[52]	Bryant Oct. 2001[53]	Total
Outcomes valid/ objective/ independent	–	✓	–	✓	–	–	–	✓	✓	–	–	✓	✓	✓	7
All relevant outcomes included (e.g. adverse effects)	–	✓	–	–	–	–	–	–	–	–	✓	–	–	–	2
Assessment pre- and post-intervention	–	–	–	–	–	–	–	–	–	–	✓	–	–	–	1
Patient input to outcomes	–	–	–	–	–	–	–	–	–	–	✓	–	–	–	1
Dose–response relationship/temporal relationship	–	–	–	✓	–	–	–	–	–	–	–	–	✓	–	2
Prognostic factors or confounders assessed/ analysis stratified	✓	✓	–	–	✓	–	–	–	✓	–	–	✓	✓	–	6
Appropriate statistical methods	–	✓	✓	✓	✓	–	–	–	–	✓	–	–	–	–	5
Appropriate interpretation of results	–	✓	–	–	–	–	–	–	–	–	–	–	–	–	1
Relevance/ generalisability assessed	✓	–	–	✓	–	–	–	✓	–	–	–	–	–	✓	4
Number of items	7	13	7	7	7	4	1	7	6	8	3	6	9	7	–

Chapter 4

Systematic review of methodological literature

Rationale

This report aimed to investigate whether or not there are characteristics of case series design that impact on the findings in a systematic way. In order to establish if this had previously been investigated in the literature, a systematic review was undertaken. This aimed to identify any studies which assessed the validity of case series in relation to the way in which they were designed, analysed or assessed for quality.

Method

Search strategy

A number of pilot search strategies were undertaken incorporating various search terms. Searching for case series studies is known to be challenging compared with searching for RCTs, which are typically well indexed. With case series it is difficult to balance inclusivity and specificity. We searched for existing methodological research on case series studies using the strategy outlined in Appendix 2. In addition to searching electronic databases, the contents pages of several journals were handsearched via their websites and full text papers obtained where appropriate (see Appendix 3).

A very large number of papers were initially found in MEDLINE with the original strategy, and specificity was extremely low. This strategy was therefore used only in methodological databases – Cochrane methodology database, the HTA National Research Register and the ESRC research register. In order to make the abstract scan manageable, a less broad search was devised and run through MEDLINE (see Appendix 2).

Inclusion and exclusion criteria

The inclusion process was performed by two independent researchers and differences were resolved by consensus. Searches were limited to English language only. Discussion articles, papers available only as abstracts and assessments of diagnostic tests were excluded. Included were primary studies which aimed to assess, in relation to the validity of studies:

1. aspects of case series design
2. aspects of case series analysis
3. aspects of case series quality.

In addition, we were looking for studies which compare the results of case series studies with those obtained through another study design.

Data extraction strategy

Data were to be extracted by one reviewer and checked by another. We planned to extract the following data from the studies: number of RCTs included in the study, number of case series studies included in the study, number and type of other study designs included, size of the studies, topic area, results from the case series, results from the RCTs, quality assessment methods used, methods of data comparison and main conclusions.

The results of the methodological studies were to be summarised and described with reference to case series studies.

Results

Initial broad searches in MEDLINE produced several thousand hits. Assessing the first hundred of these records identified no relevant references to acquire. We then ran this comprehensive first search through the Cochrane methodology database (as this is known to be produced from MEDLINE and handsearches of other sources), two research registers and also handsearched key journals. However, the team considered it important also to include a search in the key MEDLINE database. We searched MEDLINE with a less comprehensive search strategy, thereby sacrificing some inclusivity but producing a manageable number of hits that was realistic to examine. Details of both search strategies can be found in Appendix 2.

Number of studies identified

Search 1

- Cochrane methodology database: 112 hits, 11 papers obtained.
- HTA National Research Register: 13 hits, three papers obtained.

- ESRC research register (REGARD): 32 hits, one paper obtained.

Search 2

- MEDLINE: 768 hits.

When duplicates were excluded, a total of 914 potential articles were identified through the search strategies.

Handsearched journals

- *Social Science and Medicine* 1992;**34**(1) to 2003; **56**(4): no hits.
- *International Journal of Health Technology Assessment*: three hits, excluded at full text.
- *Journal of Epidemiology and Community Health* 1992;**46**(1) to 2003;**57**(2): three hits, excluded at full text.
- *American Journal of Epidemiology* 1992;**135**(1) to 2003;**157**(4): four hits, excluded at full text.
- *Controlled Clinical Trials* 1992;**13**(1) to 2003; **24**(1): four hits, excluded at full text.
- *Statistics in Medicine* 1996;**15**(1) to 2003;**22**(14): no hits.
- *Journal of the Royal Statistical Society, Series B* 1997;**59**(1) to 2003;**65**(3) and *Series C* 1997; **46**(1) to 2003;**52**(3): no hits.

Number and type of studies excluded, with reasons for specific exclusions

See Appendix 3 for details of studies identified and reasons for exclusion. Few studies were included at the abstract stage. We found no studies that specifically addressed the aims of this review. Most studies which looked at the results of non-randomised study designs compared with RCTs actually examined non-randomised but controlled studies, rather than case series. The papers that were included for background information fell into three broad categories: studies comparing the results of non-randomised but controlled observational studies and RCTs, studies which provided checklists or other quality criteria for non-RCTs, and discussion pieces. No studies were identified which empirically examined aspects of case series design, analysis or quality.

Although we were unable to identify any studies that addressed the central question of this project, there are limitations in the methods used. It is known to be difficult to design search strategies that comprehensively identify non-RCT study designs. In this case, due to time limitations, we elected not to use results from our most inclusive search strategy as it appeared from an initial examination of the results that specificity was extremely low. However, there may have been a few relevant papers from this search that were not identified through our other searches. In addition, we focused on trying to identify papers that had a methodological focus. It is possible that studies that were primarily reports of a case series study in a particular subject area in fact discussed methodology aspects of the study design and results. Such papers would not have been identified through our search strategies.

The next chapter describes an investigation into the elements of case series characteristics and the effect on study findings using four examples from published HTAs.

Chapter 5

Investigation of characteristics and findings of case series

Introduction

This section reports on the third part of our review: the empirical investigation of various aspects of case series study design and results using case examples from existing reports produced as part of the NICE HTA programme. We investigated whether particular aspects of study design and quality are associated with significant and systematic variation in results. We hoped that this could be used to inform reviewers who need to use case series studies about which methodological characteristics are likely to be of greatest significance and possibly to describe the direction of likely bias.

Hypotheses

The following hypotheses relating to the design of case series studies and the outcome frequency that they report were specified *a priori*:

1. Smaller sample sizes may be associated with more selection of those cases to include, or may show greater variation around a 'true' outcome frequency. We tested the null hypothesis that sample size has no systematic effect on outcome.
2. The outcome in case series could be better than in population-based or registry-based studies due to selection or to the healthy study effect. We tested the null hypothesis that there is no difference between these and case series studies.
3. The desirable outcome frequency in retrospective studies may be systematically better than in those which are prospective, due to selection. We tested the null hypothesis that there is no difference.
4. Multi-centre case series may show greater desirable outcome frequency than single-centre case series as there is more likely to be selection of cases. We tested the null hypothesis that there is no difference.
5. Prospective studies with consecutive enrolment may show greater desirable outcome frequency than those which employ non-consecutive enrolment. We tested the null hypothesis that there is no difference.
6. It is possible that studies where outcomes are not measured independently or blindly will show greater desirable outcome frequency than studies in which outcomes were assessed independently or in which there was outcome-assessor blinding. We postulate that lack of blind or objective outcome assessment has a systematic effect to increase outcome frequency. We tested the null hypothesis that there is no difference.
7. Length of follow-up is often related to estimate of outcome frequency, due to the natural history of the condition under study. We hypothesised that, for those outcomes that are not clearly related to length of follow-up, case series with short follow-up periods would systematically report better outcomes than those with longer follow-up periods. We tested the null hypothesis that there is no difference. It is possible that no effect would be discerned as loss to follow-up may bias the findings towards higher good outcome frequency. We planned to explore this where data were available.
8. We hypothesised that early reports of a new intervention would be likely to be published by enthusiasts and hence desirable outcome frequency would be higher in early reports of an intervention than in later reports. We tested the null hypothesis that there is no difference.
9. We tested the hypothesis that low-quality studies (as defined by a prespecified method) report higher desirable outcome frequency (null hypothesis of no effect).
10. Where possible, comparison between desirable outcome frequency estimated from case series and from RCTs was made.

Methods

Study identification

Initially, a list of projects undertaken for the UK HTA programme up to October 2002 was drawn up to show which had included case series

evidence. These assessments were found through our own project lists, published HTAs and contacting health technology assessment teams and are shown in *Table 5*. To be included, HTAs had to meet the following inclusion criteria, that had been specified *a priori*:

- at least 40 published case series studies available
- information on age of participants included (as a minimum description of the population)
- at least one good-quality controlled trial in the topic area.

Selection of topic areas

Few HTA reports both contained an RCT and had more than 40 case series studies available. We included in the study: the reviews of functional endoscopic sinus surgery (FESS) for nasal polyps, spinal cord stimulation (SCS) for chronic back pain and coronary artery bypass grafting (CABG) and percutaneous transluminal coronary angioplasty (PTCA) for chronic stable angina. The two treatments for angina were investigated separately. The HTA on hip replacement was initially considered for inclusion. However, closer examination of the studies included in this review revealed a very wide range of different interventions (hip prostheses). This suggested that any effects on outcome from methodological features may have been swamped by heterogeneity in the intervention. This concern and time constraints on the study led to a decision to restrict the investigation to only three topic areas.

For the studies on surgical treatments for angina and FESS for nasal polyps, the teams originally assessing these technologies had restricted the inclusion of case series studies by the number of participants included. In the case of angina studies this was for studies with <1000 participants and for FESS it was <50 patients with polyps. In order to examine the effect of sample size, these smaller studies were obtained. In the case of the FESS study, a full list of papers excluded owing to small study size was presented in the original report and this was used to obtain the relevant papers. In the case of angina, no list of excluded studies was provided. The search strategy described in the angina report was therefore re-run and the abstracts were examined to identify case series papers with <1000 participants with angina undergoing PTCA or CABG.

Two reviewers (RG and KS) examined the abstracts to decide which studies should be included. Studies were included if they used PTCA or CABG to treat angina, were case series studies, were published before 1999 (the date of the original report) and contained <1000 people.

Data extraction

Data were extracted from the published or unpublished reports of the included HTAs and from the original papers that had been excluded by these reports owing to the small sample size. Data were extracted by one reviewer (RG or KD) and checked by a second reviewer (KS or EC). The following data concerned with the study design and the sample achieved were extracted from each study within each HTA:

- study size
- prospective or retrospective design
- consecutive enrolment of patients or otherwise
- single-centre or multi-centre study
- date of publication
- length of follow-up
- independence or blinding of outcome assessment, that is, performed by someone not directly involved in applying the intervention or by using objective criteria [for FESS, symptomatic improvement; for SCS, pain relief; for angina, grading method for recurrent angina such as the New York Heart Association (NYHA) measure]
- age of participants
- proportion of study population who were male.

In addition, for each of the topic areas, information about the severity of the condition and other condition specific sample characteristics was also extracted:

1. FESS[57]
   (a) Percentage of sample with polyps (success of the procedure may vary among those with symptoms caused by polyps or other causes resulting in selection bias).
   (b) Percentage of the sample who had previously undergone surgery for the same condition (success rates may differ for first and subsequent surgical procedures resulting in selection bias).
2. SCS[58]
   (a) Absence of co-interventions.
   (b) Use of validated pain outcomes.
   (c) Mean pain duration.
   (d) Percentage of sample undergoing previous surgery.
   (e) Quality score for the study reports.
3. CABG and PTCA for stable angina[59]

TABLE 5 *HTAs including case series*

Subject	HTA group	No. of RCTs	Quality of RCTs	No. of case series included	Dates of publication	Case series criteria	Prospective	Consecutive	Range of sample sizes	Length of follow-up (months)	Independent outcome assessment	No. of studies with no loss to follow-up	No. of multi-centre studies
FESS	PenTAG	3	Poor	27[a]	1978 to 2001	>50 patients	7	6	50–1112 (median 130)	6–42 (median 17)	10	12	6
Temozolomide	Southampton	1	Poor	6	1996 to 2001	>45 patients	?	1	48–162 (median 89)	6–12 survival	5	2	3
Rituximab	Birmingham	0	NA	4	1998 to 2000	>10 patients	4	0	31–166 (median 95)	4–36	1	2	?
IFN for CML	PenTAG	8	Moderate/ poor	26	1987 to 2001	>20 patients	2	6	23–587 (median 81)	12–52 (median 42)	NA	10	?
Spinal cord stimulation CLBP/FBSS	Birmingham	1	Poor	76	1971 to 1998	None	8	11	1–250 (median 36)	1–120 (median 25)	6	35	7
Spinal cord stimulation CRPS	Birmingham	1	Good	21	1975 to 2001	None	3	2	1–189 (median 75)	1–87 (median 32)	3	15	1
ACI	Birmingham	0	NA	17	1998 to 2000	None	?	?	2–213 (median 25)	min 3–24 (median 12)	?	?	?
Intrathecal pumps	?	0	NA	53	1983 to 1999	None	?	?	1–429 (median 26)	?	?	?	?
Gemcitabine for pancreatic cancer	Sheffield	5	Poor	17	1991 to 2000	None	3 retrospective	?	10–3023 (median 35)	?	?	7	3
Hip replacement surgery	York/ Oxford	11	Poor	159 (15 assessed in detail)	?	5 years follow-up	?	?	?	?	NA	4	?
Chronic stable angina	Brunel/ York	3+	?	36[a]	1980 to 1986	>1000 patients	19	?	930–172,283 (median 2329)	?	?	?	10

ACI, autologous chondrocyte implantation; CLBP, chronic low back pain; CML, chronic myelogenous leukaemia; CRPS, complex regional pain syndrome; FBSS, failed back surgery syndrome; FESS, functional endoscopic sinus surgery; IFN, interferon; NA, not applicable.

[a] Although <40 case series were included, more than 40 were available, thus meeting the inclusion criteria.

(a) Percentage of the sample with unstable angina.
(b) Percentage of the sample with hypertension.
(c) Percentage of the sample with diabetes.
(d) Percentage of the sample with proximal left anterior descending coronary artery (LAD) stenosis.
(e) Percentage of the sample with left main artery disease.
(f) Percentage of the population with class 3 or 4 angina as measures on the NYHA scale.
(g) Mean ejection fraction.
(h) Percentage of the population with left ventricular dysfunction.

Finally, data on outcomes were extracted as follows:

1. FESS
   (a) Symptomatic improvement.
   (b) Polyp/disease recurrence.
   (c) Revision surgery.
   (d) Patency.
2. SCS
   (a) Percentage of patients with postoperative pain relief >50%.
   (b) Difference in mean pain scores on a visual analogue scale (VAS).
3. CABG and PTCA for angina
   (a) Percentage mortality.
   (b) Percentage of population experiencing recurrent angina postoperatively.
   (c) Survival at 5, 7 and 10 years taken from survival analysis.

Outcome measures were recorded as they were reported in the original papers and previous reviews as a mean result for the study as a whole. For the FESS studies, an alternative outcome measure using the whole cohort as a denominator was also calculated as an intention-to-treat (ITT) analysis. However, as there was large loss to follow-up in these trials, the ITT results are extremely conservative and may not be useful in this context.

Papers excluded – angina

We combined the results of a UK study reported at 1 year by Farrer and colleagues in 1997[60] and at 5 years by Skinner and colleagues in 1999[61] and these are reported under the Farrer study.

Papers for the PTCA and CABG for angina were excluded if they analysed only those in whom the operation was 'successful' or if they excluded those who died in hospital and did not report patient numbers. We also excluded studies that only reported outcome in hospital without longer term follow-up. A list of excluded papers is shown in Appendix 5.

Papers excluded – FESS

As there was a full list of papers excluded owing to small sample size in the Appendix of this report, we did not rerun the search. None of the papers identified in the Appendix were subsequently excluded for this study.

Papers excluded – spinal cord stimulation

All case series studies identified were included in this review, we did not rerun a search and all studies reported were included here.

Methods: analysis of study characteristics

The analysis of study characteristics was undertaken at a between-study level within each review. For each of the study hypotheses described in the section 'Hypotheses' (p. 25), where the potential explanatory variable is continuous a scatter plot was drawn and inspected and, if appropriate, a linear regression analysis was performed. Given the considerable heterogeneity in the data, robust regression was carried out using STATA version 8. The approach identifies single data points which have a particularly strong impact on the regression and sets these aside. The remaining data points are weighted according to the size of their residuals prior to an ordinary least-squares (OLS) regression being carried out. This method, although having lower statistical power in ideal circumstances, does not require the errors in the data to be normally, independently and identically distributed (normal i.i.d.). It is therefore a more general and flexible approach. Weighted regression analyses were also performed, weighted by sample size.

Where the potential explanatory variable was dichotomous, a box and whisker plot was drawn and a *t*-test and Mann–Whitney test were performed. Analysis of variance (ANOVA) was performed, weighted for sample size.

As variations in the population included in different studies may explain some of the differences in outcomes, data on possible explanatory study characteristics such as disease severity, mean age and proportion of the population that was male had been extracted. Multivariate analysis using these explanatory variables was performed using multivariate robust

regression or ANOVA analysis of covariance (ANCOVA) as appropriate.

A comparison between the effect on outcome shown in case series and the intervention arms of RCTs was carried out. First, multiple regression analyses including all relevant case series (robust and weighted) were carried out with the intervention as an explanatory factor. The coefficient in this analysis represents the effect of the intervention across case series. This was compared with the effect across intervention arms in RCTs using a meta-analysis, based on a random effects model. This comparison was carried out only for PTCA and CABG for angina as too few data points were available in the other data sets to permit meaningful analyses.

The outcome measures extracted from the studies were reported at different time periods, depending on the length of follow-up of the entire study. For relatively non-time-dependent outcomes (such as those with following FESS and SCS), the relationship between length of follow up and outcome was one of our hypotheses. For the angina outcomes, the natural history of the condition suggested that the outcome measure of mortality would worsen with time. Therefore, a yearly adjusted outcome measure was calculated by dividing the reported outcome by the average length of follow-up. Although this is likely to be an oversimplification of the true relationship between length of follow-up and mortality, this seemed to be a reasonable assumption. This method, as opposed to including length of follow-up in a multivariate analysis, was used because of the relatively small number of observations in the data set. For angina recurrence, the possible linear relationship was not as clear, hence both non-adjusted and adjusted outcome measures were calculated.

Results

Details of the number and type of studies that form the three data sets analysed in this section are shown in *Table 6*. The included papers were often not explicit about items such as whether the data were collected prospectively or consecutively. Where it was not possible to tell one way or another, these data were excluded from further analysis. In addition to RCTs and case series, other study designs were also examined – these included non-randomised comparative studies (using, for example, a historical control group or a group treated at the same institution) and case–control designs. Full details of the data extracted can be seen in Appendix 6. All multiple regression analyses were tested for homoscedasticity using the Cook–Weisberg test in STATA. None showed evidence of heteroscedasticity.

TABLE 6 *Summary of studies included in analysis*

	FESS	SCS	Angina PTCA	Angina CABG
Number of case series	42	76	63	72
Number of RCTs	3	1	10	10
Number of other designs	3 comparative	1 cohort	12 comparative 4 case–control	4 comparative 4 case–control
For case series:				
Median (range) in sample size	114 (5–1112)	36 (1–304)	166 (11–10785)	221 (10–172,283)
Number prospective	11	13	14	36
Number retrospective	19	38	15	28
Not clear whether retrospective or prospective	12	25	34	8
Number registry or population based	1	0	4	3
Number multi-centre	6	Not known	4	12
Number single-centre	36	1	55	53
Not clear if single- or multi-centre	0	75	4	7
Number consecutive enrolment	13	16	32	38
Number not consecutively enrolled	7	Not known	3	18
Not clear if consecutively enrolled	22	60	28	16
Median (range) in publication date	1994 (1978–2001)	1990 (1975–2001)	1991 (1982–1998)	1990 (1973–1998)
Number recording independent measure or blind outcome	15	43	24	23
Median (range) in length of follow-up (months)	17 (3–42)	24 (1–120)	27 (1–120)	48 (3–240)

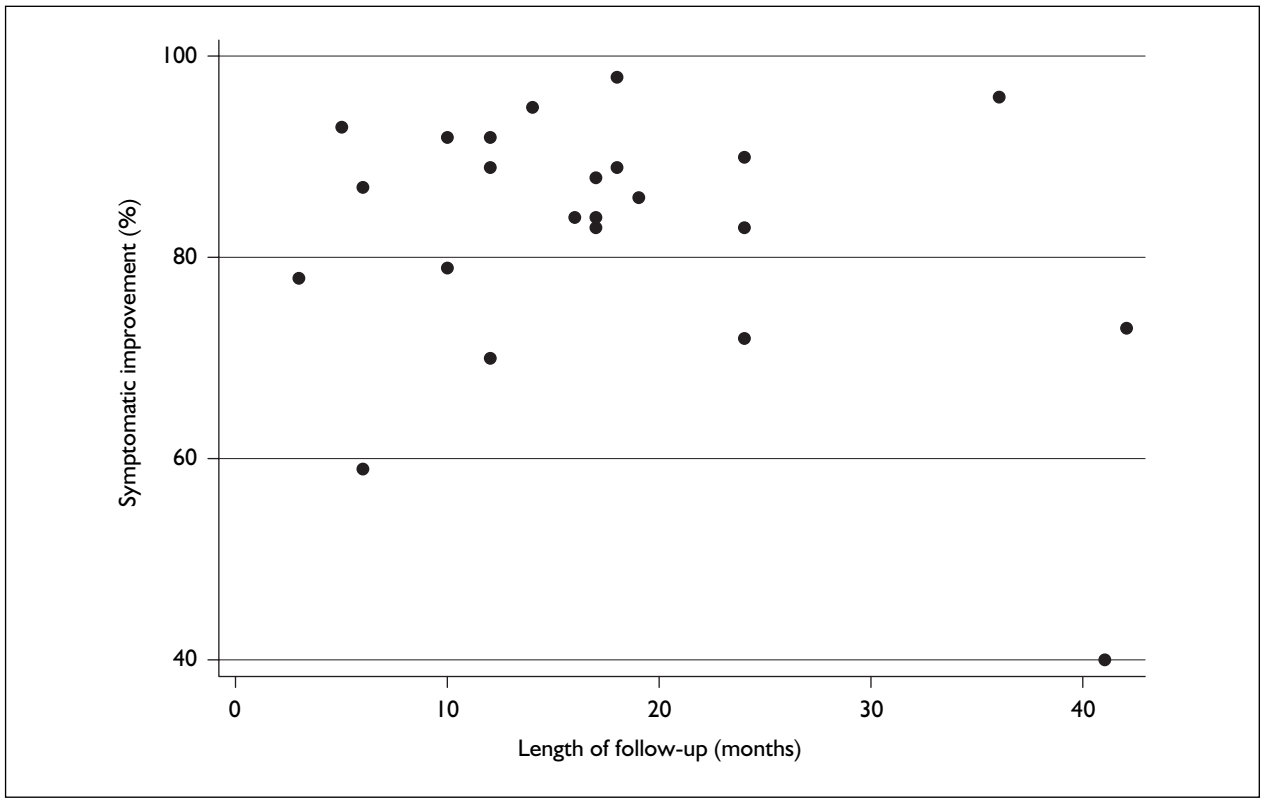

FIGURE 2 *Scatter plot of percentage symptomatic improvement after FESS and follow-up*

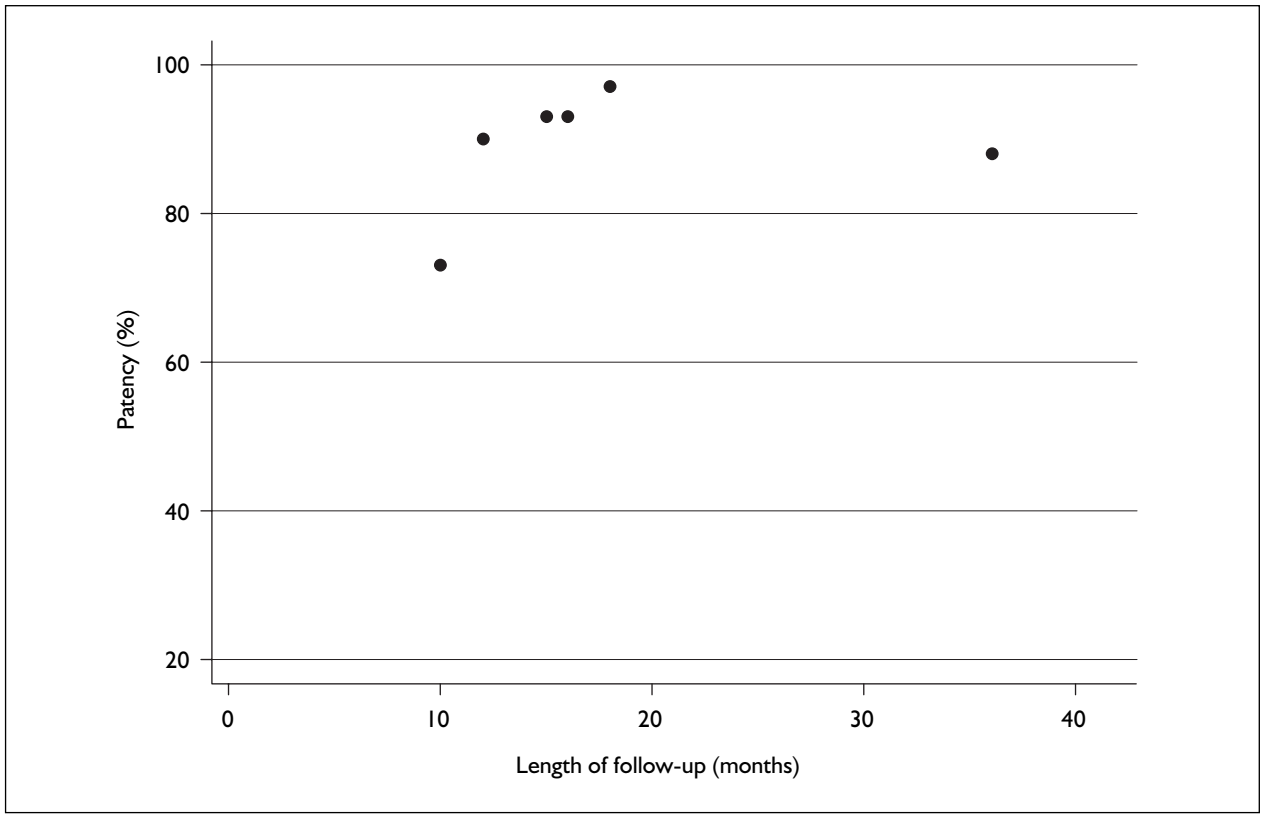

FIGURE 3 *Scatter plot of patency after FESS and length of follow-up*

Results for FESS

A total of 42 case series studies were available for analysis relating to FESS for the treatment of nasal polyps. Further details of the studies are shown in *Table 6*. The data extracted are presented in Appendix 6.

FESS – regression analysis of case series studies

Scatter plots were produced to investigate the relationship between outcomes and length of follow-up. These are shown in *Figures 2* and *3*. There is no clear linear relationship, particularly for patency, where few data points are present. Robust and non-robust regression was used to investigate the hypotheses outlined in the section Hypothesis and results are shown in *Table 7*. Where robust analysis is used, the number of studies excluded is stated. This number differs for each variable as different numbers of studies reported each variable and the outcome assessed.

TABLE 7 *Univariate analysis – investigation of hypotheses through robust regression for FESS studies*

	Outcomes	
Hypothesis	**Symptom improvement (%)**	**Patency**
Sample size has no effect on desirable outcome frequency	**No effect seen** Robust regression coefficient 0.006, $p = 0.61$ (95% CI –0.02 to 0.03) $n = 29$ excluded: Danielsen and Oloffsson (230 patients)[a]	**No effect seen** Robust regression coefficient 0.03, $p = 0.6$ (95% CI –0.12 to 0.17) $n = 8$, no studies excluded
Prospective shows lower desirable outcome frequency than retrospective	**No effect seen** Mean 83% for both groups Mann–Whitney $p = 0.23$ $n = 21$	**No effect seen** Mean 94% prospective vs 86% retrospective Mann–Whitney $p = 0.2$ $n = 7$
Case series show higher desirable outcome frequency than registry- or population-based studies	**No data for comparison**	**No data for comparison**
Multi-centre studies show higher desirable outcome frequency than single-centre studies	**Significantly higher results in multi-centre trials** Mean 92% vs 81% Mann–Whitney $p = 0.02$	**Significantly higher results in multi-centre trials** 95% vs 70% Mann–Whitney $p = 0.03$
Prospective studies with consecutive enrolment show lower desirable outcome frequency than those with non-consecutive enrolment	**No effect seen** 87% prospective and 85% retrospective Mann–Whitney $p = 0.24$ $n = 15$	**No data for comparison**
Studies with independent or blinded measurement of outcome will show lower desirable outcome frequency than those without such features	**No effect seen** 82% blinded/independent measurement, 78% not. Mann–Whitney $p = 0.24$ $n = 19$	**No effect seen** Mean 89% vs 64% Mann–Whitney $p = 0.15$ $n = 8$
Length of follow-up will be negatively associated with desirable outcome frequency	**No effect seen** Conflicting results Weighted regression coefficient –0.58, $p = 0.040$ (95% CI –1.14 to –0.03) Robust regression coefficient –0.004, $p = 0.986$ (95% CI –0.49 to 0.48) $n = 19$, no study excluded	**No effect seen** Weighted regression coefficient 0.04, $p = 0.91$ (95% CI –0.92 to 1) Robust regression coefficient 2.61, $p = 0.08$ (95% CI –0.5 to 5.8) $n = 5$, no study excluded
The date of publication will be negatively associated with desirable outcome frequency	**No effect seen** Weighted regression coefficient = 0.35, $p = 0.42$ (95% CI –0.54 to 1.25) Robust regression coefficient 0.46, $p = 0.225$ (95% CI –0.3 to 1.23) $n = 29$, excluded Danielson and Olofsson (230 patients)[a]	**No effect seen** Weighted regression coefficient –0.1, $p = 0.92$ (95% CI –2.6 to 2.4) Robust regression coefficient 0.005, $p = 0.99$ (95% CI –4.7 to 4.7) $n = 8$, no study excluded

[a] Here and in subsequent tables and text, for details of studies see Appendix 6.

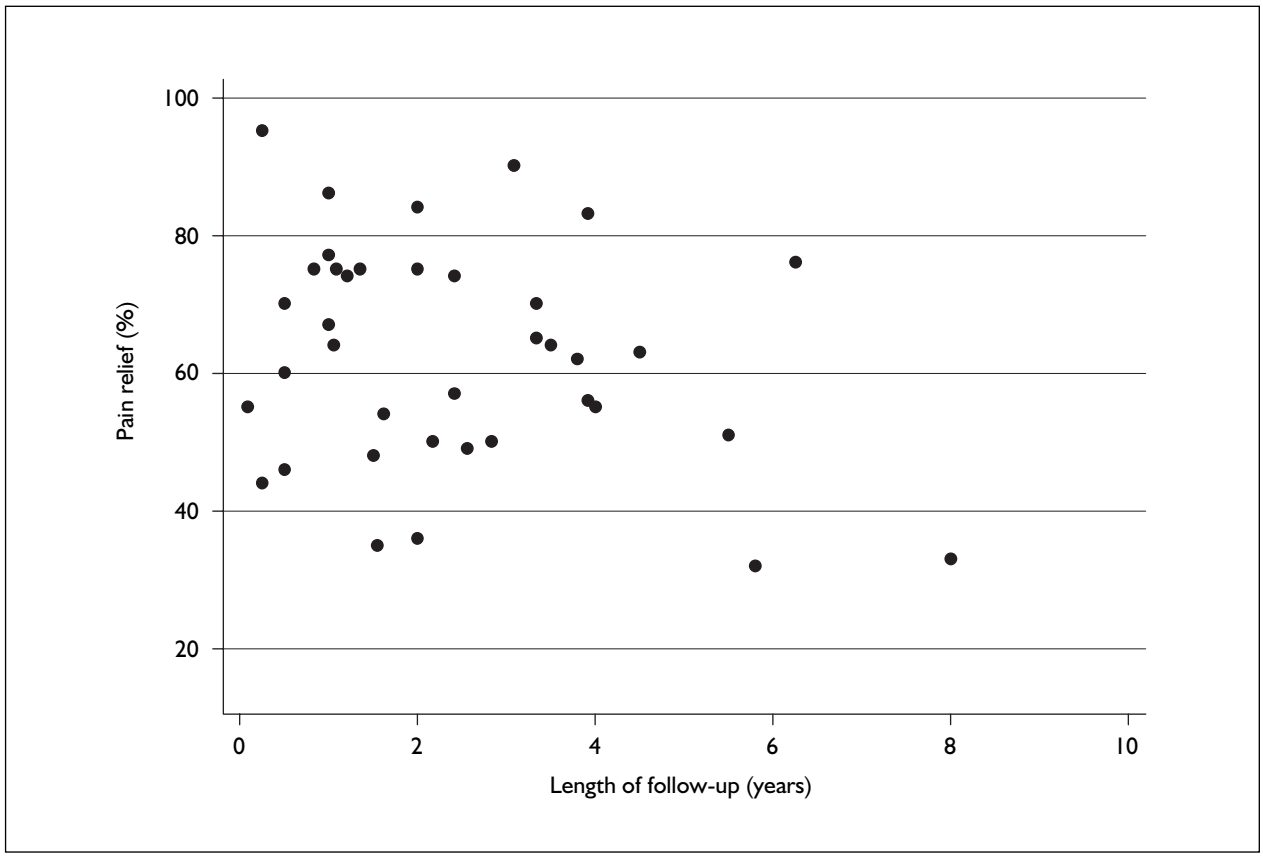

FIGURE 4 *Scatter plot of percentage pain relief after SCS and length of follow-up*

For most variables, no effect was seen with robust and non-weighted regression. Length of follow-up appeared to be negatively associated with symptomatic improvement in a weighted, but not a robust regression. Insufficient data were available for comparisons using patency and consecutively enrolled prospective studies and for both outcomes and registry-based studies. Multi-centre design showed a significant effect and this was found for both outcomes ($p = 0.02$; $p = 0.03$).

FESS – multivariate analysis of case series studies

A negative regression coefficient was found between age and symptom improvement and patency, but this was non-significant. A positive regression coefficient was found between these outcomes and the percentage of males in the sample; this was non-significant for patency, but significant for symptom improvement, $b = 0.68$ ($p = 0.003$). The addition of age and sex to the univariate analyses performed above did not change any of the findings with the exception of multi-centre study design, which remained significant for patency ($p = 0.05$) but not for symptomatic improvement.

Results for SCS

There were 75 case series studies of SCS for chronic back pain available for analysis. See *Table 6* for further details.

SCS – regression analysis of case series studies

Reported proportion of people achieving >50% postoperative pain relief plotted against average length of follow-up in years was explored in a scatter plot (*Figure 4*). Regression analysis showed a non-significant small negative coefficient for length of follow-up (regression coefficient –0.0021, $p = 0.126$, 95% CI –0.005 to 0.0013) (*Table 8*).

There were insufficient data to undertake the analysis for the VAS outcome difference in pre- and postoperative pain. As there were no registry- or population-based studies identified, it was not possible to explore this hypothesis. Insufficient data also prevented the exploration of the effect of use of blinded or independent measures of outcome, multi-centre versus single-centre trials and the effect of consecutive enrolment.

The quality score used in the SCS review was derived from the Jadad score for RCTs[31] whereas the quality of case series was designed for this

TABLE 8 *Univariate analysis – investigation of hypotheses through robust regression analysis using SCS studies*

Hypothesis	Pain relief >50%
Sample size has no effect on desirable outcome frequency	**No effect seen** Robust regression coefficient –0.00013, $p = 0.845$ (95% CI –0.0014 to 0.0012) $n = 56$, no studies excluded Weighted regression 0.0005, $p = 0.22$ (95% CI –0.0003 to 0.0013)
Prospective shows lower desirable outcome frequency than retrospective	**No effect seen** Mean 65% prospective and 62% retrospective Mann–Whitney $p = 0.81$ $n = 34$
Case series show higher desirable outcome frequency than registry- or population-based studies	**Insufficient data**
Multi-centre studies show higher desirable outcome frequency than single-centre studies	**Insufficient data**
Prospective studies with consecutive enrolment show lower desirable outcome frequency than those with non-consecutive enrolment	**Insufficient data**
Studies with independent or blinded measurement of outcome will show lower desirable outcome frequency than those without such features	**Insufficient data**
Length of follow-up will be negatively associated with desirable outcome frequency	**No effect seen** Robust regression coefficient –0.0021, $p = 0.126$ (95% CI –0.005 to 0.0006) $n = 38$, no studies excluded Weighted regression –0.00075, $p = 0.475$ (95% CI –0.003 to 0.0013)
The date of publication will be negatively associated with desirable outcome frequency	**No effect seen** Weighted regression coefficient 0.002, $p = 0.651$ (95% CI –0.005 to 0.008) Robust regression coefficient –0.002, $p = 0.51$ (95% CI –0.009 to 0.005) $n = 57$, no studies excluded
Quality score will show a negative association with desirable outcome frequency	**Negative correlation** Weighted regression coefficient –0.06, $p = 0.002$ Robust regression coefficient –0.053, $p = 0.04$ Weighted ANOVA, $p = 0.038$

assessment by the authors and was based on previously published checklists including items on selection bias, attrition bias, performance bias and detection bias. It included some of the items that we investigated (consecutive recruitment, prospective recruitment, independent/blind outcome assessment) and others not included here (see Appendix 6 for details). A significant negative correlation was seen between the study quality score and desirable outcome frequency (i.e. poorer quality studies overestimated treatment effect compared with higher quality studies).

SCS – multivariate analysis of case series studies

No significant associations with any of the possible explanatory or confounding variables were identified – age, proportion male, duration of pain, number of previous operations (*Table 9*). Inclusion of age and sex in the multivariable analysis did not alter the findings.

Results for CABG and PTCA for angina

Unlike the results for FESS and SCS, which reported outcomes relating to treatment success, the reported outcomes in the angina studies were undesirable (mortality and recurrence of angina). Insufficient studies reported on 5-year or longer survival to perform statistical analyses.

A total of 72 case series studies were identified for CABG for angina and 63 case series for PTCA for angina. The quality of the RCTs identified is not reported in the original review.

TABLE 9 *SCS – robust regression of sample characteristics and pain relief*

	Robust regression coefficient	95% CI	*p*-Value	No. of studies on which analysis based
Age	–0.0203	–0.046 to 0.005	0.110	29, no studies excluded
Proportion of males	0.3055	–0.238 to 0.84	0.258	38, no studies excluded
Duration of pain	–0.042	–0.087 to 0.003	0.067	21, no studies excluded
No. of previous operations	–0.07	–0.23 to 0.09	0.35	15, no studies excluded

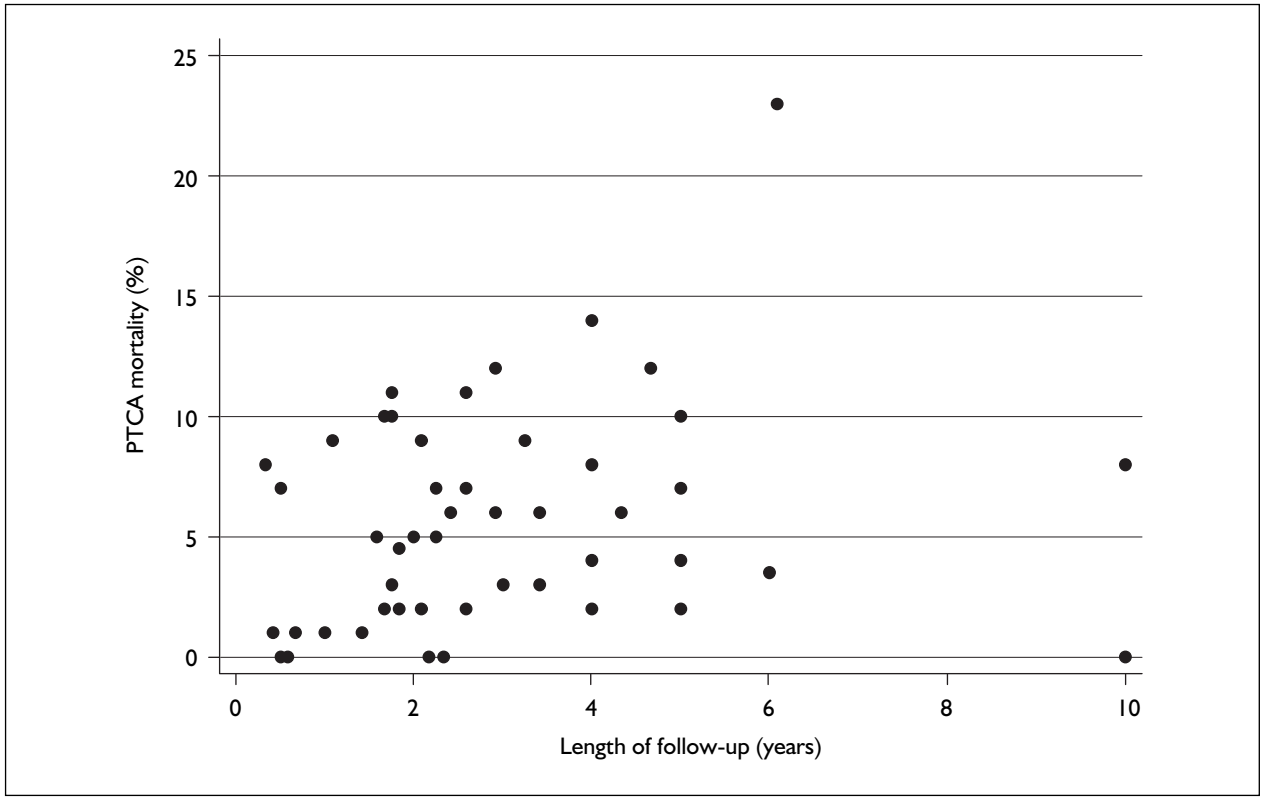

FIGURE 5 *Scatter plot of mortality after PCTA for angina and length of follow-up*

PTCA – regression analysis of case series studies

A scatter plot showing reported mortality (proportion) against length of follow-up in years was plotted for PCTA and is shown in *Figure 5*. There was a positive regression coefficient (0.007, $p = 0.03$). Adjusted yearly mortality was therefore used in the analysis.

A scatter plot of recurrent angina (proportion) and length of follow-up showed no linear relationship (*Figure 6*). The natural history of the condition suggests that angina recurrence should increase rather than decrease over time, so this finding is surprising. Loss to follow-up, including deaths, may explain the apparent lack of relationship, although the inclusion of mortality in a multivariate analysis did not demonstrate that this factor had a significant confounding effect. This is explored further in comparison with RCTs in the section 'Comparison of case series results with RCT results' (p. 44). We performed analysis of angina results both adjusted for length of follow-up and non-adjusted.

Table 10 shows the results of the robust regression analysis for case series studies of PTCA for the treatment of angina. The table notes the number of studies excluded through using robust regression for each investigated study characteristic. This number may be different for the same hypothesis investigated for different outcomes as different numbers of studies report each outcome and methodological characteristics that we are investigating (for example, length of follow-up). For continuous variables (sample size, length of follow-up, date of publication) robust

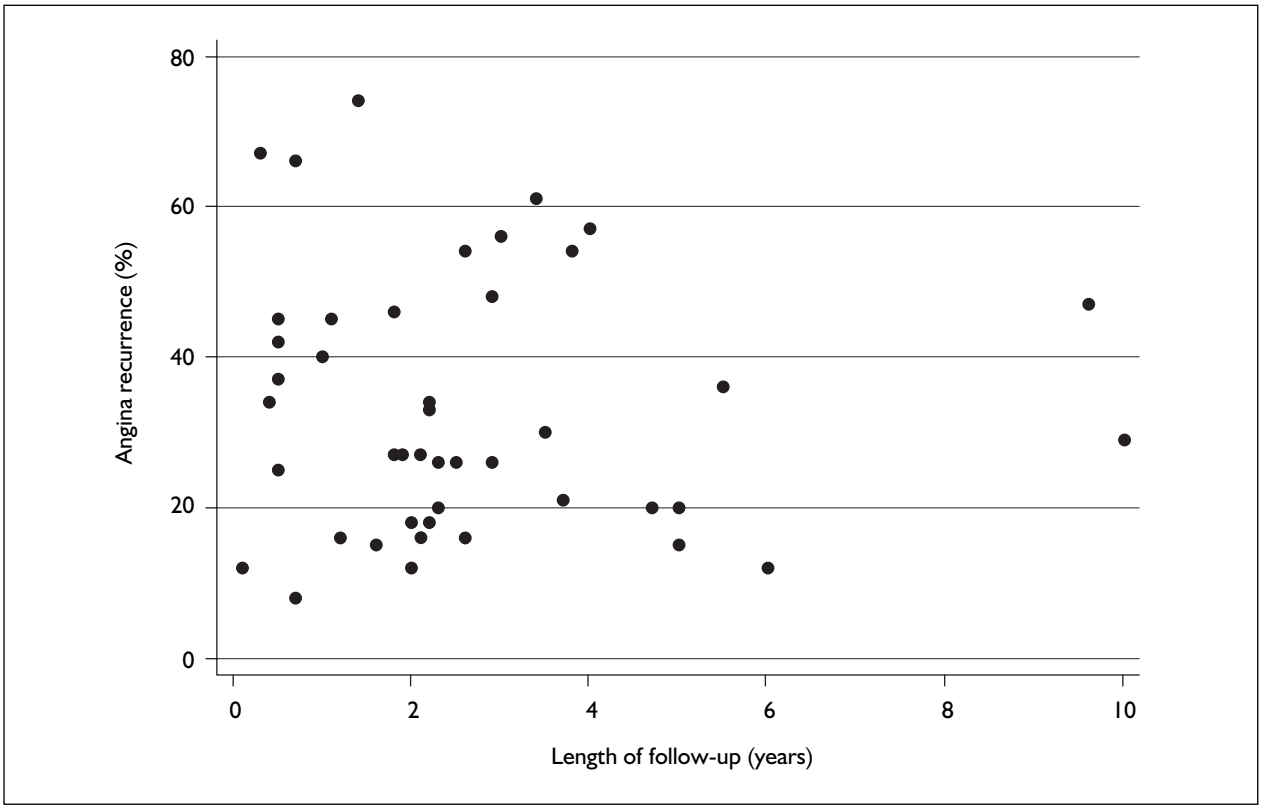

FIGURE 6 *Scatter plot of angina recurrence after PTCA and length of follow-up*

regression coefficients are reported. For dichotomous variables (prospective data collection, registry-based data collection, use of independent outcome measurement), means and medians are reported.

Due to the nature of the outcome, it is not appropriate to assess the effect of independent or blinded measurement of mortality and this analysis has not been undertaken. Only three studies involved multi-centre data collection and comparative statistics were not calculated owing to insufficient data. As only a few studies collected data from a disease register, there were insufficient data to analyse the possible effect of this characteristic. For most other study characteristics, no effect was seen in the level of reported outcome and this was true for angina outcomes whether adjusted or non-adjusted data were examined. However, a significantly higher rate of recurrent angina was seen in the non-adjusted analysis for studies that measured this outcome independently compared to those that did not (weighted ANOVA, $p = 0.005$). When the data were adjusted for length of follow-up, no effect was noted for this variable. Publication date was negatively associated with undesirable outcome frequency for adjusted angina recurrence (i.e. earlier publication date showed less favourable results, not more favourable results as we had hypothesised). For further details, see *Table 10*.

PTCA – multivariate analysis of case series studies

Population characteristics that may act as confounders and for which data were available are shown in *Table 11*. The proportion of patients with single-vessel disease in the studies exhibited a trimodal distribution, because number of diseased vessels was an inclusion criterion in some studies. This variable was excluded from further analysis. Sample age, the proportion of male patients and the proportion of patients with more severe angina, as measured by the NYHA criteria, were the variables for which most data were available and these were therefore used in the multivariate analysis.

PTCA – multivariate analysis of case series studies – mortality

The multivariate analyses included age, proportion of males and proportion of people in NYHA grade 3 or 4. These are all significant or marginally significant in univariate robust regression shown in *Table 12*. The data show the expected effects of these variables with greater age, female sex and more severe angina related to

TABLE 10 *Univariate analysis – investigation of hypotheses through regression of PCTA outcomes and study characteristics*

Hypothesis	Yearly adjusted mortality	Yearly adjusted angina recurrence	Unadjusted angina recurrence
Sample size has no effect on undesirable outcome frequency	**No effect noted** Robust regression coefficient <–0.0000, $p = 0.60$ $n = 49$, 1 study excluded from regression: Safian and Urban	**No effect noted** Robust regression coefficient <–0.0000, $p = 0.54$ $n = 48$, 8 studies excluded from regression Anderson and Ward, Myler, Mata, Melchior, Krajcer, Urban, Gaylani, Safian	**No effect noted** Robust regression coefficient <–0.0000, $p = 0.85$ $n = 50$, no studies excluded from regression
Prospective studies show greater undesirable outcome frequency than retrospective studies	**No effect noted** 12 studies reported retrospective and 11 prospective measure of outcome Median (0.01 vs 0.02) and mean (0.02 vs 0.04) both lower in prospective studies *t*-Test mean difference = 0.02 (SE 0.02), $p = 0.29$ Mann–Whitney test, $p = 0.44$ Weighted ANOVA, $p = 0.44$	**No effect noted** 11 studies reported retrospective and 10 prospective measure of outcome Median (0.11 vs 0.11) the same in both studies, mean (0.37 vs 0.12) lower in prospective studies. *t*-Test mean difference = 0.25 (SE 0.2), $p = 0.19$ Mann–Whitney test, $p = 0.57$ Weighted ANOVA, $p = 0.85$	**No effect noted** 11 studies reported retrospective and 10 prospective measure of outcome Median (0.24 vs 0.26) and mean (0.28 vs 0.32) both lower in prospective studies. *t*-Test mean difference = 0.04 (SE 0.07), $p = 0.62$ Mann–Whitney test, $p = 0.44$ Weighted ANOVA, $p = 0.15$
Case series show lower undesirable outcome frequency than registry- or population-based studies	**Insufficient data for analysis**	**Insufficient data for analysis**	**Insufficient data for analysis**
Multi-centre studies show lower undesirable outcome frequency than single-centre studies	**Comparative statistics not calculated** 45 centres reported single-centre and 3 reported multi-centre enrolment Median higher in multi-centre studies (0.02 vs 0.15) but mean lower in multi-centre studies (0.03 vs 0.02)	**Comparative statistics not calculated** 42 centres reported single-centre and 3 reported multi-centre enrolment Median lower in multi-centre studies (0.10 vs 0.12) and mean lower in multi-centre studies (0.19 vs 0.27)	**Comparative statistics not calculated** 44 centres reported single-centre and 3 reported multi-centre enrolment Median lower in multi-centre studies (0.26 vs 0.3) and mean lower in multi-centre studies (0.28 vs 0.34)
Prospective studies with consecutive enrolment show higher undesirable outcome frequency than those with non-consecutive enrolment	**Comparative statistics not calculated** 27 studies stated consecutive and 2 stated non-consecutive enrolment	**Comparative statistics not calculated** 24 studies stated consecutive and 2 stated non-consecutive enrolment	**Comparative statistics not calculated** 26 studies stated consecutive and 2 stated non-consecutive enrolment

continued

TABLE 10 *Univariate analysis – investigation of hypotheses through regression of PCTA outcomes and study characteristics (cont'd)*

Hypothesis	Yearly adjusted mortality	Yearly adjusted angina recurrence	Unadjusted angina recurrence
Studies with independent or blinded measurement of outcome will show higher undesirable outcome frequency than those without such features (for recurrent angina this refers to graded levels of angina using a recognised scale such as the NYHA)	**Not applicable for mortality outcome**	**No effect noted** 41 studies reported whether there was objective measure of outcome – 20 had no objective measure and 21 had objective measure *t*-Test mean difference = 0.2 (SE 0.13), $p = 0.15$ Mann–Whitney test, $p = 0.08$ Weighted ANOVA, $p = 0.25$	**Significantly higher reported rate of angina recurrence in studies that independently measured outcomes** 41 studies reported whether there was objective measure of outcome – 20 had no objective measure and 21 had objective measure *t*-Test mean difference = 0.17 (SE 0.05), $p = 0.0007$ Mann–Whitney test, $p = 0.001$ Weighted ANOVA, $p = 0.005$
Length of follow-up will be positively associated with undesirable outcome frequency	**Significant positive association** Robust regression coefficient 0.007, $p = 0.03$, (95% CI 0.0008 to 0.01) $n = 51$, no studies excluded Weighted regression coefficient = 0.007, $p = 0.04$, (95% CI 0.004 to 0.01)	**Not applicable** Data adjusted for length of follow-up	**No effect seen** Robust regression coefficient –0.02, $p = 0.08$, (95% CI –0.05 to 0.003) $n = 48$, no studies excluded Weighted regression coefficient = –0.004, $p = 0.78$, (95% CI –0.03 to 0.02)
The date of publication will be positively associated with undesirable outcome frequency	**No effect seen** Robust regression coefficient 0.001, $p = 0.16$ (95% CI –0.0004 to 0.002) $n = 49$, 2 studies excluded from regression: Safian, Urban Weighted regression = 0.015, $p = 0.21$ (95% CI –0.0009 to 0.004)	**Significant result in weighted analysis** Robust regression coefficient –0.004, $p = 0.27$ (95% CI –0.01 to 0.003) $n = 47$, 7 studies excluded from regression: Safian, Mata, Gaylani, Urban, Krajcer, Anderson, Melchior) Weighted regression –0.02, $p = 0.005$ (95% CI –0.037 to –0.007)	**No effect seen** Robust regression coefficient –0.005, $p = 0.42$, (95% CI –0.02 to 0.004) $n = 49$, no studies excluded from regression Weighted regression coefficient –0.005, $p = 0.53$ (95% CI –0.01 to 0.008)

n, number of studies included in analysis; SE, standard error.

TABLE 11 *Sample characteristics of PTCA for angina case series*

	Range	Median	Mean (SD)	*n*
Mean age (years)	46	58	56.3 (9.3)	55
Proportion male	0.42	0.77	0.77 (0.08)	61
Proportion left ventricular dysfunction	0.34	0.32	0.28 (0.15)	6
Mean ejection fraction	0.21	0.58	0.58 (0.05)	17
Proportion with single-vessel disease	1	0.53	0.49 (0.31)	50
Proportion of patients in NYHA grade 3 or 4	0.74	0.62	0.61 (0.18)	35
Proportion with left main stem disease	0.02	0.01	0.02 (0.01)	5

TABLE 12 *PTCA – robust regression of sample characteristics and adjusted mortality*

	Robust regression coefficient	95% CI	*p*-Value	No. of studies on which analysis based
Age	0.0007	0.0002 to –0.001	0.01	42 (excluded: Safian, Urban)
Proportion male	–0.12	–0.15 to –0.08	0.000	47 (excluded: Urban, Safian, Maiello)
Proportion of patients in NYHA grade 3 or 4	0.03	–0.005 to 0.06	0.09	27 (excluded: Safian)

TABLE 13 *PTCA – robust regression of sample characteristics and angina recurrence (non-adjusted)*

	Regression coefficient	95% CI	*p*-Value	No. of studies included in analysis
Age	0.003	–0.003 to 0.008	0.37	44, no studies excluded
Proportion male	0.35	–0.28 to 0.97	0.28	49, no studies excluded
Proportion of patients in NYHA grade 3 or 4	–0.17	–0.58 to 0.24	0.4	28, no studies excluded

worse outcomes for angina treatment. However, the effect of more severe angina only approaches significance.

Using unadjusted mortality as the outcome, multivariate analysis found length of follow-up to have a significant positive coefficient of 0.01 (95% CI 0.0007 to 0.02), $p = 0.04$. This analysis was based on 24 studies. This shows little change from univariate regression (see *Table 10*). However, robust regression showed no effect with adjusted mortality [–0.001; 95% CI 0.255 to –0.002; $p = 0.255$, two studies excluded (Safian and Urban)].

Sample size, publication date and objective measurement of outcome continue to have no effect on outcome, as frequently seen in the univariate analysis shown in *Table 10* (based on 24, 24 and 20 studies, respectively).

It was not possible to investigate prospective, consecutive enrolment and multi-centre vs single-centre enrolment in multivariate analysis as all had insufficient data for analysis.

PTCA – multivariate analysis of case series studies – angina recurrence

Potential explanatory sample variables for the non-adjusted rate of angina recurrence were examined through robust regression analysis as shown in *Table 13* and adjusted for length of follow-up (*Table 14*). None were found to be significant in either type of analysis. The direction of association of severity of angina with recurrence is counter-intuitive. However, this may be due to a healthy survivor effect where those with more severe angina are more likely to die, leaving the proportion alive with recurrent angina reduced. Length of follow-up showed a significant negative coefficient based on 19 studies in non-adjusted

TABLE 14 *PTCA – robust regression of sample characteristics and adjusted angina recurrence*

	Regression coefficient	95% CI	*p*-Value	No. of studies included in analysis
Age	0.002	–0.0005 to 0.003	0.13	42,excluded: Urban, Krajcer, Myler, Melchior, Mata, Gaylani, Safian, Anderson
Proportion male	–0.15	–0.39 to 0.08	0.2	47, excluded: Holmes, Myler, Anderson, Krajcer, Urban, Mata, Safian, Melchior
Proportion of patients in NYHA grade 3 or 4	–0.025	–0.17 to 0.12	0.72	27, excluded: Anderson, Safian, Urban, Myler, Melchior, Krajcer

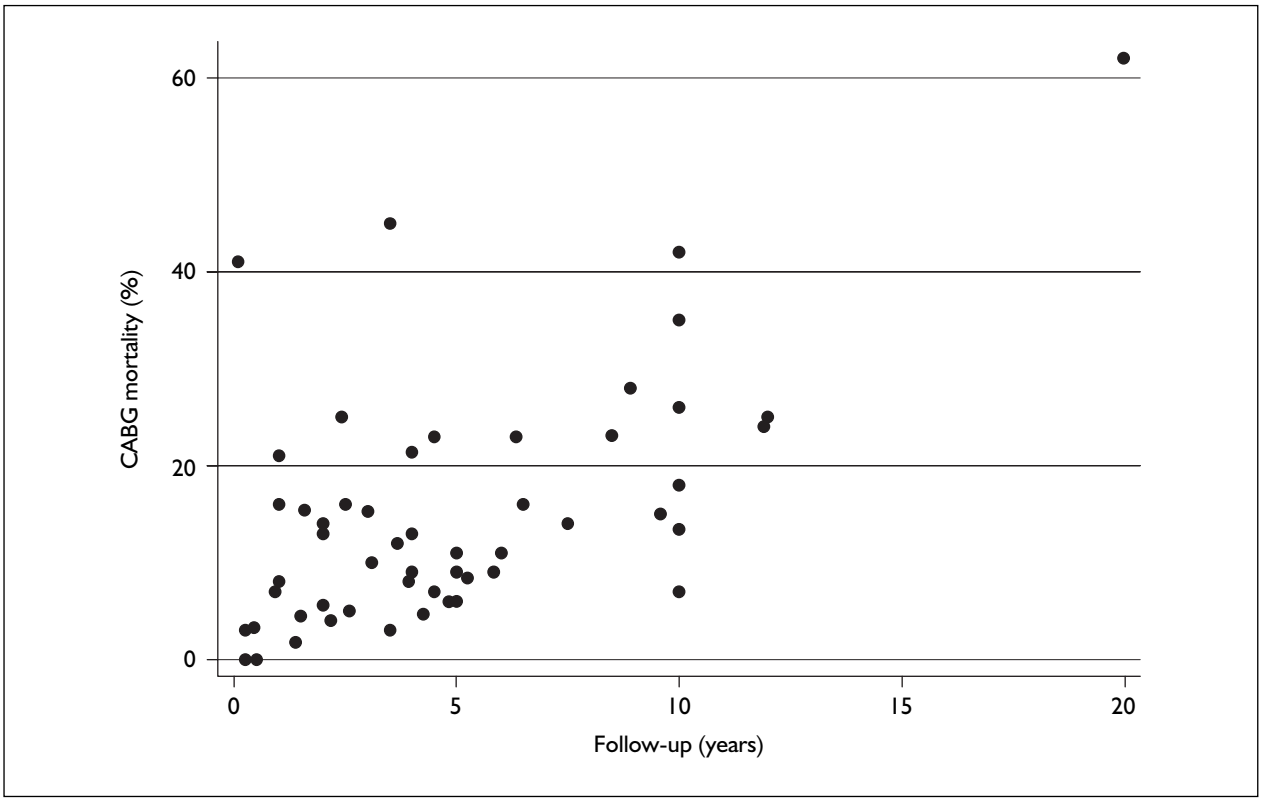

FIGURE 7 *Scatter plot of mortality after CABG for angina and length of follow-up*

analysis (negative coefficient –0.03; 95% CI –0.05 to –0.01; $p = 0.004$). In addition, the effect of male sex is different with adjusted and non-adjusted analyses.

Multivariate analysis which included age, proportion male and proportion with NYHA grade 3 or 4 angina showed that sample size continued to have no effect (based on 23 studies) and publication date had no effect (based on 22 studies).

Insufficient data were available about other variables to allow analyse any of the other hypotheses. See *Table 10* for comparison with the results of univariate analysis.

CABG – regression analysis of case series studies

A scatter plot of mortality proportion against length of follow-up in years for CABG is shown in *Figure 7*. There was a positive regression coefficient (0.02, $p = 0.000$) and adjusted yearly mortality was therefore used in the analyses. This meant

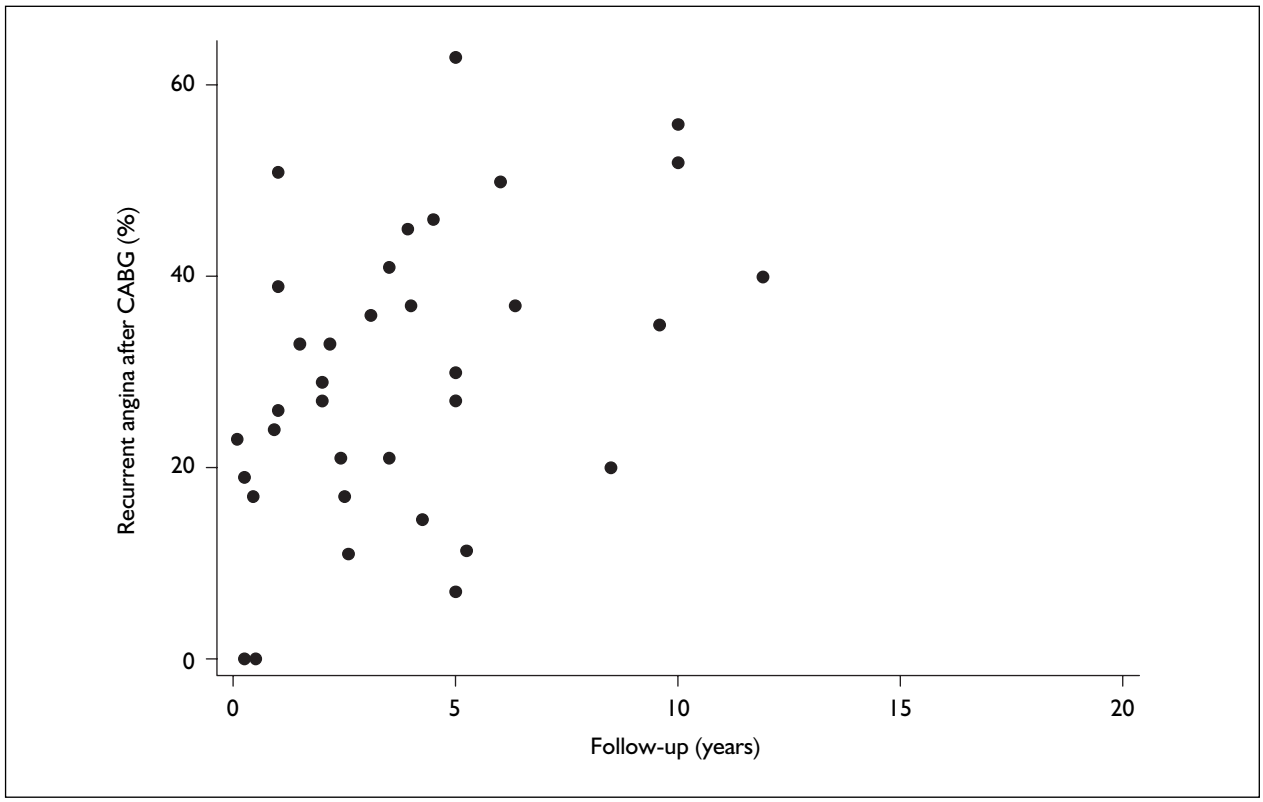

FIGURE 8 *Scatter plot of recurrent angina following CABG for angina against length of follow-up*

that one study (Gelbfish, $n = 28$) was excluded from the analysis as the adjusted mortality was equal to one.

A scatter plot of reported recurrent angina (proportion) against length of follow-up is shown in *Figure 8*. A significant positive association was found (robust regression coefficient 0.02, $p = 0.02$) and adjusted outcomes were therefore used.

Table 15 shows the results of the analyses for case series studies of CABG for the treatment of angina. The table notes the number of studies excluded as outliers through using robust regression for each investigated study characteristic. This number may be different for the same hypothesis investigated for different outcomes, as different numbers of studies report each outcome and record the aspects of the hypotheses that we are investigating.

As before, since blinding or independent outcome measurement is not an appropriate description of the outcome mortality and was not reported by any study, this has not been investigated. Comparative statistics were not calculated about registry studies as insufficient data were available (only two registry-based studies were reported).

Comparative statistics have not been calculated for the effect of consecutive enrolment as there are insufficient studies reporting non-consecutive enrolment ($n = 2$) to allow analysis. For multi-centre study design, only two studies reporting recurrent angina as an outcome also reported multi-centre enrolment, and there were therefore insufficient data for analysis.

Sample size was shown not to effect outcome frequency for either mortality or recurrent angina. In investigating sample size, six studies were excluded as outliers in the robust regression analysis. It should be noted that two of these are very large (Acinapura, $n = 3853$; Weintraub, $n = 2030$).

Compared with the results of regression for PTCA treatment for angina, the results for CABG are more discrepant. Although no effect on reported mortality and adjusted angina recurrence was noted with prospective enrolment compared with retrospective enrolment, discrepant results were seen for unadjusted angina recurrence. This was also true for independent outcome measurement.

Date of publication did not affect reported mortality. Adjusted recurrent angina rates showed significant results for robust regression analyses.

TABLE 15 *Investigation of study characteristics in CABG for angina studies*

Hypothesis	Yearly adjusted mortality	Adjusted angina recurrence	Angina recurrence (non-adjusted)
Sample size has no effect on undesirable outcome frequency	**No effect noted** Robust regression coefficient <0.0000, $p = 0.98$ $n = 49$, 6 studies excluded from regression: Weintraub, Mullany, MacDonald, Egstrup, Acinapura, Ruygrok	**No effect noted** Robust regression coefficient is <0.0000, $p = 0.37$ $n = 35$, 3 studies excluded from regression: Simmons, Ruygrok, Egstrup	**No effect noted** Robust regression coefficient <0.0000, $p = 0.69$ $n = 41$, no studies excluded from regression
Prospective studies show greater undesirable outcome frequency than retrospective studies	**No effect noted** 20 studies reported retrospective and 17 prospective measure of outcome. Median lower in prospective studies (0.01 vs 0.02) and means equal (0.04 vs 0.04) *t*-Test mean difference = 0.003 (SE 0.01), $p = 0.8$ Mann–Whitney test, $p = 1.0$ Weighted ANOVA, $p = 0.34$	**No effect noted** 20 studies reported retrospective and 17 prospective measure of outcome. Median (0.03 vs 0.024) higher in retrospective studies and mean (0.04 vs 0.044) higher in prospective studies *t*-Test mean difference = 0.003 (SE 0.013), $p = 0.8$ Mann–Whitney test, $p = 1$ Weighted ANOVA, $p = 0.34$	**Discrepant results** 17 studies reported retrospective and 9 prospective measure of outcome. Median (0.37 vs 0.24) and mean (0.34 vs 0.26) both higher in prospective studies *t*-Test mean difference = 0.07 (SE 0.07), $p = 0.33$ Mann–Whitney test, $p = 0.31$ Weighted ANOVA, $p = 0.002$
Case series show lower undesirable outcome frequency than registry- or population-based studies	**Comparative statistics not calculated** 12 studies clearly indicated no registry base and 2 from a registry Median (0.03 vs 0.02) higher from registry-based studies and means equal (0.03 vs 0.03)	**Insufficient data for analysis**	**Insufficient data for analysis**
Multi-centre studies show lower undesirable outcome frequency than single-centre studies	**No effect noted** 30 centres reported single-centre and 7 reported multi-centre enrolment Median (0.03 vs 0.02) and mean (0.05 vs 0.03) both higher in single-centre studies *t*-Test mean difference = 0.01 (SE 0.02), $p = 0.47$ Mann–Whitney test, $p = 0.68$ Weighted ANOVA, $p = 0.51$	**Comparative statistics not calculated** 23 centres reported single-centre and 2 reported multi-centre enrolment Comparative statistics not calculated. Median (0.1 vs 0.02) and mean lower in multi-centre studies (0.17 vs 0.02)	**Comparative statistics not calculated** 23 centres reported single-centre and 2 reported multi-centre enrolment Median lower in multi-centre studies (0.29 vs 0.14) and mean lower in multi-centre studies (0.30 vs 0.14)

continued

TABLE 15 *Investigation of study characteristics in CABG for angina studies (cont'd)*

Hypothesis	Yearly adjusted mortality	Adjusted angina recurrence	Angina recurrence (non-adjusted)
Prospective studies with consecutive enrolment show higher undesirable outcome frequency than those with non-consecutive enrolment	**Comparative statistics not calculated** 25 studies stated consecutive and 2 stated non-consecutive enrolment	**Comparative statistics not calculated** 21 studies stated consecutive and 1 stated non-consecutive enrolment	**Comparative statistics not calculated** 24 studies stated consecutive and 2 stated non-consecutive enrolment
Studies with independent or blinded measurement of outcome will show higher undesirable outcome frequency than those without such features	**Not applicable** No analysis undertaken	**No effect noted** 32 studies reported on objective measure of outcome, 19 no objective measure and 13 objective measure Median lower (0.09 vs 0.11) and mean higher (0.3 vs 0.19) in studies with objective outcome *t*-Test mean difference = 0.1 (SE 0.19), $p = 0.6$ Mann–Whitney test, $p = 0.48$ Weighted ANOVA, $p = 0.72$	**Discrepant results** 32 studies reported on objective measure of outcome, 19 no objective measure and 13 objective measure *t*-Test mean difference = 0.02 (SE 0.05), $p = 0.73$ Mann–Whitney test, $p = 0.88$ Weighted ANOVA, $p = 0.02$
Length of follow-up will be positively associated with undesirable outcome frequency	**Significant positive association** Robust regression coefficient 0.02, $p = 0.000$ (95% CI 0.009 to 0.02) $n = 50$, No studies excluded Weighted regression coefficient = 0.02, $p = 0.000$ (95% CI 0.02 to 0.03)	**Not applicable** Data adjusted for length of follow-up	**Discrepant results** Robust regression coefficient 0.02, $p = 0.02$ (95% CI 0.004 to 0.04) $n = 35$, no studies excluded Weighted regression coefficient = –0.01, $p = 0.40$ (95% CI –0.04 to 0.02)
The date of publication will be positively associated with undesirable outcome frequency	**No effect noted** Robust regression coefficient –0.005, $p = 0.32$, (95% CI –0.001 to 0.0005) $n = 53$, 3 studies excluded: Weintraub, MacDonald, Ruygrok Weighted regression coefficient = 0.002, $p = 0.72$ (95% CI –0.007 to 0.01) Difference between these two explained by Weintraub as being a large study with a high mortality, a clear outlier on scatter plot (see Appendix 6)	**Significant negative association in one analysis** Robust regression coefficient –0.004, $p = 0.01$, (95% CI –0.007 to –0.001) $n = 35$, 5 studies excluded: Ruygrok, Simmons, Gelbfish, Carter, Egstrup Weighted regression coefficient = –0.0004, $p = 0.95$, (95% CI –0.012 to 0.013)	**Significant negative association in one analysis** Robust regression coefficient –0.008, $p = 0.04$, (95% CI –0.02 to –0.0005) $n = 41$, no studies excluded Weighted regression coefficient = –0.003, $p = 0.60$, (95% CI –0.01 to 0.008)

TABLE 16 *Sample characteristics of CABG for angina case series studies*

	Range	Median	Mean (SD)	*n*
Mean age (years)	46	57	58.6 (8.6)	53
Proportion male	0.55	0.85	0.82 (0.11)	64
Proportion left ventricular dysfunction	0.55	0.51	0.47 (0.15)	19
Mean ejection fraction	0.14	0.59	0.58 (0.04)	20
Proportion with single-vessel disease[a]	1	0.09	0.15 (0.18)	40
Proportion of patients in NYHA grade 3 or 4	0.71	0.79	0.76 (0.18)	24
Proportion with left main stem disease	0.38	0.15	0.17 (0.09)	23

SD, standard deviation.
[a] Trimodal distribution, not suitable for regression analysis.

TABLE 17 *CABG – robust regression analysis of sample characteristics and adjusted mortality*

	Robust regression coefficient	95% CI	*p*-Value	No. of studies included in analysis
Age	0.001	0 to 0.002	0.05	42, excluded: Gelbfish
Proportion male	–0.10	–0.16 to –0.04	0.001	48, excluded: Ruygrok, Weinstraub, MacDonald
Proportion of patients in NYHA grade 3 or 4	–0.004	–0.04 to 0.03	0.85	16, excluded: Mullany, Weinstraub, Farrer, MacDonald

However, the size and direction of the regression coefficient were similar in all analyses, earlier publication date being associated with less favourable results.

CABG – multivariate analysis of case series studies

Population characteristics that may act as confounders and for which data were available are shown in *Table 16*. The proportion of patients with single-vessel disease in the studies exhibited a trimodal distribution, because the number of diseased vessels was an inclusion criterion in some studies. This variable was excluded from further analysis. Sample age, the proportion of male patients and the proportion of patients with more severe angina, as measured by the NYHA criteria, were the variables for which most data were available and these were therefore used in the multivariate analysis.

CABG – multivariate analysis of case series studies for mortality

Robust regression was used to investigate the sample characteristics of age, proportion of males and proportion of patients with NYHA grade 3 or 4 angina and the outcome of adjusted mortality. Age and proportion males were significant in univariate analyses (see *Table 17*) and showed effects in the expected direction. However, the proportion of patients with more severe angina was shown to have a non-significant negative effect on mortality, which is counter-intuitive. However, this analysis was only based on 16 studies and confidence intervals were wide.

Multivariate analysis was therefore conducted, including age and proportion of males in the sample. Sample size had no effect, with similar results to the univariate analysis shown in *Table 15*. Similar results to univariate analyses were also seen for length of follow-up and publication date. Length of follow-up (based on 39 studies, excluding Ruygrok and Weintraub) had a significant positive coefficient 0.022 (95% CI 0.016 to 0.028) $p = 0.000$. Publication date had a significant negative coefficient, –0.0015 (95% CI –0.003 to –0.0002), $p = 0.02$ (based on 38 studies, excluding Ruygrok, Gelbfish and MacDonald). There were insufficient data to analyse the other hypotheses.

CABG – multivariate analysis of angina recurrence

Robust regression analysis was performed to explore possible explanatory sample characteristic variables associated with non-adjusted angina recurrence (*Table 18*) and adjusted angina recurrence (*Table 19*). None of the potential

TABLE 18 *CABG – robust regression for sample characteristics and angina recurrence (unadjusted)*

	Robust regression coefficient	95% CI	*p*-Value	No. of studies included in analysis
Age	–0.004	–0.01 to 0.003	0.24	30, no studies excluded
Proportion male	0.16	–0.41 to 0.72	0.58	37, no studies excluded
Proportion of patients in NYHA grade 3 or 4	–0.02	–0.78 to 0.72	0.95	13, no studies excluded

TABLE 19 *CABG – robust regression for sample characteristics and adjusted angina recurrence*

	Robust regression coefficient	95% CI	*p*-Value	No. of studies included in analysis
Age	<–0.0000	–0.004 to 0.004	0.99	22, excluded: Simmons, Egstrup)
Proportion male	–0.37	–0.74 to 0.01	0.056	30, excluded: Ruygrok
Proportion of patients in NYHA grade 3 or 4	–0.02	–0.7 to 0.073	0.095	13

explanatory variables were significant in univariate analyses for unadjusted angina recurrence.

Multivariate analyses were undertaken including age and proportion of males for completeness. As was seen in univariate analysis (*Table 15*), no effect was seen for sample size (based on 28 studies unadjusted; 22 studies adjusted, excluding Ruygrok). Publication date had a negative coefficient of –0.005 (95% CI –0.015 to 0.005), $p = 0.37$ (based on 28 studies, unadjusted data). This coefficient value was similar to that in the univariate analysis but no longer significant. For adjusted data, publication date showed a significant negative coefficient (–0.007, 95% CI –0.01 to –0.004, $p = 0.00$) based on 19 studies (excluding Simmons, Gelbfish, Farrer and Ruygrok).

The coefficient for the length of follow-up based on unadjusted data was similar to that in univariate analysis at 0.02 (95% CI –0.0034 to 0.04), $p = 0.09$. For adjusted data, similar results were obtained (coefficient –0.019, 95% CI –0.037 to –0.002, $p = 0.032$; based on 22 studies). There were insufficient data to analyse the other hypotheses.

Comparison of case series results with RCT results

It was hypothesised that case series data would show higher desirable outcome frequency than a similar intervention in one arm of an RCT. This section reports on the results of an analysis addressing this hypothesis.

FESS – case series and RCT comparison

There were only three RCT studies relating to FESS. One of these reported only disease recurrence as an outcome and the other two only reported symptomatic improvement, further restricting the amount of data available for investigation. All the RCTs were described as of poor quality by the authors of the original TAR, having inadequate randomisation, variation in applied intervention and loss to follow-up, and two studies also had limited study power.

There were insufficient data to analyse the differences between RCT and case series in this example. However, the mean patency result across case series studies was higher than that seen in RCTs (*Table 20*).

Spinal cord stimulation – comparison of RCTs and case series results

Only one RCT was identified by the report on SCS. This study was assessed by the authors of that report as being of poor quality. Details of randomisation and allocation were absent and the study did not adequately describe baseline patient characteristics. Comparison of RCT and case series study results was not undertaken.

PTCA – comparison of RCTs and case series results for mortality

For angina treated by PTCA, the mean time at which mortality was measured was similar for case series and RCTs but the range was skewed by a small number of case series studies with very long

TABLE 20 *FESS – comparison of results from RCT and case series studies*

Hypothesis	Symptom improvement (%)	Patency
Case series will show higher desirable outcome frequency than a similar intervention arm in an RCT	Insufficient data for regression analysis, no difference in means	Insufficient data for regression analysis, mean in case series 88% vs RCT arm 38%

TABLE 21 *PTCA – details of time in years at which mortality was measured in different study designs*

Study design	Range	Median	Mean	SD	No. of studies reporting outcome
RCT	4.33	2.92	2.82	1.46	10
Case series	9.67	2.25	2.81	2.03	59
Case–control	0	3.33	3.33	0	2
Comparative	7.67	2	2.23	2.06	11

TABLE 22 *PTCA – yearly mortality reported by different study designs (unadjusted)*

Study design	Range	Median	Mean	SD	No. of studies included
RCTs	0.13	0.04	0.061	0.043	9
Case series	0.23	0.0425	0.055	0.046	54
t-Test difference = 0.009 (SE 0.016), $p = 0.58$. Mann–Whitney test, $p = 0.46$. Weighted ANOVA, $p = 0.59$.					

TABLE 23 *PTCA – yearly mortality reported by different study designs adjusted for length of follow-up*

Study design	Range	Median	Mean	SD	No. of studies included
RCTs	0.04	0.020	0.024	0.01	9
Case series	0.24	0.017	0.027	0.04	51
t-Test difference = 0.003 (SE 0.01), $p = 0.83$. Mann–Whitney test, $p = 0.30$. Weighted ANOVA, $p = 0.72$.					

TABLE 24 *PTCA – comparison of RCT and case series results for angina recurrence (unadjusted)*

	Range	Median	Mean	SD	No. of studies reporting outcome
RCTs	0.33	0.25	0.26	0.11	6
Case series	0.66	0.30	0.34	0.17	50
t-Test difference = 0.08 (SE 0.07), $p = 0.25$. Mann–Whitney test, $p = 0.33$. Weighted ANOVA, $p = 0.10$.					

follow-up (see *Table 21* for comparison of the range). Median follow-up was greater for RCTs than for case series. Other study designs were included for completeness.

Table 22 shows unadjusted mortality and *Table 23* shows mortality adjusted for length of follow-up. In both cases, the case series show a much greater range, a lower median but a higher mean mortality. None of the differences are significant.

PTCA – comparison of RCTs and case series studies for angina recurrence

Table 24 shows the unadjusted rates of recurrent angina following PTCA as reported in RCTs and case series.

Table 25 shows the rates adjusted for length of follow-up. In both analyses, case series show higher levels of recurrent angina (median and mean) but also a larger range. However, none of these differences were statistically significant.

TABLE 25 *PTCA – comparison of RCT and case series results for adjusted angina recurrence*

	Range	Median	Mean	SD	No. of studies reporting outcome
RCTs	0.25	0.12	0.13	0.089	6
Case series	0.97	0.10	0.28	0.23	48
t-Test difference = 0.148 (SE 0.17), $p = 0.37$. Mann–Whitney test, $p = 0.76$. Weighted ANOVA, $p = 0.9$.					

TABLE 26 *CABG – details of time in years at which mortality was measured in different study designs*

Study design	Range	Median	Mean	SD	No. of studies reporting outcome
RCT	4.33	2.5	2.68	1.59	10
Case series	19.92	4	5.00	4.09	61
Case–control	2.5	3.75	3.75	1.77	2
Comparative	6.08	2.9	3.8	2.3	6

TABLE 27 *CABG – crude mortality reported by different study designs (unadjusted)*

Study design	Range	Median	Mean	SD	No. of studies reporting outcome
RCTs	0.11	0.047	0.055	0.04	9
Case series	0.62	0.115	0.15	0.14	56
t-Test difference = –0.1046 (SE 0.048), $p = 0.033$. Mann–Whitney test, $p = 0.0061$. Weighted ANOVA, $p = 0.41$.					

TABLE 28 *CABG – yearly mortality reported by different study designs adjusted for length of follow-up*

Study design	Range	Median	Mean	SD	No. of studies reporting outcome
RCTs	0.12	0.02	0.03	0.04	9
Case series	0.21	0.03	0.04	0.04	56
t-Test difference = 0.009 (SE 0.01), $p = 0.53$. Mann–Whitney test, $p = 0.36$. Weighted ANOVA, $p = 0.28$.					

CABG – comparison of RCTs and case series for mortality

Table 26 shows the average times at which the outcome mortality was measured for different study designs. Both mean and median follow-up times were longer for case series than for RCTs, but case series also had a much larger range. Hence, unadjusted mortality would be expected to be higher in case series studies. Other study designs were included for completeness.

Table 27 shows post-CABG mortality reported by RCTs and case series unadjusted for follow-up and *Table 28* shows mortality adjusted for length of follow-up. The case series studies show a greater range, a higher median and a higher mean mortality. Differences are significant in unadjusted analyses.

CABG – comparison of RCTs and case series for angina recurrence

Table 29 shows the unadjusted rates of recurrent angina reported after CABG treatment in RCTs and case series. Again, the case series show a larger range, a higher median and a higher mean proportion of angina recurrence. These differences are statistically significant [*t*-test difference = 0.16 (SE 0.06), $p = 0.01$; Mann–Whitney test, $p = 0.01$; weighted ANOVA, $p = 0.06$]. However, when these figures are adjusted for length of follow-up, the difference is non-significant [*t*-test difference = –0.12 (SE 0.17), $p = 0.5$; Mann–Whitney test, $p = 0.34$; weighted ANOVA, $p = 0.90$] (*Table 30*).

TABLE 29 *CABG – comparison of RCT and case series for angina recurrence (unadjusted)*

Study design	Range	Median	Mean	SD	No. of studies reporting outcome
RCTs	0.24	0.15	0.15	0.08	8
Case series	0.63	0.30	0.30	0.17	41

TABLE 30 *CABG – comparison of RCT and case series for adjusted angina recurrence*

Study design	Range	Median	Mean	SD	No. of studies reporting outcome
RCTs	0.25	0.05	0.08	0.08	8
Case series	0.76	0.09	0.13	0.16	35

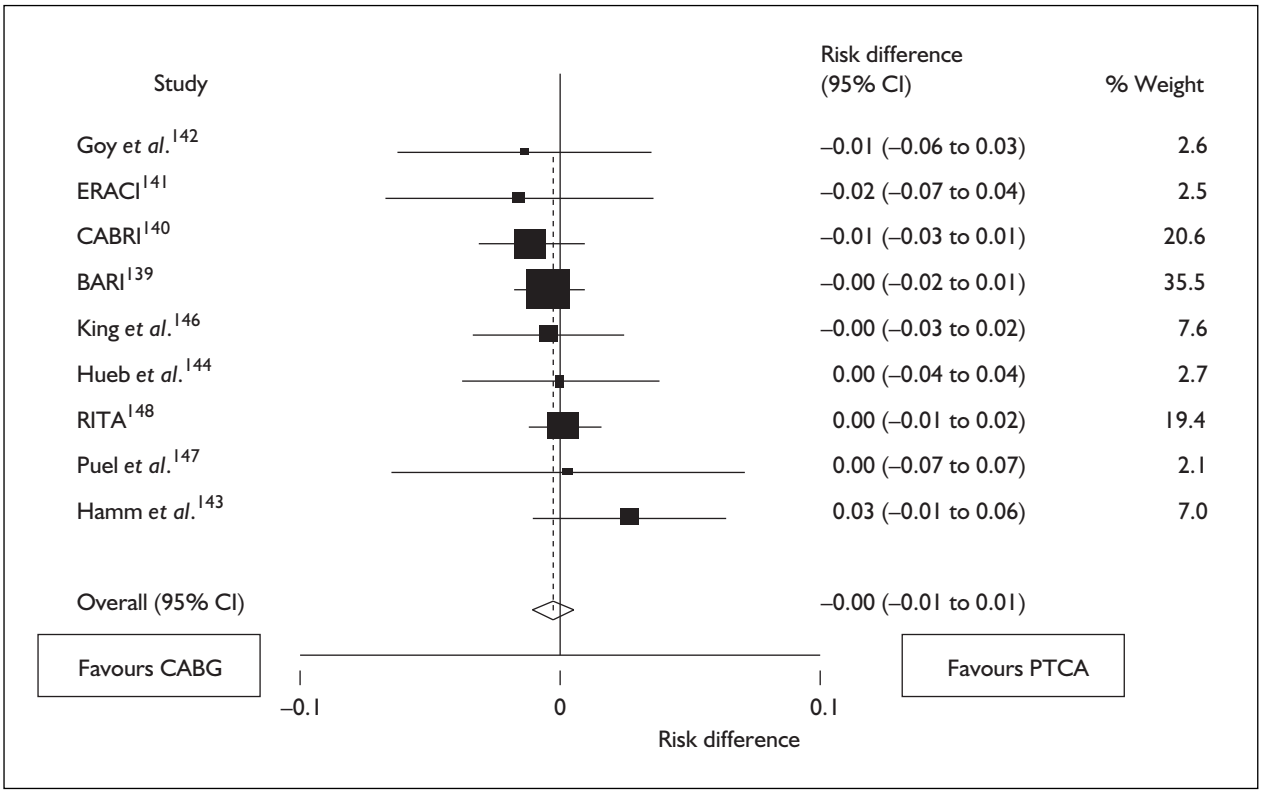

FIGURE 9 *Meta-analysis showing risk difference for adjusted mortality CABG versus PTCA*

CABG versus PTCA – estimate of outcome frequency

Mortality – case series results

Using robust regression, the coefficient was estimated as 0.008 (SE 0.003) for CABG compared with PTCA, based on 99 observations. Nine studies were excluded (Ruygrok, Mullany, Acinapura, Safian, Gelbfish, MacDonald, Urban, Weintraub and Egstrub). The 95% CI were 0.0009 to 0.014, $p = 0.03$. This is a significant difference, with CABG, on average, being associated with a 0.8% higher yearly adjusted mortality.

Weighted regression estimated the coefficient as 0.024 (95% CI 0.01 to 0.04), $p = 0.001$. This suggested a higher yearly adjusted mortality associated with CABG of 2%.

Mortality – RCT results

For the RCTs, a random effect meta-analysis showing the risk difference for mortality following CBAG and PTCA was performed. No difference between the two treatments was found ($p = 0.570$) (*Figure 9*).

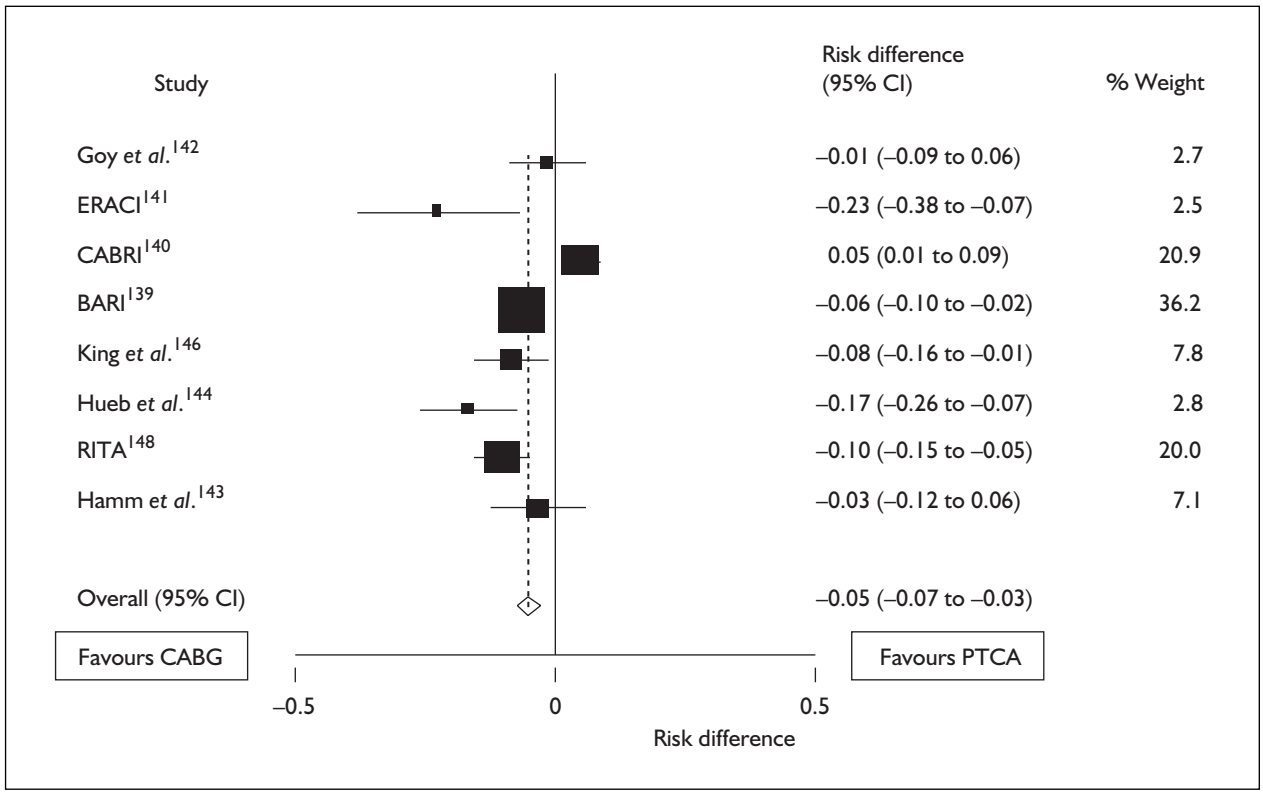

FIGURE 10 *Meta-analysis showing risk difference for recurrent angina (unadjusted) with CABG and PTCA*

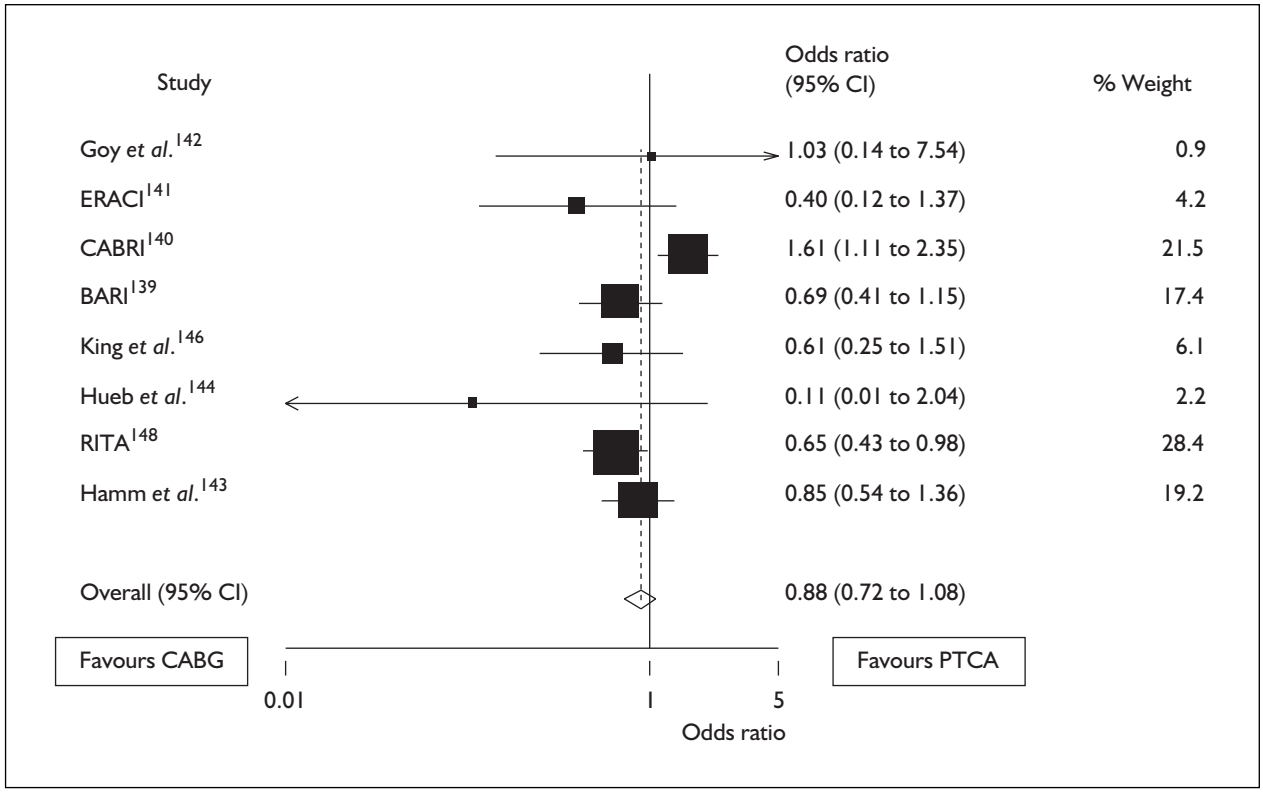

FIGURE 11 *Meta-analysis showing risk difference for angina (adjusted) with CABG and PTCA*

TABLE 31 *Summary of results*

	FESS		SCS	Angina PTCA		Angina CABG	
	Symptom improvement	**Patency**	**Pain relief**	**Mortality**	**Angina recurrence**	**Mortality**	**Angina recurrence**
Sample size	No effect	No effect	No effect	No effect	No effect	No effect	No effect
Prospective	No effect	No effect	No effect	No effect	No effect	No effect	Discrepant
Registry	Insufficient data	Insufficient data	Insufficient data	Insufficient data	Insufficient data	Insufficient data	Insufficient data
Multi-centre	Higher in multi-centre series	Higher in multi-centre series	No effect	Insufficient data	Insufficient data	No effect	Insufficient data
Consecutive enrolment	No effect	No effect	Insufficient data	Insufficient data	Insufficient data	Insufficient data	Insufficient data
Independent measure	No effect	No effect	Insufficient data	No effect	Independent measured higher recurrence	No effect	Discrepant
Length of follow-up	No effect	No effect	Longer follow-up less pain relief	Higher mortality longer follow-up	No effect	Higher mortality longer follow-up	Discrepant
Publication date	No effect	No effect	No effect	No effect	Early publication higher recurrence	No effect	Early publication higher recurrence
Quality score	Insufficient data	Insufficient data	Better quality score lower pain relief	Insufficient data	Insufficient data	Insufficient data	Insufficient data

Angina recurrence – case series results

Using robust regression, the coefficient was estimated as –0.026 (SE 0.039) for CABG compared with PTCA based on 91 observations. No studies were excluded. The 95% CI were –0.10 to 0.05, $p = 0.50$. This is a non-significant difference. Weighted regression estimated the coefficient as 0.009 (95% CI –0.05 to 0.07), $p = 0.7$.

Angina recurrence – RCT results

For RCTs a meta-analysis was carried out to examine the risk difference between CABG and PTCA for recurrent angina. This was performed for both adjusted (for length of follow-up) and unadjusted rates of recurrent angina (*Figures 10* and *11*). For unadjusted rates of recurrent angina, CABG appeared to have a significantly lower rate of recurrent angina (–0.05, 95% CI –0.07 to –0.03) However, this was not apparent in the adjusted calculation ($p = 0.217$).

Summary of main findings

There were insufficient data in the case series studied to address all the hypotheses set out at the start of the project and the findings shown were not consistent across the different series and are subject to considerable uncertainty. Poor reporting of methods severely constrained the analyses.

Main findings are listed in *Table 31*. We found no relationship between sample size and outcome in the cases studied. Whether a study was prospective or retrospective was not shown to be associated with outcome frequency. In one analysis (FESS), multi-centre studies showed a significant association with outcome. In one analysis (angina recurrence following PTCA) we found an association between independent measurement of outcome and outcome frequency. In the analysis of case series examining PTCA and CABG we found some evidence that earlier publication may be associated with less favourable outcomes. As expected, length of follow-up was related to outcome in three of the analyses, although in the case of angina recurrence after PTCA this was not significant and we included adjusted and unadjusted analyses for completeness. In the analysis of case series for spinal cord surgery the quality score of studies was associated with

outcome, with studies scoring higher on the quality score reporting lower pain relief. The quality score used by the original researchers included the items whose impact on outcome was investigated separately, without evidence of an effect.

In the comparison between case series and RCTs, only data from the CABG and PTCA series were used. For PTCA, there were no significant differences between the mean mortality or recurrence of angina reported across case studies and that reported in the treatment arms of the RCTs. However, the range reported for all outcomes was greater in the case series than RCTs. For CABG, case series showed higher mortality than RCTs. However, this was related to length of follow-up, and when adjusted for this factor the differences shown were no longer statistically significant.

We compared CABG and PTCA using case series and carried out meta-analysis of RCTs to investigate differences in the potential conclusions of such a comparison using different study designs. Using case series, CABG was associated with a 1–2% increase in mortality compared with PTCA. The meta-analysis showed no difference between interventions. Both case series and meta-analysis demonstrated no difference between interventions for angina recurrence when length of follow-up was taken into account in the RCTs. Great caution should be exercised when making indirect comparisons as we have here.

Chapter 6
Discussion

Main results

It has been argued that case series cannot be used to assess effectiveness as, in the absence of a control group, it is impossible to conclude that any observed outcome is caused by the treatment given. By using evidence from case series in decision-making, there is the risk that study results will be misleading. Such misinformation may lead to treatments which are not beneficial and possibly harmful being adopted. Examples exist in the literature, such as observational studies of hormone replacement therapy (HRT), which suggested a potential benefit,[62] whereas later RCTs (the HERS trial[63]) revealed no net benefit. As the women who take HRT are more likely to be from wealthier backgrounds, this was a major confounding factor in the observational data.

However, we identified 14 cases among NICE assessments where data taken from case series had been considered. Other non-randomised study designs (such as case–control and cohort studies) were also included in half of these reports. The most common reason ($n = 5$) for including case series was the absence of RCT data. Ten different methods of quality assessment were used. From this, we conclude that, despite their critical methodological weaknesses, case series will continue to play a significant role in health technology assessments, particularly in systems such as the NICE appraisal process, which predominantly consider new technologies. The plethora of approaches to quality assessment of case series reflects uncertainty about the importance of different methodological features of case series and supports our subsequent attempts to investigate the relationship between methodology and outcome in a small number of case studies. Our literature review did not identify any previous attempts to address this issue.

The data in our case series were limited and it was not possible to address several of the hypotheses set out at the beginning of the study. Insufficient data were available to compare case series with population registries in all the analyses. No analysis addressed all the hypotheses. Confounding is almost certain to be present, in addition to ecological bias.

Overall, we found limited evidence of association between methodological features and outcome in the analyses carried out. However, a consistent finding across all the case studies was of no relationship between sample size and outcome frequency. Although the number of examples studied was very small, this finding, if replicated, may have important implications for health technology assessments. Hitherto, sample size has been used as a criterion for the inclusion or exclusion of case series from reviews. The lack of relationship between study size and outcome suggests that this approach may not be justified. Where case series are included in reviews it is likely that they will be more numerous than RCTs or other designs (as in the cases we report on here). If reviews are being carried out to a limited timescale, as in the NICE appraisal process, there is therefore a strong incentive for researchers to limit the number of case series included in the review, supported perhaps by the view that this design is necessarily less likely to result in robust conclusions. Our findings tentatively suggest that setting a cut-off in terms of sample size may be less justified than including all studies or taking a random sample of those available.

We found no evidence that prospective series, or those in which consecutive cases were enrolled, were associated with different outcome frequency to studies not having these features. Again, these criteria are frequently used to judge the quality of case series. These analyses were particularly constrained by inadequate reporting in the original studies. However, all the examples explored were surgical interventions and it may be that where retrospective designs were used case ascertainment was good, reflecting the ease of identifying patients following surgical procedures from hospital records. Where ascertainment is more difficult retrospectively, for example for drug technologies, a greater difference may be shown between retrospective and prospective studies or those in which recruitment was or was not consecutive. A further consideration in this and all the analyses showing no association between methodological features and outcome is the limited power to detect a significant difference afforded by the small number and heterogeneity of studies included in the examples studied.

In the case of SCS, a significant association between the quality score used by reviewers and outcome was demonstrated. This was the only example in which such a score had been used and may suggest that the use of quality scoring systems can differentiate between studies. However, we found no relationship between the individual study factors which made up the score and outcome in the case series. It is therefore difficult to conclude whether the score is acting as a valid measure of study quality. Since the relationship between methodological features and validity is not clear and how item scores should be summed into a single measure of study quality remains uncertain, it may be unwise to use such single scoring systems to judge the quality of case series.

Our other findings were inconsistent across the case studies and the small number of associations demonstrated cannot be taken as good evidence on which to base any change in approach to the appraisal of case series. The finding, in one analysis, that independent measurement of outcome may be important in determining study quality is consistent with the findings of Juni and colleagues regarding blinding in relation to the quality of RCTs.[64] This is potentially important, but further evidence is required of the importance of this factor in other case series.

The failure to demonstrate any relationship between date of publication and outcome may, as with the other negative findings in this study, be related to limited statistical power. Three other explanations are possible. First, the impact of early adopters and any effects of selection of cases in early studies may be short-lived and therefore not apparent when a longer historical perspective is taken. Second, the effect of the learning curve in the early stages of use of a technology may counteract the effects of case selection. Third, technological improvements may have a very marked effect on successful outcomes.

The comparison between case series and RCTs for PTCA and CABG showed that the case series reported a greater range in outcome frequency than RCTs and higher mortality, although the difference between study designs was not statistically significant. This supports the view that case series include more heterogeneous populations that may be a better representation of routine practice than is achieved in RCTs designed to evaluate efficacy. The higher reported rates of adverse outcome in case series probably reflects a broader case mix in these studies than RCTs and does not support the view that case series are likely to provide more optimistic estimates of treatment effectiveness. The direction of bias introduced by the greater variance in results from case series is unlikely to be consistent, as has been shown in comparisons of randomised and non-randomised trials,[22] although this was not studied in the current project.

The potential difficulties in comparing treatments through the use of case series is demonstrated by the comparison of relative effectiveness of PTCA and CABG using case series and RCTs. Mortality was judged to be higher for CABG than PTCA in the case series analysis whereas no difference was shown in the meta-analysis of RCTs. Interpretations of this finding may be conflicting. On the one hand, the estimate of effectiveness from the RCT is less likely to be biased and so a conclusion based only on the results of the case series analysis would be unwise. However, if the case series include a more typical population then the finding of higher mortality may reflect the 'real world' case more accurately.

We suggest that there are complementary positions for different methodological approaches in the ongoing evaluation of health technologies. Although researchers should always consider carrying out RCTs to establish efficacy and effectiveness, this is clearly not currently the case for the technologies examined in this project where only a tiny proportion of the total study populations were included in RCTs. In some cases, where the natural history of the condition is well understood and a dramatic effect is shown by a technology, comparative studies may not be considered necessary or ethical. We expect that such cases will be very few. It is more likely that case series will continue to be carried out in the early stages of technology diffusion, particularly in surgery where there is a less stringent regulatory framework governing adoption. Such case series will be important in identifying whether technologies are likely to be efficacious. Early assessment of case series may therefore identify technologies which should be subject to more rigorous evaluation. Efficacy may then be established through well-conducted RCTs. This, however, may be insufficient to inform practice and policy and for some technologies it may be necessary to continue to collect data through case series or, more systematically, through the use of comprehensive registries. These hold a number of potential advantages over case series led, more conventionally, by the clinicians delivering the intervention. Standardisation of data collection

and reporting is more feasible, investigation of the effects of centre and operator are more feasible and the establishment of an ongoing system for reporting of process and outcomes would demonstrate changes in the nature of the technology, which is a particular issue in the development of surgical techniques. A key advantage of ongoing collection of data through large case series is the identification of uncommon side-effects in practice and a high degree of external validity. Using registry or case series data to make a comparison between technologies will continue to be necessarily and severely constrained by the non-direct nature of such comparisons and the effect of a large range of known and unknown confounders. However, the collection of data on the performance of technologies in undifferentiated populations over long periods will complement and may extend the knowledge yielded in the generally short timescales and selected populations of RCTs.

Assumptions, limitations and uncertainty

Despite the inclusion of handsearching of key journals, the literature review for methodological studies found no examples of published relevant work and it is possible that our search failed to identify relevant studies. In particular, we concentrated on identifying studies which focused on methodological issues. We may therefore have missed relevant methodological considerations included in papers whose focus was clinical rather than methodological. However, we think it is unlikely that a large volume of literature on this subject exists because of the generally low level of interest in case series designs by methodologists, confirmed by the negative results of the searches carried out.

In the investigation of possible impact of methodological aspects of case series, our examples were all surgical interventions. This means that our findings may not be generalisable to evaluations of other types of technology using case series. As noted above, the effects of learning curve and the possibility of bias arising from enthusiastic early adopters may act in opposite directions, making it difficult to discern any effect relating to timing of publication. A more important problem arising from the nature of the technologies examined is the introduction of further variance in the data as a result of operator effects which would not be apparent in, for example, drug technologies.

The small number of cases examined and the relatively limited number of studies in each set of case series are important limitations to precision and generalisability which may be addressed by further research. However, it is likely that empirical opportunities for investigation will be few, as has been shown in the comparisons of randomised and non-randomised controlled trials. Under these circumstances modelling studies may be valuable. Our analysis was necessarily limited to the aggregate reports of individual studies and there is therefore the potential for ecological bias.

A general problem in the data examined is the very low 'signal-to-noise' ratio. In other words, it is difficult to identify the effects of methodological factors from the potential confounding effects of heterogeneity between studies in aspects of the populations and interventions. This is a particular problem where reporting of population and intervention characteristics was limited. The impact of unknown confounders, the fundamental reason for favouring RCTs over other study designs, is also an important consideration.

The potential role of publication bias should be considered. As stated earlier, case series may be particularly prone to publication bias, although this was not formally explored in the current review. Case series are recognised as much less robust than comparative or experimental designs and they may therefore be less likely to achieve publication in any journal or in journals indexed on major electronic databases. Small case series are likely to be more prone to this bias, as in other study designs, and those with less impressive findings and small size are likely to be at greatest risk. However, two findings in this study suggest that publication bias may not be a particular problem in the examples studies. First, the finding of no association between sample size and outcome suggests that smaller studies are not more likely to be positive. Second, the very large range in sample sizes among studies suggests that even small studies are achieving publication. The extent to which these findings are likely to be replicated in other reviews of case series is unknown and further research into the extent and impact of publication bias in different study designs is required.

The general finding of poor reporting of methodological features in case series is a cause for concern and will continue to hamper research into case series and the ability of decision-makers to consider the appropriate influence of case series evidence on policy. We chose to constrain our

analyses to reported data, that is, where a methodological feature was not reported in a study this was excluded from the analysis. The reporting of other study designs, notably RCTs and systematic reviews, has been improved considerably in recent years. Although case series rightly occupy a position low in the hierarchy of evidence, their continued use in health technology assessments strongly suggests the need to improve the quality of reporting, such as whether a study was prospective or whether cases were enrolled consecutively.

Some statistical considerations should be borne in mind when interpreting the results of our analyses. Weighting for study size in regression is generally favoured in meta-analysis, and gives greater weight to larger studies in order to improve precision. Although sample size is not the only determinant of variance in studies, in the current context it is likely to dominate other factors. As we did not have data on the variance of individual studies, we were constrained to using sample size alone. We found no relationship with study size which does not support concern over increased bias in smaller observational studies. Hence we included weighted regression for completeness. Our main statistical approach was robust regression. This is a slightly more conservative technique than OLS regression but performs better than OLS with non-ideal data. This technique resists the influence of extreme outliers, but results in slightly larger standard errors. Hence the power of robust regression to detect true differences is slightly reduced compared with OLS regression. Given the nature of our data we consider this to be a reasonable analytic approach. The use of several methods of analysis has led, in some cases, to apparently discrepant results. Given the large number of analyses performed, the usual level of significance of $p = 0.05$ should be viewed with caution. This further demonstrates the tentative nature of our findings and that more work is required to investigate the potential impact of methodological features of case series on their results.

Need for further research

The current study is exploratory and largely inconclusive. Further research examining the relationship between methodological features of case series and their outcome is justified. The case study approach that we have taken could be replicated in a wider sample of technologies, seeking larger sets of case series.

A wider study of the use of case series in technology appraisal systems would be of value in demonstrating the impact of this study design on decision making in contexts outside the NICE appraisal process.

Further research is needed into the extent and impact of publication bias on reviews including different study designs.

Value of information analyses including case series data is a methodological area that could be further explored.

Further exploration of the differences between case series results and RCT results would be valuable, preferably by using registry or comprehensive case series data.

Chapter 7
Conclusions

Case series are incorporated in a significant proportion of health technology assessments.

A wide range of quality criteria have been used to appraise the quality of case series and decide on their inclusion in reviews of studies using this design. In a small series of case studies drawn from health technology assessments carried out for the NHS HTA programme, we found little evidence to support the use of many of the factors included in quality assessment tools. Importantly, we found no relationship between study size and outcome across the four examples studied.

Isolated examples of a potentially important relationship between other methodological factors and outcome were shown, for example blinding of outcome measurement, but these were not shown consistently across the small number of examples studied.

Comparison of case series and RCT data was possible in only two examples studied but demonstrated a greater range in outcomes reported in case series, reflecting the likelihood that this design includes different populations. However, outcomes were not better in case series, contrary to expectations.

Estimates of comparative efficacy of alternative techniques by comparing case series studies were shown to be different from analyses based on RCTs. However, it is not clear from this whether this is an effect of confounding or indicates different efficacy in different populations. This study is based on a very small sample of studies and should therefore be considered as exploratory. Further investigation of the relationship between methodological features and outcome are justified given the frequency of use of case series in health technology assessments.

Acknowledgements

We are grateful to the authors of the previous HTA reports which were included in the empirical study and to the expert reviewers who provided careful and helpful comments on the final draft of this report.

Contributions of authors

Kim Dalziel wrote the protocol, undertook the review of case series used in NICE HTAs and contributed to writing the report. Ali Round undertook joint management of the project, contributed to the writing of the protocol and the report and conducted the analyses. Ken Stein undertook joint management of the project, contributed to the writing of the protocol and report and checked extracted data. Ruth Garside checked and extracted data and contributed to the writing of the report. Emanuela Castelnuovo checked and extracted data and conducted the analyses. Liz Payne designed and carried out the literature searches.

This report was commissioned by the NHS R&D HTA programme.

References

1. Clarke M, Oxman A. Cochrane Reviewers' Handbook 4.1.5 [updated April 2002]. *Cochrane Library, Issue 2, 2002*. Oxford: Update Software; 2002.
2. Garside R, Round A, Dalziel K, Stein K, Royle P. The effectiveness and cost-effectiveness of imatinib in chronic myeloid leukaemia: a systematic review. *Health Technol Assess* 2002;**6**(33).
3. Prevost TC, Abrams KR, Jones DR. Hierarchical models in generalized synthesis of evidence: an example based on studies of breast cancer screening. *Stat Med* 2000;**19**:3359–76.
4. Geddes JR, Game D, Jenkins NE, Peterson LA, Pottinger GR, Sackett DL. What proportion of primary psychiatric interventions are based on evidence form randomised controlled trials? *Qual Health Care* 1996;**5**:215–17.
5. Ellis J, Mulligan I, Rowe J, Sackett DL. Inpatient general medicine is evidence based. *Lancet* 1995; **346**:407–10.
6. Lewis JA, Facey KM. Statistical shortcomings in licensing applications. *Stat Med* 1998;**17**:1663–73.
7. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 2000; **342**:1887–92.
8. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med* 2000;**342**:1878–86.
9. 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).
10. MacLehose RR, Reeves BC, Harvey IM, Sheldon TA, Russell IT, Black AM. A systematic review of comparisons of effect sizes derived from randomised and non-randomised studies. *Health Technol Assess* 2000;**4**(34).
11. Ioannidis, JP, Haidich AB, Lau J. Editorial. Any casualties in the clash of randomised and observational evidence? *BMJ* 2001;**322**:879–80.
12. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* 1998;**317**:1185–90.
13. Linde K, Scholz M, Melchart D, Willich SN. Should systematic reviews include non-randomized and uncontrolled studies? The case of acupuncture for chronic headache. *J Clin Epidemiol* 2002; **55**:77–85.
14. Hrobjartsson A, Gotzsche PC. Is the placebo powerless? *N Engl J Med* 2001;**344**:1594–602.
15. Harkness E, Ernst E. The enigmatic placebo effect – a systematic review to define its determinants. *Perfusion* 2000;**13**:164–70.
16. Avis M. Reading research critically. II. An introduction to appraisal: assessing the evidence. *J Clin Nurs* 1994;**3**:272–7.
17. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Comm Health* 1998;**52**:377–84.
18. Fletcher RH. Evaluation of interventions. *J Clin Epidemiol* 2002;**55**:1183–90.
19. Gorkin L, Schron EB, Handshaw K, Shea S, Kinney MR, Branyon M, *et al*. Clinical trial enrollers vs. non-enrollers: the cardiac arrhythmia suppression trial (CAST) recruitment and enrollment assessment in clinical trials (REACT) project. *Control Clin Trials* 1996;**17**:46–59.
20. Brambilla P, Barale F, Caverzasi E, Tognoni G, Barbui C. Clozapine-treated subjects with treatment-resistant schizophrenia: a systematic review of experimental and observational studies. *Int Clin Psychopharmacol* 2002;**17**:189–95.
21. Chen CI, Skingley P, Meyer RM. A comparison of elderly patients with aggressive histology lymphoma who were entered or not entered on to a randomized phase II trial. *Leuk Lymphoma* 2000; **38**:327–34.
22. McKee M, Britton A, Black N, McPherson K, Sanderson C, Bain C. Methods in health services research. Interpreting the evidence: choosing between randomised and non-randomised studies. *BMJ* 1999;**319**:312–15.
23. Crome P, Gray R, Hills RK, Bentham P, Sellwood E, for the AD2000 Study Group. No evidence of co-morbidity or age influencing response to, or compliance with, donepezil. Results of the AD2000 trial. *Age Ageing* 2203; **32** (Suppl 1):7 (abstract).
24. Begg CB, Cho MK, Eastwood S, Horton R, Moher D, Olkin I, *et al*. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *JAMA* 1996;**276**:637–9.

25. Chalmers TC, Celano P, Sacks HS, Smith HJ. Bias in treatment assignment in controlled clinical trials. *N Engl J Med* 1999;**309**:1359–61.

26. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;**273**:408–12.

27. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, *et al*. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998; **352**:609–13.

28. Kjaergerd LL, Willumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Ann Intern Med* 2001;**135**:982–9.

29. Juni P, Witcshi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials in meta-analysis. *JAMA* 1999;**282**:1054–60.

30. Moher D, Jadad AR, Nichol G, Fenman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials* 1995;**16**:62–73.

31. Jadad A, Moore R, Carroll D, Jenkinson C, Reynolds D, Gavaghan D, *et al*. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996; **17**:1–12.

32. Shapiro S. Is meta-analysis a valid approach to the evaluation of small effects in observational studies? *J Clin Epidemiol* 1997;**50**:223–9.

33. Spitzer WO, Lawrence V, Dales R, Hill G, Archer MC, Clark P, *et al*. Links between passive smoking and disease: a best-evidence synthesis (a report of the Working Group on Passive Smoking). *Clin Invest Med* 1990;**13**:17–43.

34. Boulware LE, Daumit GL, Frick KD, Minkovitz CS, Lawrence RS, Powe NR. Quality of clinical reports on behavioral interventions for hypertension. *Prev Med* 2002;**34**:463–75.

35. Cowley D. Prostheses for primary total hip replacement: a critical appraisal of the literature. *Int J Technol Assess Health Care* 1995;**11**:770–8.

36. DuRant RH. Checklist for the evaluation of research articles. *J Adolesc Health* 1994;**15**:4-8.

37. Coleridge Smith P. The management of chronic venous disorders of the leg: an evidence-based report of an International Task Force. *Phlebology* 1999;**14** (Suppl 1):1–131.

38. Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori S, Tektonidou MG, *et al.* Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA* 2001;**286**:821–30.

39. Eggar M, Smith GD, Altman DG. *Systematic reviews in health care. Meta-analysis in context*. London: BMJ Books; 1995.

40. Hyde C, Wake B, Bryan S, Barton P, Fry-Smith A, Davenport C, *et al*. Fludarabine as second-line therapy for B-cell chronic lymphocytic leukaemia: a technology assessment. *Health Technol Assess* 2002;**6**(2).

41. Forbes C, Wilby J, Richardson G, Sculpher M, Mather L, Riemsma R. A rapid and systematic review and economic evaluation of pegylated liposomal doxorubicin hydrochloride for ovarian cancer. *Health Technol Assess* 2002;**6**(23).

42. Vale L Wyness L, McCormack K, Mckenzie L, Brazzelli M, Stearns SC. A systematic review of the effectiveness and cost-effectiveness of metal-on-metal hip resurfacing arthroplasty for treatment of hip disease. *Health Technol Assess* 2002;**6**(15).

43. Peters J, Stevenson M, Beverley C, Lim JNW, Smith S. The clinical effectiveness and cost-effectiveness of inhaler devices used in the routine management of chronic asthma in older children: a systematic review and economic evaluation. *Health Technol Assess* 2002;**6**(5).

44. Vardulaki KA, Bennett-Lloyd BD, Parfitt J, Normond C, Paisley S, Darzi A, *et al*. A systematic review of the effectiveness and cost-effectiveness of laparoscopic surgery for colorectal cancer. A report to NICE. 2000. URL: http://www.nice.org.uk/pdf/htareportonlapsurgcoloreccanc.pdf

45. Wake B, Hyde C, Bryan S, Barton P, Song F, Fry-Smith A, *et al*. Rituximab as third-line treatment for refractory or recurrent stage III or IV follicular non-Hodgkin's lymphoma: a systematic review and economic evaluation. *Health Technol Assess* 2002;**6**(3).

46. Ward S, Morris E, Bansback N, Calvert N, Crellin A, Forman D, *et al*. A rapid and systematic review of the clinical effectiveness and cost-effectiveness of gemcitabine for the treatment of pancreatic cancer. *Health Technol Assess* 2001;**5**(24).

47. Dinnes J, Cave C, Huang S, Major K, Milne R. The effectiveness and cost-effectiveness of temozolomide for the treatment of recurrent malignant glioma: a rapid and systematic review. *Health Technol Assess* 2001;**5**(13).

48. Lewis R, Bagnall A-M, Forbes C, Shirran E, Duffy S, Kleijnen J, *et al*. The clinical effectiveness and cost-effectiveness of trastuzumab for breast cancer: a systematic review *Health Technol Assess* 2002;**6**(13).

49. Fitzpatrick R, Shortall E, Sculpher M, Murray D, Morris R, Lodge M, *et al*. Primary total hip replacement surgery: a systematic review of outcomes and modelling of cost-effectiveness associated with different prostheses. *Health Technol Assess* 1998;**2**(20).

50. Jobanputra P, Parry D, Fry-Smith A, Burls A. Effectiveness of autologous chondrocyte transplantation for hyaline cartilage defects in knees: a rapid and systematic review. *Health Technol Assess* 2001;**5**(11).

51. Bagnall A-M, Jones L, Ginnelly L, Lewis R, Glanville J, Gilbody S, *et al*. A rapid and systematic review of atypical antipsychotic drugs in schizophrenia. *Health Technol Assess* 2003;**7**(13).

52. Woolacott NF, Jones L, Forbes CA, Mather LC, Sowden AJ, Song FJ, *et al*. The clinical effectiveness and cost-effectiveness of bupropion SR and nicotine replacement therapy (NRT) for smoking cessation: a systematic review and economic evaluation. *Health Technol Assess* 2002;**6**(16).

53. Bryant J, Cave C, Mihaylova B, Chase D, Mclntyre L, Gerard K, *et al*. Clinical effectiveness and cost-effectiveness of growth hormone in children: a systematic review and economic evaluation. *Health Technol Assess* 2002;**6**(18).

54. Crombie IK. The standard appraisal questions. In *The pocket guide to critical appraisal.* London: BMJ Publishing Group; 1996. pp. 23–9.

55. Morris RW. A statistical study of papers in the journal of bone and joint surgery (BR) 1984. *J Bone Joint Surg Br* 1988;**70**:242–6.

56. Young J, Fry-Smith A, Hyde C. Lung volume reduction surgery (LVRS) for chronic obstructive pulmonary disease (COPD) with underlying severe emphysema. *Thorax* 1999;**54**:779–89.

57. Dalziel K, Stein K, Round A, Garside R, Royle P. Systematic review of endoscopic sinus surgery for nasal polyps. *Health Technol Assess* 2003;**7**(17).

58. Taylor R, Van Buyten J-P, Buchser E. Spinal cord stimulation for chronic low back pain/failed back surgery syndrome and complex regional pain syndrome – systematic review of the clinical effectiveness and cost-effectiveness literature. 2002.

59. Sculpher MJ, Petticrew M, Kelland JL, Elliott RM, Holdright DR, Buxton MJ. Resource allocation for chronic stable angina: a systematic review of effectiveness, costs and cost-effectiveness of alternative interventions. *Health Technol Assess* 1998;**2**(10).

60. Farrer M, Skinner JS, Albers CJ, Alberti KG, Adams PC. Outcome after coronary artery surgery in women and men in the north of England. *QJM* 1997;**90**:203–11.

61. Skinner JS, Farrer M, Albers CJ, Neil HA, Adams PC. High apolipoprotein AI concentrations are associated with lower mortality and myocardial infarction five years after coronary artery bypass graft surgery. *Heart* 1999;**81**:488–94.

62. Grodstein F, Stampfer MJ, Colditz GA. Postmenopausal hormone therapy and mortality. *N Engl J Med* 1997;**336**:1769–75.

63. Hulley S, Grady D, Bush T, for the Heart and Estrogen/Progestin Replacement Study (HERS) Research Group. Randomised trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. *JAMA* 1998;**280**:605–13.

64. Juni P, Altman D, Egger M. Assessing the quality of randomised controlled trials. In Egger M, Davey-Smith G, Altman D, editors. *Systematic reviews in health care*. London: BMJ Publishing Group; 2000. pp. 87–108.

65. Littenberg B, Weinstein LP, McCarren M, Mead T, Swiontkowski MF, Rudicel SA, *et al*. Closed fractures of the tibial shaft: a meta-analysis of three methods of treatment. *J Bone Joint Surg* 1998;**80–A**:174–83.

66. McAweeney MJ, Tate DG, McAweeney W. Psychosocial interventions in the rehabilitation of people with spinal cord injury: a comprehensive methodological enquiry. *SCI Psychosocial Process* 1997;**10**:58–65.

67. Nielsen ME, Reilly PL. A guide to understanding and evaluating research articles. *Gifted Child Q* 1985;**29**:90–2.

68. Shay CB, Zimmerman WS, Michael WB. The factorial validity of a rating scale for the evaluation of research articles. *Educ Psychol Meas* 1972; **32**:453–7.

69. Sheldon TA, Song F, Davey-Smith G. Critical appraisal of the medical literature: how to assess whether health-care interventions do more harm than good. In Drummond MF, Maynard AE, editors. *Purchasing and providing cost-effective health care*. London: Churchill Livingstone; 1993.

70. Drummond MF, Maynard AE. Applying the principles of cost-effective purchasing. *Purchasing and providing cost-effective health care*. London: Churchill Livingstone; 1993.

71. The Venus Forum of the Royal Society of Medicine and Societas Phlebologica Scandinavia. Methodology of the Task Force. *Phlebology* 1999; 1**4** Suppl 1:3–19

72. Bass JL, Christoffel KK, Widome M, Boyle W, Scheidt P, Stanwick R, *et al*. Childhood injury prevention counseling in primary care settings: a critical review of the literature. *Paediatrics* 1993; **97**:544–50.

73. Brown S. Measurement of quality of primary studies for meta-analysis. *Nurs Res* 1991;**40**:352–5.

74. Bracken M. Reporting observational studies. *Br J Obstet Gynaecol* 1989;**96**:383–8.

75. Brown SA, Upchurch S, Anding R, Winter M, Ramirez G. Promoting weight loss in type II diabetes. *Diabetes Care* 1996;**19**:613–24.

76. Campos-Outcalt D, Senf J, Watkins AJ, Bastacky S. The effects of medical school curricula, faculty role models, and biomedical research support on choice of generalist physician careers: a review and quality assessment of the literature. *Acad Med* 1995;**70**:611–9.

77. Cho MK, Bero L. Instruments for assessing the quality of drug studies published in the medical literature. *JAMA* 1994;**272**:101–4.

78. Cox GL, Merkel WT. A qualitative review of psychosocial treatments for bulimia. *J Nerv Ment Dis* 1989;**177**:77–84.

79. Cuijpers P. Psychological outreach programmes for the depressed elderly: a meta-analysis of effects and dropout. *Int J Geriatr Psychiatry* 1998;1**3:**41–8.

80. Fowkes FGR, Fulton PM. Critical appraisal of published research: introductory guidelines. *BMJ* 1991;**302**:1136–40.

81. Garber *et al.*? Adult respiratory distress syndrome: a systematic overview of incidence and risk factors. *Crit Care Med* 1996;**24**:687–95.

82. Haynes RB, Sackett DL, Taylor DW, Hackett BC, Luterbach E, Cloak JR. An annotated bibliography (including notes on methodologic standards for compliance research). In Sackett DL, Haynes RB, editors. *Compliance with therapeutic regimens*. Baltimore, MD: Johns Hopkins University Press; 1975. pp. 193–278.

83. Kreulen CM, Creugers HJ, Meijering AC. Meta-analysis of anterior veneer restorations in clinical studies. *J Dent* 1998;**26**:345–53.

84. Krogh CL. A checklist system for critical review of medical literature. *Med Educ* 1985;**19**:392–5.

85. Massy ZA, Ma JZ, Louis TA, Kasiske BL. Lipid-lowering therapy in patients with renal disease. *Kidney Int* 1995;**48**:188–98.

86. Meijman FJ, de Melker RA. The extent of inter- and intrareviewer agreement on the classification and assessment of designs of single-practice research. *Fam Pract* 1995;**12**:93–7.

87. Morris SB, Turner CW, Szykula SA. Psychotherapy outcome research: an application of a new method for evaluating research methodology. *Psychotherapy* 1988;**25**:18–26.

88. Rey JM, Walter G. Half a century of ECT use in young people. *Am J Psychiatry* 1997;**154**:595–601.

89. Rowe DE, Bernstein SM, Riddick MF, Adler F, Emans J, Gardner-Bonneau D. A meta-analysis of the efficacy of non-operative treatments for idiopathic scoliosis. *J Bone Joint Surg* 1997; **79A**:664–74.

90. Stock SR. Workplace ergonomic factors and the development of musculoskeletal disorders of the neck and upper limbs: a meta-analysis. *Am J Ind Med* 1991;**19**:87–107.

91. Lawrie GM, Morris GCJ, Calhoon JH, Safi H, Zamora JL, Beltengady M *et al*. Clinical results of coronary bypass in 500 patients at least 10 years after operation. *Circulation* 1982;**66**:I1–I5.

92. ten Berg JM, Voors AA, Suttorp MJ, Ernst SM, Mast EG, Bal E, *et al*. Long-term results after successful percutaneous transluminal coronary angioplasty in patients over 75 years of age. *Am J Cardiol* 1996;**77**:690–5.

93. Kurent Z, Zargi M. Endoscopic polypectomy vs. endoscopic ethmoidectomy in combined treatment of massive bilateral nasal polyposis. In Stammberger H, Wolf G, editors. *XVII European Rhinologic Society & International Symposium of Infection and Allergy of the Nose – ERS & ISIAN Meeting '98*, Vienna, Austria, 28 July–1 August 1998. Bologna: Medimond, Monduzzi Editore International Proceedings Division. pp. 187–90.

94. Penttila M, Rautiainen M, Pukander J, Kataja M. Functional vs. radical maxillary surgery. Failures after functional endoscopic sinus surgery. *Acta Otolaryngol Suppl* 1997;173–6.

95. Venkatachalam VP, Bhat A. Comparative evaluation of endoscopic and conventional surgical techniques in the management of nasal polyposis. *JK Pract* 1998;**5**:295–9.

96. Jankowski R, Pigret D, Decroocq F. Comparison of functional results after ethmoidectomy and nasalization for diffuse and severe nasal polyposis. *Acta Otolaryngol* 1997;**117**:601–8.

97. Unlu HH, Caylan R, Nalca Y, Akyar S. An endoscopic and tomographic evaluation of patients with sinusitis after endoscopic sinus surgery and Caldwell–Luc operation: a comparative study. *J Otolaryngol* 1994;**23**:197–203.

98. Harkness P, Brown P, Fowler S, Topham J. A national audit of sinus surgery. Results of the Royal College of Surgeons of England comparative audit of ENT surgery. *Clin Otolaryngol* 1997; **22**:147–51.

99. Danielsen A. Functional endoscopic sinus surgery on a day case out-patient basis. *Clin Otolaryngol* 1992;**17**:473–7.

100. Danielsen A, Olofsson J. Endoscopic endonasal sinus surgery. A long-term follow-up study. *Acta Otolaryngol* 1996;**116**:611–19.

101. Davis WE, Templer JW, Lamear WR, Davis WE, Craig SB. Middle meatus anstrostomy: patency rates and risk factors. *Otolaryngol Head Neck Surg* 1991;**104**:467–72.

102. Delank KW, Stoll W. Olfactory function after functional endoscopic sinus surgery for chronic sinusitis. *Rhinology* 1998;**36**:15–19.

103. Fortune DS, Duncavage JA. Incidence of frontal sinusitis following partial middle turbinectomy. *Ann Otol Rhinol Laryngol* 1998;1**07**:447–53.

104. Franzen G, Klausen OG. Post-operative evaluation of functional endoscopic sinus surgery with computed tomography. *Clin Otolaryngol* 1994; **19**:332–9.

105. Friedman M, Landsberg R, Schults RA, Tanyeri H, Caldarelli DD. Frontal sinus surgery: endoscopic technique and preliminary results. *Am J Rhinol* 2000;**14**:393–403.

106. Friedman M, Landsberg R, Tanyeri H. Middle turbinate medialization and preservation in endoscopic sinus surgery. *Otolaryngol Head Neck Surg* 2000;**123**:76–80.

107. Frisch T, Arndal H, Fons M. Outcome for the first 85 patients treated with the functional endoscopic sinus surgery technique. *Rhinology* 1995;**33**:236–9.

108. Jacobs JB. 100 years of frontal sinus surgery. *Laryngoscope* 1997;**107**:1–36.

109. Jakobsen J, Svendstrup F. Functional endoscopic sinus surgery in chronic sinusitis – a series of 237 consecutively operated patients. *Acta Otolaryngol Suppl* 2000;**543**:158–61.

110. Jiang RS, Hsu CY. Endoscopic sinus surgery for the treatment of chronic sinusitis in geriatric patients. *Ear Nose Throat J* 2001;**80**:230–2.

111. Katsantonis GP, Friedman WH, Bruns M. Intranasal sphenoethmoidectomy: an evolution of technique. *Otolaryngol Head Neck Surg* 1994; **111**:781–6.

112. Kennedy DW. Prognostic factors, outcomes and staging in ethmoid sinus surgery. *Laryngoscope* 1992;**102**:1–18.

113. Klossek JM, Peloquin L, Friedman WH, Ferrier JC, Fontanel JP. Diffuse nasal polyposis: postoperative long-term results after endoscopic sinus surgery and frontal irrigation. *Otolaryngol Head Neck Surg* 1997;**117**:355–61.

114. Lawson W. The intranasal ethmoidectomy – an experience with 1,077 procedures. *Laryngoscope* 1991;**101**:367–71.

115. Levine HL. Functional endoscopic sinus surgery: evaluation, surgery, and follow-up of 250 patients. *Laryngoscope* 1990;**100**:79–84.

116. Lund VJ, Mackay IS. Outcome assessment of endoscopic sinus surgery. *J R Soc Med* 1994; **87**:70–2.

117. Massegur H, Adema JM, Lluansi J, Fabra JM, Montserrat JM. Endoscopic sinus surgery in sinusitis. *Rhinology* 1995;**33**:89–92.

118. Moses RL, Cornetta A, Atkins JPJ, Roth M, Rosen MR, Keane WM. Revision endoscopic sinus surgery: the Thomas Jefferson University experience. *Ear Nose Throat J* 1998; **77**:190, 193–5, 199–202.

119. Nishioka GJ, Cook PR, Davis WE, McKinsey JP. Functional endoscopic sinus surgery in patients with chronic sinusitis and asthma. *Otolaryngol Head Neck Surg* 1994;**110**:494–500.

120. Park AH, Lau J, Stankiewicz J, Chow J. The role of functional endoscopic sinus surgery in asthmatic patients. *J Otolaryngol* 1998;**27**:275–80.

121. Rice DH. Endoscopic sinus surgery: results at 2-year follow up. *Otolaryngol Head Neck Surg* 1989; **101**:476–9.

122. Roth Y, Shoshan JB, Kronenberg J. Functional endoscopic sinus surgery: experience with the first 100 patients. *Int Surg* 1995;**80**:278–9.

123. Ryan RM, Whittet HB, Norval C, Marks NJ. Minimal follow-up after functional endoscopic sinus surgery. Does it affect outcome? *Rhinology* 1996;**34**:44–5.

124. Sato K, Nakashima T. Endoscopic sinus surgery for chronic sinusitis with antrochoanal polyp. *Laryngoscope* 2000;**110**:1581–3.

125. Schaefer SD. An anatomic approach to endoscopic intranasal ethmoidectomy. *Laryngoscope* 1998; **108**:1628–34.

126. Schaefer SD, Manning S, Close LG. Endoscopic paranasal sinus surgery: indications and considerations. *Laryngoscope* 1989;**99**:1–5.

127. Schaitkin B, May M, Shapiro A, Fucci M, Mester SJ. Endoscopic sinus surgery: 4-year follow-up on the first 100 patients. *Laryngoscope* 1993; **103**:1117–20.

128. Shapshay SM, Rebeiz EE, Pankratov MM. Holmium:yttrium aluminum garnet laser-assisted endoscopic sinus surgery: clinical experience. *Laryngoscope* 1992;**102**:1177–80.

129. Sipila J, Antila J, Suonpaa J. Pre- and postoperative evaluation of patients with nasal obstruction undergoing endoscopic sinus surgery. *Eur Arch Otorhinolaryngol* 1996;**253**:237–9.

130. Sobol SE, Wright ED, Frenkiel S. One-year outcome analysis of functional endoscopic sinus surgery for chronic sinusitis. *J Otolaryngol* 1998; **27**:252–7.

131. Stammberger H, Posawetz W. Functional endoscopic sinus surgery. Concept, indications and results of the Messerklinger technique. *Eur Arch Otorhinolaryngol* 1990;**247**:63–76.

132. Stoop AE, van der Heijden HA, Biewenga J, van der Baan S. Clinical aspects and distribution of immunologically active cells in the nasal mucosa of patients with nasal polyps after endoscopic sinus surgery and treatment with topical corticosteroids. *Eur Arch Otorhinolaryngol* 1992; **249**:313–17.

133. Venkatachalam VP, Bhat A. Functional endoscopic sinus surgery – a newer surgical concept in the

management of chronic sinusitis. *Indian J Otolaryngol Head Neck Surg* 1999;**52**:13–16.

134. Vleming M, de Vries N. Endoscopic sinus surgery for antrochoanal polyps. *Rhinology* 1991;**29**:77–8.

135. Weber R, Draf W, Keerl R, Schick B, Saha A. Endonasal microendoscopic pansinus operation in chronic sinusitis. II. Results and complications. *Am J Otolaryngol* 1997;**18**:247–53.

136. Wigand ME, Hosemann W. Microsurgical treatment of recurrent nasal polyposis. *Rhinol Suppl* 1989; **8**:25–9.

137. Wigand ME, Steiner W, Jaumann MP. Endonasal sinus surgery with endoscopical control: from radical operation to rehabilitation of the mucosa. *Endoscopy* 1978;**10**:255–60.

138. Wolf G, Greistorfer K, Jebeles JA. The endoscopic endonasal surgical technique in the treatment of chronic recurring sinusitis in children. *Rhinology* 1995;**33**:97–103.

139. Writing Group for the Bypass Angioplasty Revascularization Investigation (BARI) Investigators. Five-year clinical and functional outcome comparing bypass surgery and angioplasty in patients with multivessel coronary disease. A multicenter randomized trial. *JAMA* 1997;**277**:715–21.

140. CABRI Trial Participants. First-year results of CABRI (Coronary Angioplasty versus Bypass Revascularisation Investigation). *Lancet* 1995; **346**:1179–84.

141. Rodriguez A, Mele E, Peyregne E, Bullon F, Perez-Balino N, Liprandi MI, *et al*. Three-year follow-up of the Argentine Randomized Trial of Percutaneous Transluminal Coronary Angioplasty Versus Coronary Artery Bypass Surgery in Multivessel Disease (ERACI). *J Am Coll Cardiol* 1996;**27**:1178–84.

142. Goy JJ, Eeckhout E, Burnand B, Vogt P, Stauffer JC, Hurni M, *et al*. Coronary angioplasty versus left internal mammary artery grafting for isolated proximal left anterior descending artery stenosis. *Lancet* 1994;**343**:1449–53.

143. Hamm CW, Reimers J, Ischinger T, Rupprecht HJ, Berger J, Bleifeld W. A randomized study of coronary angioplasty compared with bypass surgery in patients with symptomatic multivessel coronary disease. German Angioplasty Bypass Surgery Investigation (GABI). *N Engl J Med* 1994; **331**:1037–43.

144. Hueb WA, Bellotti G, de Oliveira SA, Arie S, de Albuquerque CP, Jatene AD, *et al*. The Medicine, Angioplasty or Surgery Study (MASS): a prospective, randomized trial of medical therapy, balloon angioplasty or bypass surgery for single proximal left anterior descending artery stenoses. *J Am Coll Cardiol* 1995;**26**:1600–5.

145. Jones EL, Weintraub WS. The importance of completeness of revascularization during long-term follow-up after coronary artery operations. *J Thorac Cardiovasc Surg* 1996;**112**:227–37.

146. 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 versus Surgery Trial (EAST). *N Engl J Med* 1994; **331**:1044–50.

147. Puel J, Karouny E, Marco F. Angioplasty versus surgery in multi vessel disease: immediate results and in hospital outcome in a randomised prospective study. *Circulation* 1992;**86**:372.

148. Coronary angioplasty versus coronary artery bypass surgery: the Randomized Intervention Treatment of Angina (RITA) trial. *Lancet* 1993; **341**:573–80.

149. Bonnier H, de Vries C, Michels R, el Gamal M. Initial and long-term results of coronary angioplasty and coronary bypass surgery in patients of 75 or older. *Br Heart J* 1993;**70**:122–5.

150. Mick MJ, Simpfendorfer C, Arnold AZ, Piedmonte M, Lytle BW. Early and late results of coronary angioplasty and bypass in octogenarians. *Am J Cardiol* 1991;**68**:1316–20.

151. Tyras DH, Barner HB, Kaiser CA, Codd JE, Pennington DG, Willman VL. Bypass grafts to the left anterior descending coronary artery: saphenous vein grafts versus internal mammary artery. *J Thorac Cardiovasc Surg* 1980;**80**:327–33.

152. Ullyot DJ, Wisneski J, Sullivan RW, Gertz EW. Improved survival after coronary artery surgery in patients with extensive coronary artery disease. *J Thorac Cardiovasc Surg* 1975;**70**:405–13.

153. Acar C, Ramsheyi A, Pagny JY, Jebara V, Barrier P, Fabiani JN *et al*. The radial artery for coronary artery bypass grafting: clinical and angiographic results at five years. *J Thorac Cardiovasc Surg* 1998; **116**:981–9.

154. Acinapura AJ, Rose DM, Jacobowitz IJ, Kramer MD, Robertazzi RR, Feldman J, *et al*. Internal mammary artery bypass grafting: influence on recurrent angina and survival in 2,100 patients. *Ann Thorac Surg* 1989;**48**:186–91.

155. Acinapura AJ, Jacobowitz IJ, Kramer MD, Zisbrod Z, Cunningham JN. Internal mammary artery bypass: thirteen years of experience. Influence of angina and survival in 5125 patients. *J Cardiovasc Surg (Torino)* 1992;**33**:554–9.

156. Arnold M, Wathanacharoen S, Reed WA, Killen DA, Crockett J, King JT, *et al*. Five-year follow-up of patients undergoing coronary artery bypass. *Ann Thorac Surg* 1979;**27**:225–9.

157. Ashor GW, Meyer BW, Lindesmith GG, Stiles QR, Walker GH, Tucker BL. Coronary artery disease:

surgery in 100 patients 65 years of age and older. *Arch Surg* 1973;**107**:30–3.

158. Azariades M, Fessler CL, Floten HS, Starr A. Five-year results of coronary bypass grafting for patients older than 70 years: role of internal mammary artery. *Ann Thorac Surg* 1990;**50**: 940–5.

159. Baldwin DR, Slaughter MS, Park S, McFalls E, Ward HB. Coronary bypass grafting for single-vessel coronary artery disease: a 17-year review with short- and long-term follow-up. *Chest* 1998; **113**:676–80.

160. Barner HB, Standeven JW, Reese J. Twelve-year experience with internal mammary artery for coronary artery bypass. *J Thorac Cardiovasc Surg* 1985;**90**:668–75.

161. Bathgate AJ, Irving JB. Ten year follow up of patients referred for coronary artery bypass grafting from a single district general hospital. *Heart* 1997;**78**:584–6.

162. Bell MR, Gersh BJ, Schaff HV, Holmes DRJ, Fisher LD, Alderman EL, *et al*. Effect of completeness of revascularization on long-term outcome of patients with three-vessel disease undergoing coronary artery bypass surgery. A report from the Coronary Artery Surgery Study (CASS) Registry. *Circulation* 1992;**86**:446–57.

163. Beretta L, Lemma M, Vanelli P, Botta M, Antonacci C, Bevilacqua M, *et al*. Gastroepiploic artery free graft for coronary bypass. *Eur J Cardiothorac Surg* 1990;**4**:323–7.

164. Bergsma TM, Grandjean JG, Voors AA, Boonstra PW, den Heyer P, Ebels T. Low recurrence of angina pectoris after coronary artery bypass graft surgery with bilateral internal thoracic and right gastroepiploic arteries. *Circulation* 1998;**97**:2402–5.

165. Brandrup-Wognsen G, Haglid M, Karlsson T, Berggren H, Herlitz BJ. Preoperative risk indicators of death at an early and late stage after coronary artery bypass grafting. *Thorac Cardiovasc Surg* 1995;**43**:77–82.

166. Cameron AA, Green GE, Brogno DA, Thornton J. Internal thoracic artery grafts: 20-year clinical follow-up. *J Am Coll Cardiol* 1995;**25**:188–92.

167. Canver CC, Nichols RD, Cooler SD, Heisey DM, Murray EL, Kroncke GM. Influence of increasing age on long term survival after coronary artery bypass grafting. *Ann Thorac Surg* 1996;**62**:1123–7.

168. Carter MJ. The use of the right gastro-epiploic artery in coronary artery bypass grafting. *Aust N Z J Surg* 1987;**57**:317–21.

169. Christakis GT, Rao V, Fremes SE, Chen E, Naylor CD, Goldman BS. Does coronary endarterectomy adversely affect the results of bypass surgery? *J Card Surg* 1993;**8**:72–8.

170. Christenson JT, Schmuziger M. Sequential venous bypass grafts: results 10 years later. *Ann Thorac Surg* 1997;**63**:371–6.

171. Egstrup K. Asymptomatic myocardial ischemia as a predictor of cardiac events after coronary artery bypass grafting for stable angina pectoris. *Am J Cardiol* 1988;**61**:248–52.

172. Fitzgibbon GM, Kafka HP, Leach AJ, Keon WJ, Hooper GD, Burton JR. Coronary bypass graft fate and patient outcome: angiographic follow-up of 5,065 grafts related to survival and reoperation in 1,388 patients during 25 years. *J Am Coll Cardiol* 1996;**28**:616–26.

173. French JK, Scott DS, Whitlock RM, Nisbet HD, Vedder M, Kerr AR, *et al*. Late outcome after coronary artery bypass graft surgery in patients <40 years old. *Circulation* 1995;**92**:II14–II19.

174. Gale AW, Shanahan MX, Chang VP, Windsor HM. Coronary artery surgery. *Med J Aust* 1977;**2**:271–6.

175. Gelbfish J, Jacobowitz IJ, Rose DM, Connolly MW, Acinapura AJ, Zisbrod Z, *et al*. Cryopreserved homologous saphenous vein: early and late patency in coronary artery bypass surgical procedures. *Ann Thorac Surg* 1986;**42**:70–3.

176. Gelfand ET, Wartak J, Callaghan JC. Coronary artery bypass in patients under 40 years of age. *Can J Surg* 1983;**26**:188–9.

177. Green GE, Kemp HG, Alam SE, Pierson RNJ, Friedman MI, David I. Coronary bypass surgery. Five-year follow-up of a consecutive series of 140 patients. *J Thorac Cardiovasc Surg* 1979;**77**:48–56.

178. Higginbothom M, Hunt D, White A, Clarebrough J. Surgical treatment of angina pectoris in the elderly. *Med J Aust* 1981;**2**:664–6.

179. Horgan JH, Teo KK, Darby JF, Teehan M, Harte M, Gallagher T, *et al*. Coronary artery bypass graft surgery: initial therapeutic response. *Ir Med J* 1981;**74**:353–5.

180. Ivert T, Lindblom D, Welti R. Coronary artery bypass grafting in patients 70 years of age and older. Early and late results. *Eur J Cardiothorac Surg* 1989;**3**:52–7.

181. Jenkins CD, Stanton BA, Savageau JA, Denlinger P, Klein MD. Coronary artery bypass surgery. Physical, psychological, social, and economic outcomes six months later. *JAMA* 1983;**250**:782–8.

182. Killen DA, Reed WA, Conn R, Harbrecht J, Arnold M. Myocardial revascularization: results relative to extensiveness. *South Med J* 1982; **75**:1525–30.

183. Killen DA, Arnold M, McConahay DR, Wathanacharoen S, Reed WA. Fifteen-year results of coronary artery bypass for isolated left anterior descending coronary artery disease. *Ann Thorac Surg* 1989;**47**:595–9.

184. Killen DA, Wathanacharoen S, Reed WA, Piehler JM, Borkon AM, Gorton ME, *et al*. Coronary artery bypass for isolated disease of the left anterior descending artery. Late survival of 648 patients. *Tex Heart Inst J* 1998;**25**:181–4.

185. Kornfeld DS, Heller SS, Frank KA, Wilson SN, Malm JR. Psychological and behavioral responses after coronary artery bypass surgery. *Circulation* 1982;**66**:III24–III28.

186. Laird-Meeter K, van Domburg R, van den Brand MJ, Lubsen J, Bos E, Hugenholtz PG. Incidence, risk, and outcome of reintervention after aortocoronary bypass surgery. *Br Heart J* 1987;**57**:427–35.

187. Laks H, Kaiser GC, Barner HB, Codd JE, Willman VI. Coronary revascularization under age 40 years. Risk factors and results of surgery. *Am J Cardiol* 1978;**41**:584–9.

188. Liao Y, Cooper RS, Ghali JK, Szocka A. Survival rates with coronary artery disease for black women compared with black men. *JAMA* 1992; **268**:1867–71.

189. Lytle BW, Kramer JR, Golding LR, Cosgrove DM, Borsh JA, Goormastic M, *et al*. Young adults with coronary atherosclerosis: 10 year results of surgical myocardial revascularization. *J Am Coll Cardiol* 1984;**4**:445–53.

190. Maddern GJ, Craddock D, Leppard PI, Ross IK, Stubberfield J, Waddy JL. Twelve years of coronary artery surgery in South Australia. *Med J Aust* 1984; **140**:136-40.

191. MacDonald P, Stadnyk K, Cossett J, Klassen G, Johnstone D, Rockwood K. Outcomes of coronary artery bypass surgery in elderly people. *Can J Cardiol* 1998;**14**:1215–22.

192. Morin JE, Hedderich G, Poirier NL, Sampalis J, Symes JF. Coronary artery bypass using internal mammary artery branches. *Ann Thorac Surg* 1992;**54**:911–14.

193. Morris JJ, Smith LR, Glower DD, Muhlbaier LH, Reves JG, Wechsler AS, *et al*. Clinical evaluation of single versus multiple mammary artery bypass. *Circulation* 1990;**82**:IV214–IV223.

194. Mullany CJ, Darling GE, Pluth JR, Orszulak TA, Schaff HV, Ilstrup DM, *et al*. Early and late results after isolated coronary artery bypass surgery in 159 patients aged 80 years and older. *Circulation* 1990;**82**:IV229–IV236.

195. Nicholson IA, Paterson HS. Modified T graft for triple-vessel disease. *Ann Thorac Surg* 1997; **64**:451–3.

196. Ochsner JL, Moseley PW, Mills NL, Bower PJ. Long-term follow-up of internal mammary artery myocardial implantation. *Ann Thorac Surg* 1977; **23**:118–21.

197. Palatianos GM, Bolooki H, Horowitz MD, Lowery MH, Rosenthal SP, Chandarlapaty SK, *et al*. Sequential internal mammary artery grafts for coronary artery bypass. *Ann Thorac Surg* 1993; **56**:1136–40.

198. Patel D, Mulcahy D, Curzen N, Sullivan A, Cunningham D, Sparrow J, *et al*. Prognostic significance of transient ST segment changes after coronary artery bypass surgery: a long-term (4–10 year) follow up study. *Br Heart J* 1993;**70**:337–41.

199. Peterson ED, Cowper PA, Jollis JG, Bebchuk JD, DeLong ER, Muhlbaier LH, *et al*. Outcomes of coronary artery bypass graft surgery in 24461 patients aged 80 and older. *Circulation* 1995; **92**:85–91.

200. Pinna-Pintor P, Torta R, Bartolozzi S, Borio R, Caruzzo E, Cicolin A, *et al*. Clinical outcome and emotional-behavioural status after isolated coronary surgery. *Qual Life Res* 1992;**1**:177–85.

201. Rahimtoola SH, Fessler CL, Grunkemeier G, Starr A. Survival 15 to 20 years after coronary artery bypass surgery for angina. *J Am Coll Cardiol* 1993;**21**:151–7.

202. Rahimtoola SH, Bennet A.J., Grunkemeier GL, Block P, Starr A. Survival at 15 to 18 years after coronary bypass surgery for angina in women. *Circulation* 1993;**88**:II71–II78.

203. Richardson JV, Cyrus R.J. Reduced efficacy of coronary artery bypass in women. *Ann Thorac Surg* 1986;**42**:S16–S21.

204. Risum O, Abdelnoor M, Svennevig JL, Nitter-Hauge S, Levorstad K, Foerster A, *et al*. Risk factors for early and late mortality in surgical treatment of coronary artery disease. *Cardiovasc Surg* 1995;**3**:537–44.

205. Risum O, Abdelnoor M, Svennevig JL, Levorstad K, Gullestad L, Bjornerheim R, *et al*. Diabetes mellitus and morbidity and mortality risks after coronary artery bypass surgery. *Scand J Thorac Cardiovasc Surg* 1996;**30**:71–5.

206. Ruygrok PN, Agnew TM, Coverdale HA, Kerr AR, Graham KJ, Whitlock RM. Coronary artery surgery in the elderly: long-term follow-up. *Aust N Z J Med* 1993;**23**:489–93.

207. Saatvedt K, Froysaker T, Vatne K, Nordstrand K. Retrograde internal mammary artery as coronary bypass. A prospective study with postoperative angiographic evaluation. *Scand Cardiovasc J* 1997; **31**:101–3.

208. Salomon NW, Page US, Bigelow JC, Krause AH, Okies JE, Metzdorff MT. Reoperative coronary surgery. Comparative analysis of 6591 patients undergoing primary bypass and 508 patients undergoing reoperative coronary artery bypass. *J Thorac Cardiovasc Surg* 1990;**100**:250–9.

209. Schaff HV, Gersh BJ, Pluth JR, Danielson GK, Orszulak TA, Puga FJ, *et al*. Survival and functional status after coronary artery bypass grafting: results 10 to 12 years after surgery in 500 patients. *Circulation* 1983;**68**:II200–II204.

210. Schmuziger M, Christenson JT, Maurice J, Mosimann E, Simonet F, Velebit V. Reoperative myocardial revascularization: an analysis of 458 reoperations and 2645 single operations. *Cardiovasc Surg* 1994;**2**:623–9.

211. Sheldon WC, Loop FD. Coronary artery bypass surgery. The Cleveland Clinic Experience 1967–1982. *Postgrad Med* 1984;**75**:108–21.

212. Simmons BE, Castaner A, Santhanam V, Ghali J, Silverman NA, Goldfaden DM, *et al*. Outcome of coronary artery bypass grafting in black persons. *Am J Cardiol* 1987;**59**:547–51.

213. Sterling RP, Graeber GM, Albus RA, Burton NA, Lough FC, Fleming AW. Results of myocardial revascularization in black males. *Am Heart J* 1984; **108**:695–9.

214. Tector AJ, Schmahl TM, Canino VR. Expanding the use of the internal mammary artery to improve patency in coronary artery bypass grafting. *J Thorac Cardiovasc Surg* 1986;**91**:9–16.

215. Tschan W, Hoffmann A, Burkart F, Meier C, Burckhardt D. Coronary artery bypass grafts. Influence of preoperative risk factors on the late postoperative course. *Chest* 1985;**88**:185–9.

216. Ullyot DJ, Wisneski J, Sullivan RW, Gertz EW, Ryan C. The impact of coronary artery bypass on late myocardial infarction. *J Thorac Cardiovasc Surg* 1977;**73**:165–75.

217. Verhiest W, Piessens J, Suy R, Kesteloot H, De Geest H. Initial experience with coronary artery bypass graft surgery. A follow-up study. *Acta Cardiol* 1977;**32**:157–86.

218. Weintraub WS, Jones EL, Craver JM, Grosswald R, Guyton RA. In-hospital and long-term outcome after reoperative coronary artery bypass graft surgery. *Circulation* 1995;**92**:II50–II57.

219. Wright JS. Late survival after coronary surgery. *Med J Aust* 1979;**2**:615–16.

220. Cohen DJ, Basamania C, Graeber GM, Deshong JL, Burge JR. Coronary artery bypass grafting in young patients under 36 years of age. *Chest* 1986;**89**:811–16.

221. Horneffer PJ, Gardner TJ, Manolio TA, Hoff SJ, Rykiel MF, Pearson TA, *et al*. The effects of age on outcome after coronary bypass surgery. *Circulation* 1987;**76**:V6–V12.

222. Cavallini C, Risica G, Olivari Z, Marton F, Franceschini E, Giommi L. Clinical and angiographic follow-up after coronary angioplasty in patients with two-vessel disease: influence of completeness and adequacy of revascularization on long-term outcome. *Am Heart J* 1994;**127**:1504–9.

223. Jeroudi MO, Kleiman NS, Minor ST, Hess KR, Lewis JM, Winters WLJ, *et al*. Percutaneous transluminal coronary angioplasty in octogenarians. *Ann Intern Med* 1990;**113**:423–8.

224. Kamp O, Beatt KJ, de Feyter PJ, van den Brand M, Suryapranata H, Luijten HE, *et al*. Short-, medium- and long-term follow-up after percutaneous transluminal coronary angioplasty for stable and unstable angina pectoris. *Am Heart J* 1989;**117**:991–6.

225. Meyer J, Schmitz HJ, Kiesslich T, Erbel R, Krebs W, Schulz W, *et al*. Percutaneous transluminal coronary angioplasty in patients with stable and unstable angina pectoris: analysis of early and late results. *Am Heart J* 1983;**106**:973–80.

226. Perry RA, Seth A, Singh A, Shiu MF. Success and complication rates of coronary angioplasty in patients with and without previous myocardial infarction. *Eur Heart J* 1988;**9**:37–42.

227. Simpfendorfer C, Raymond R, Schraider J, Badhwar K, Dorosti K, Franco I, *et al*. Early and long-term results of percutaneous transluminal coronary angioplasty in patients 70 years of age and older with angina pectoris. *Am J Cardiol* 1988; **62**:959–61.

228. Thomas ES, Most AS, Williams DO. Coronary angioplasty for patients with multivessel coronary artery disease: follow-up clinical status. *Am Heart J* 1988;**115**:8–13.

229. Thompson RC, Holmes DRJ, Gersh BJ, Mock MB, Bailey KR. Percutaneous transluminal coronary angioplasty in the elderly: early and long-term results. *J Am Coll Cardiol* 1991;**17**:1245–50.

230. Anderson MH, Ward DE. Early experience with low speed rotational angioplasty. *Br Heart J* 1991; **66**:130–3.

231. Arnold AM, Mick MJ, Piedmonte MR, Simpfendorfer C. Gender differences for coronary angioplasty. *Am J Cardiol* 1994;**74**:18–21.

232. Bell MR, Grill DE, Garratt KN, Berger PB, Gersh BJ, Holmes DRJ. Long-term outcome of women compared with men after successful coronary angioplasty. *Circulation* 1995;**91**:2876–81.

233. Bentivoglio LG, Holubkov R, Kelsey SF, Holmes DRJ, Sopko G, Cowley MJ, *et al*. Short and long term outcome of percutaneous transluminal coronary angioplasty in unstable versus stable angina pectoris: a report of the 1985–1986 NHLBI PTCA Registry. *Catheter Cardiovasc Diagn* 1991;**23**:227–38.

234. Berger E, Williams DO, Reinert S, Most AS. Sustained efficacy of percutaneous transluminal coronary angioplasty. *Am Heart J* 1986;**111**:233–6.

235. Buffet P, Danchin N, Juilliere Y, Feldmann L, Marie PY, Selton-Suty C, *et al*. Percutaneous transluminal coronary angioplasty in patients more than 75 years old: early and long-term results. *Int J Cardiol* 1992;**37**:33–9.

236. Buffet P, Colasante B, Feldmann L, Danchin N, Juilliere Y, Anconina J, *et al*. Long-term follow-up after coronary angioplasty in patients younger than 40 years of age. *Am Heart J* 1994;**127**:509–13.

237. Burton JR, Haraphongse M, Hsu L, Kappagoda CT, Rossall RE, Schlaut B, *et al*. Risk stratification after percutaneous transluminal coronary angioplasty. *Cardiovasc Drugs Ther* 1990;**4**:687–93.

238. Ciampricotti R, el Gamal M, Van Gelder B, Bonnier J, Taverne R. Coronary angioplasty of bifurcational lesions without protection of large side branches. *Catheter Cardiovasc Diagn* 1992;**27**:191–6.

239. Cowley MJ, Vetrovec GW, DiSciascio G, Lewis SA, Hirsh PD, Wolfgang TC. Coronary angioplasty of multiple vessels: short-term outcome and long-term results. *Circulation* 1985;**72**:1314–20.

240. Cowley MJ, Vandermael M, Topol EJ, Whitlow PL, Dean LS, Bulle TM, *et al*. Is traditionally defined complete revascularization needed for patients with multivessel disease treated by elective coronary angioplasty? Multivessel Angioplasty Prognosis Study (MAPS) Group. *J Am Coll Cardiol* 1993;**22**:1289–97.

241. de Jaegere P, de Feyter P, van Domburg R, Suryapranata H, van den Brand M, Serruys PW. Immediate and long term results of percutaneous coronary angioplasty in patients aged 70 and over. *Br Heart J* 1992;**67**:138–43.

242. Deligonul U, Vandormael MG, Kern MJ, Zelman R, Galan K, Chaitman BR. Coronary angioplasty: a therapeutic option for symptomatic patients with two and three vessel coronary disease. *J Am Coll Cardiol* 1988;**11**:1173–9.

243. Dorros G, Janke LM. Coronary angioplasty in patients who undergo multiple vessel angioplasties, and in patients who have multivessel coronary disease. *Herz* 1985;**10**:281–91.

244. Dorros G, Lewin RF, Mathiak LM, Johnson WD, Brenowitz J, Schmahl T, *et al*. Percutaneous transluminal coronary angioplasty in patients with two or more previous coronary artery bypass grafting operations. *Am J Cardiol* 1988;**61**:1243–7.

245. Ellis SG, Cowley MJ, DiSciascio G, Deligonul U, Topol EJ, Bulle TM, *et al*. Determinants of 2-year outcome after coronary angioplasty in patients with multivessel disease on the basis of comprehensive preprocedural evaluation. Implications for patient selection. The Multivessel Angioplasty Prognosis Study Group. *Circulation* 1991;**83**:1905–14.

246. Ellis CJ, French JK, White HD, Ormiston JA, Whitlock RM, Webster MW. Results of percutaneous coronary angioplasty in patients <40 years of age. *Am J Cardiol* 1998;**82**:135–9.

247. Ernst SM, van der Feltz TA, Bal ET, van Bogerijen L, van den Berg E, Ascoop CA, *et al*. Long-term angiographic follow up, cardiac events, and survival in patients undergoing percutaneous transluminal coronary angioplasty. *Br Heart J* 1987;**57**:220–5.

248. el Gaylani N, McAdam BF, White U, Gearty GF, Walsh MJ, Crean PA. Immediate and follow-up results of coronary angioplasty – lessons for the future. *Ir Med J* 1996;**89**:60–1.

249. Glazier JJ, Williams MG, Madden S, Rickards AF. Clinical outcome following coronary balloon angioplasty in 100 consecutive patients with multivessel coronary artery disease. *J R Coll Physicians London* 1990;**24**:292–4.

250. Grigg LE, Valentine PA, Manolas EG, Flower DJ, Hunt D. Long-term clinical and angiographic results following percutaneous transluminal coronary angioplasty. *Aust N Z J Med* 1988; **18**:689–92.

251. Gurbel PA, Anderson RD, Peels HO, van Boven AJ, Den Heijer P. Coronary artery angioplasty with a helical autoperfusion balloon catheter. *Catheter Cardiovasc Diagn* 1997;**40**:179–85.

252. Henderson RA, Karani S, Dritsas A, Sowton E. Long-term results of coronary angioplasty for single vessel, proximal, left anterior descending disease. *Eur Heart J* 1991;**12**:642–7.

253. Holmes DRJ, Vlietstra RE, Smith HC, Vetrovec GW, Kent KM, Cowley MJ, *et al*. Restenosis after percutaneous transluminal coronary angioplasty (PTCA): a report from the PTCA Registry of the National Heart, Lung, and Blood Institute. *Am J Cardiol* 1984;**53**:77C–81C.

254. Ilsley C, Wilkins G, Amarasingham R, Restieaux N, Ablett M. Percutaneous transluminal coronary angioplasty: the early Dunedin experience. *N Z Med J* 1985;**98**:47–9.

255. Ivanhoe RJ, Weintraub WS, Douglas JSJ, Lembo NJ, Furman M, Gershony G, *et al*. Percutaneous transluminal coronary angioplasty of chronic total occlusions. Primary success, restenosis, and long-term clinical follow-up. *Circulation* 1992;**85**:106–15.

256. Jost S, Nolte CW, Simon R, Amende I, Gulba DC, Wiese B, *et al*. Angioplasty of subacute and chronic total coronary occlusions: success, recurrence rate, and clinical follow-up. *Am Heart J* 1991; **122**:1509–14.

257. Kelsey SF, James M, Holubkov AL, Holubkov R, Cowley MJ, Detre KM. Results of percutaneous transluminal coronary angioplasty in women. 1985–1986 National Heart, Lung, and Blood Institute's Coronary Angioplasty Registry. *Circulation* 1993;**87**:720–7.

258. King SB, Schlumpf M. Ten-year completed follow-up of percutaneous transluminal coronary angioplasty: the early Zurich experience. *J Am Coll Cardiol* 1993;**22**:353–60.

259. Kofflard MJ, de Jaegere PP, van Domburg R, Ruygrok P, van den Brand M, Serruys PW, *et al*. Immediate and long-term clinical outcome of coronary angioplasty in patients aged 35 years or less. *Br Heart J* 1995;**73**:82–6.

260. Krajcer Z, Boskovic D, Angelini P, Leatherman LL, Springer A, Leachman RD. Transluminal coronary angioplasty of the right coronary artery brachial cutdown approach. *Catheter Cardiovasc Diagn* 1982;**8**:553–64.

261. Leisch F, Schutzenberger W, Kerschner K, Herbinger W. Influence of a variant angina on the results of percutaneous transluminal coronary angioplasty. *Br Heart J* 1986;**56**:341–5.

262. Maiello L, Colombo A, Gianrossi R, Thomas J, Finci L. Percutaneous transluminal coronary angioplasty in patients aged 70 years and older: immediate and long-term results. *Int J Cardiol* 1992;**36**:1–8.

263. Mata LA, Bosch X, David PR, Rapold HJ, Corcos T, Bourassa MG. Clinical and angiographic assessment 6 months after double vessel percutaneous coronary angioplasty. *J Am Coll Cardiol* 1985;**6**:1239–44.

264. Melchior JP, Meier B, Urban P, Finci L, Steffenino G, Noble J, *et al*. Percutaneous transluminal coronary angioplasty for chronic total coronary arterial occlusion. *Am J Cardiol* 1987;**59**:535–8.

265. Morton BC, Higginson LA, Chambers RJ, Beanlands DS. Late reassessment of percutaneous transluminal coronary angioplasty results: clinical, exercise and angiographic follow-up. *Can J Cardiol* 1989;**5**:291–4.

266. Myler RK, Topol EJ, Shaw RE, Stertzer SH, Clark DA, Fishman J, *et al*. Multiple vessel coronary angioplasty: classification, results, and patterns of restenosis in 494 consecutive patients. *Catheter Cardiovasc Diagn* 1987;**13**:1–15.

267. Piovaccari G, Fattori R, Marzocchi A, Marrozzini C, Prati F, Magnani B. Percutaneous transluminal coronary angioplasty of the very proximal left anterior descending artery lesions: immediate results and follow-up. *Int J Cardiol* 1991;**30**:151–5.

268. Richardson M, Pitney MR, Gibbons F, Cope GD, Cumpston GN, Mews GC. Short and long-term results of coronary angioplasty in patients over 75 years. *Aust N Z J Med* 1994;**24**:55–60.

269. Ruygrok PN, de Jaegere PT, van Domburg RT, van den Brand MJ, Serruys PW, de Feyter PJ. Clinical outcome 10 years after attempted percutaneous transluminal coronary angioplasty in 856 patients. *J Am Coll Cardiol* 1996;**27**:1669–77.

270. Ruygrok P, Barron G, Ormiston J, Meredith I, Webster M, Harper R, *et al*. Introduction of the Multilink stent into routine angioplasty practice: early angiographic and clinical outcome. *Catheter Cardiovasc Diagn* 1998;**43**:147–52.

271. Safian RD, Grines CL, May MA, Lichtenberg A, Juran N, Schreiber TL, *et al*. Clinical and angiographic results of transluminal extraction coronary atherectomy in saphenous vein bypass grafts. *Circulation* 1994;**89**:302–12.

272. Sahni R, Maniet AR, Banka VS. Long-term efficacy of percutaneous transluminal angioplasty on incidence of myocardial infarction, relief of symptoms and survival. *Clin Cardiol* 1989; **12**:427–31.

273. Scott SM, Deupree RH, Sharma GV, Luchi RJ. VA study of unstable angina. 10-year results show duration of surgical advantage for patients with impaired ejection fraction. *Circulation* 1994; **90**:II120–II123.

274. Simpfendorfer C, Tuzcu EM, Badhwar K, Dorosti K, Franco I, Hollman J, *et al*. Percutaneous transluminal coronary angioplasty in the young adult. *Cleve Clin J Med* 1989;**56**:569–72.

275. Stammen F, Piessens J, Vrolix M, Glazier JJ, De Geest H, Willems JL. Immediate and short-term results of a 1988–1989 coronary angioplasty registry. *Am J Cardiol* 1991;**67**:253–8.

276. Stein B, Weintraub WS, Gebhart SP, Cohen-Bernstein CL, Grosswald R, Liberman HA, *et al*. Influence of diabetes mellitus on early and late outcome after percutaneous transluminal coronary angioplasty. *Circulation* 1995;**91**:979–89.

277. Suryapranata H, Hoorntje JC, de Boer MJ, Zijlstra F. Coronary angioplasty with monorail technique: experience in more than 2000 cases. *Cor Vasa* 1993;**35**:188–90.

278. Talley JD, Hurst JW, King SB, Douglas JSJ, Roubin GS, Gruentzig AR, *et al*. Clinical outcome 5 years after attempted percutaneous transluminal coronary angioplasty in 427 patients. *Circulation* 1988;**77**:820–9.

279. Tan KH, Sulke N, Taub N, Karani S, Sowton E. Percutaneous transluminal coronary angioplasty in patients 70 years of age or older: 12 years' experience. *Br Heart J* 1995;**74**:310–17.

280. Thompson RC, Holmes DRJ, Gersh BJ, Bailey KR. Predicting early and intermediate-term outcome of coronary angioplasty in the elderly. *Circulation* 1993;**88**:1579–87.

281. Urban P, Fox K, Crean P, Shapiro L, Rickards A. Coronary balloon angioplasty for elderly patients with severe angina. *Br Heart J* 1987;**58**:465–8.

282. Valentine PA, Manolas EG. Balloon dilatation of coronary arteries. Four years' experience of coronary angioplasty. *Med J Aust* 1984;**140**:700–5.

283. Vandormael M, Deligonul U, Taussig S, Kern MJ. Predictors of long-term cardiac survival in patients with multivessel coronary artery disease undergoing percutaneous transluminal coronary angioplasty. *Am J Cardiol* 1991;**67**:1–6.

284. Voudris V, Antonellis J, Salachas A, Ifantis G, Sionis D, Margaris N, *et al*. Coronary angioplasty in the elderly: immediate and long-term results. *Angiology* 1993;**44**:933–7.

285. Webb JG, Myler RK, Shaw RE, Anwar A, Stertzer SH. Coronary angioplasty in young adults: initial results and late outcome. *J Am Coll Cardiol* 1990;**16**:1569–74.

286. Weintraub WS, Ghazzal ZM, Douglas JSJ, Liberman HA, Morris DC, Cohen CL, *et al*. Long-term clinical follow-up in patients with angiographic restudy after successful angioplasty. *Circulation* 1993;**87**:831–40.

287. Weintraub WS, Wenger NK, Kosinski AS, Douglas JSJ, Liberman HA, Morris DC, *et al*. Percutaneous transluminal coronary angioplasty in women compared with men. *J Am Coll Cardiol* 1994;**24**:81–90.

288. Weintraub WS, King SB, Douglas JSJ, Kosinski AS. Percutaneous transluminal coronary angioplasty as a first revascularization procedure in single-, double- and triple-vessel coronary artery disease. *J Am Coll Cardiol* 1995;**26**:142–51.

289. Wilson WS, Stone GW. Late results of percutaneous transluminal coronary angioplasty of two or more major native coronary arteries. *Am J Cardiol* 1994; **73**:1041–6.

290. Yamaguchi T. Early and long-term results of percutaneous transluminal coronary angioplasty for angina pectoris. *Jpn J Med* 1990;**29**:669–71.

291. Tebbe U, Schatz RA, Scholz KH, Brune S, Sauer G, Kreuzer H. Koronare Angioplastie mit Implantation einer Gefassstutze (Stent). [Coronary angioplasty with the implantation of a vascular support (stent)]. *Dtsch Med Wochenschr* 1990;**115**:1054–8.

292. Cribier A, Korsatz L, Koning R, Rath P, Gamra H, Stix G, *et al*. Improved myocardial ischemic response and enhanced collateral circulation with long repetitive coronary occlusion during angioplasty: a prospective study. *J Am Coll Cardiol* 1992;**20**:578–86.

293. Thomas ES, Williams DO. The long-term followup results of coronary artery angioplasty. *Cardiovasc Clin* 1988;**19**:253–63.

294. Dario A, Fortini G, Bertollo D, Bacuzzi A, Grizzetti C, Cuffari S. Treatment of failed back surgery syndrome. *Neuromodulation* 2001;**4**:105–10.

295. North RB, Kidd DH, Lee MS. Spinal cord stimulation versus operation for failed back surgery syndrome: a prospective, randomised controlled trial. *Acta Neurochir Suppl* 1995;**64**:106–8.

296. Barolat G. A prospective multicentre study to assess the efficacy of spinal cord stimulation utilizing a multi-channel radio-frequency system for the treatment of intractable low and low extremity back pain. Initial considerations and methodology. *Neuromodulation* 1999;**2**:179–83.

297. Barolat G, Oakley JC, Law JD, North RB, Ketcik B, Sharan A. Epidural spinal cord stimulation with a multiple electrode paddle lead is effective in treating intractable low back pain. *Neuromodulation* 2001;**4**:59–66.

298. Batier C, Frerebeau P, Kong A, Siou D, Privat JM, Roquefeuil B, *et al*. Neurostimulation cordonale postérieure dans les lombo-radiculalgies. *Agressologie*1989;**30**:137–8.

299. Bel S, Bauer BL. Dorsal column stimulation (DCS): cost to benefit analysis. *Acta Neurochir Suppl* 1991;**52**:121–3.

300. Blond S, Armignies P, Parker F, Dupard T, Guieu JD, Duquesnoy B, *et al*. Sciatalgies chroniques par desafferentation sensitive après chirurgie de la hernie discale lombaire: aspects cliniques et therapeutiques. *Neurochirurgie* 1991;**37**:86–95.

301. Blond S, Armignies P, Veys B, Duneton O, Buisset N, Touzet G. Postoperative neuropathic pain. *Douleur Analgesie* 1998;**11**:120–30.

302. Blume H, Richardson R, Rojas C. Epidural nerve stimulation of the lower spinal cord and cauda equina for the relief of intractable pain in failed low back surgery. *Neurosurgery* 1992;**45**:456–60.

303. Burchiel KJ, Brown FD, Erickson RK, *et al*. Correlates of pain relief in spinal cord stimulator recipients. In Gebhart GF, Hammond DL, Jensen TS, editors. *Proceedings of the 7th World Congress on Pain*, August 1993. Seattle, WA: IASP Press. Abstract 335.

304. Burchiel, KJ, Anderson, VC, Wilson BJ, Denison DB, Olson KA, Shatin D. Prognostic factors of spinal cord stimulation for chronic back and leg pain. *Neurosurgery* 1995;**36**:1101–10.

305. Clark K. Electrical stimulation of the nervous system for control of pain. University of Texas Southwestern Medical School experience. *Surg Neurol* 1975;**4**:164–6.

306. Dam Hieu HP, Person H, Hoidi K, Rodriguez V, Vallee B, Besson G. Treatment of chronic low back and radicular pain by spinal cord stimulation. Long-term results. *Revue du Rhumatisme (English Edition)* 1994;**61**:255–60.

307. De La Porte C, Siegfried J. Lumbosacral spinal fibrosis (spinal arachnoiditis). Its diagnosis and treatment by spinal cord stimulation. *Spine* 1983; **8**:593–603.

308. De La Porte C, Van De Kelft E. Spinal cord stimulation in failed back surgery syndrome. *Pain* 1993;**52**:55–61.

309. Demirel T, Braun W, Reimers CD. Results of spinal cord stimulation in patients suffering from chronic pain after two year observation period. *Neurochirurgie* 1984;**27**:47–50.

310. Devulder J, De Colvenaer L, Rolly G, Caemaert J, Calliauw L, Martens F. Spinal cord stimulation in chronic pain therapy. *Clin J Pain* 1990;**6**:51–6.

311. Devulder J, Vermeulen H, De Colvenaer L, Rolly G, Calliauw L, Caemaert J. Spinal cord stimulation in chronic pain: evaluation of results, complications, and technical considerations in sixty-nine patients. *Clin J Pain* 1991;**7**:21–8.

312. Devulder J, De Laat M, Van Bastelaere M, Rolly G. Spinal cord stimulation: a valuable treatment for chronic failed back surgery patients. *J Pain Symptom Manage* 1997;**13**:296–301.

313. Fassio B, Sportes J, Romain M, Vallespir R. La neuro-stimulation medullaire chronique dans les arachnoidites rachidiennes lombo-sacrées. *Rev Chir Orthop* 1988;**74**:473–9.

314. Gonzalez-Darder JM, Canela P, Gonzales-Matinez V, Ortiz E, Gomez R, Ceijas F. Estimulacion electrica medular. Indicationes acurales y resultados an una seie de 46 pacientes. *Rev Esp Anesthiol Reanim* 1992;**39**:86–90.

315. Hassenbusch SJ, Stanton-Hicks M, Covington EC. Spinal cord stimulation versus spinal infusion for low back and leg pain. *Acta Neurochir* 1995;**64**(Suppl):109–11.

316. Heidecke V, Rainov NG, Burkert W. Hardware failures in spinal cord stimulation for failed back surgery syndrome. *Neuromodulation* 2000; **3**:27–30.

317. Hoppenstein R. Electrical stimulation of the ventral and dorsal columns of the spinal cord for relief of chronic intractable pain: preliminary report. *Surg Neurol* 1975;**4**:187–94.

318. Hunt WE, Goodman JH, Bingham WG. Stimulation of the dorsal spinal cord for treatment of intractable pain: a preliminary report. *Surg Neurol* 1975;**4**:153–6.

319. Kalin M-T, Winkelmuller W. Chronic pain after multiple lumbar discectomies – significance of intermittent spinal cord stimulation. *Pain* 1994; Suppl 2:S241.

320. Kavar B, Rosenfeld JV, Hutchinson A. The efficacy of spinal cord stimulation for chronic pain. *J Clin Neurosci* 2000;**7**:409–13.

321. Kay AD, McIntyre MD, Macrae WA, Varma TRK. Spinal cord stimulation – a long-term evaluation in patients with chronic pain. *Br J Neurosurg* 2001;**15**:335–41.

322. Kim JA, Keravel Y, Ioku M. The effect of epidural spinal cord stimulation for the 'failed back surgery syndrome' and on the prognostic factors: an analysis of fifty-eight cases. *Jpn J Neurosurg* 1994; **3**:294–300.

323. Kim SH, Tasker RR, Oh MY, Burchiel KJ, Hsu F, Kanpolat Y, *et al*. Spinal cord stimulation for non-specific limb pain versus neuropathic pain and spontaneous versus evoked pain. *Neurosurgery* 2001;**48**:1056–65.

324. Kumar K, Wyant GM, Ekong CEU. Epidural spinal cord stimulation for relief of chronic pain. *Pain Clin* 1986;**1**:91–9.

325. Kumar K, Nath R, Wyant GM. Treatment of chronic pain by epidural spinal cord stimulation: a 10-year experience. *J Neurosurg* 1991;**75**:402–7.

326. Kumar K, Toth C, Nath RK. Spinal cord stimulation for chronic pain in peripheral neuropathy. *Surg Neurol* 1996;**46**:363–9

327. Kumar K, Toth C. The role of spinal cord stimulation in the treatment of chronic pain postlaminectomy. *Curr Rev Pain* 1998;**2**:85–92.

328. Kumpulainen T, Ronty H, Koivukangas J. Management of patients with pain. *Ann Clin Res* 1986;**18** (Suppl 47):97–101.

329. Law JD. Percutaneous spinal cord stimulation for the failed back surgery syndrome. *Pain Manage Update* 1. 1993;**1**:1–2.

330. Law JD. Clinical and technical results from spinal stimulation for chronic pain of diverse pathophysiologies. *Stereot Functi Neurosurg* 1992;**59**:21–4.

331. Lazorthes Y, Siegfield J, Verdie JC, Casaux J. La stimulation medullaire chronique dans le traitment des douleurs neurogenes. *Neurochirugie* 1995;**41**:73–88

332. Leclercq TA, Russo E. La stimulation épidurale dans le traitement des douleurs chroniques. *Neurochirurgie* 1981;**27**:125–8.

333. Leclercq TA. Epidural stimulation for pain control in the 'failed disc' syndrome. *R I Med J* 1982; **65**:25–8.

334. LeDoux MS, Langford KH. Spinal cord stimulation for the failed back syndrome. *Spine* 1993;**18**:191–4.

335. Leibrock LG, Meilman P, Cuka D, Green C. Spinal cord stimulation in the treatment of chronic back and lower extremity pain syndromes. *Nebr Med J* 1984;**69**:180–3.

336. LeRoy PL. Stimulation of spinal neuraxis by biocompatible electrical current in the human. *Appl Neurophysiol* 1981;**44**:187–93.

337. Leveque J-C, Villavicencio AT, Bulsara KR, Rubin L, Gorecki JP. Spinal cord stimulation for failed back surgery syndrome. *Neuromodulation* 2000;**4**:1–9.

338. Long DM, Erickson DE. Stimulation of the posterior columns of the spinal cord for relief of intractable pain. *Surg Neurol* 1975;**4**:134–41.

339. Long, DM. Patient selection and the results of spinal cord stimulation for chronic pain. In Hosobuchi Y, Corbin T, editors. *Indications for spinal cord stimulation*. Amsterdam: Excerpta Medica; 1981. pp. 1–16.

340. Meglio M, Cioni B, Visocchi M. Spinal cord stimulation in low back pain and leg pain. *Stereotact Funct Neurosurg* 1994;**62**:263–4.

341. Meilman PW, Leibrock LG, Leong FTL. Outcome of implanted spinal cord stimulation in the treatment of chronic pain: arachnoiditis versus single nerve root injury and mononeuropathy. *Clin J Pain* 1989;**5**:189–93.

342. Mittal B, Thomas D, Walton P, Calder I. Dorsal column stimulation (DCS) in chronic pain: report of 31 cases. *Ann R Coll Surg Engl* 1987;**69**:104–9.

343. Mundinger F, Neumuller H. Programmed stimulation for control of chronic pain and motor diseases. *Appl Neurophysiol* 1982;**45**:102–11.

344. Nielson KD, Adams JE, Hosobuchi Y. Experience with dorsal column stimulation for relief of chronic intractable pain: 1968–1973. *Surg Neurol* 1975;**4**:148–52

345. North RB, Fischell TA, Long DM. Chronic dorsal column stimulation via percutaneously inserted epidural electrodes. *Appl Neurophysiol* 1977; **40**:184–191.

346. North RB, Ewend MG, Lawton MT, Piantadosi S. Spinal cord stimulation for chronic, intractable pain: superiority of 'multi-channel' devices. *Pain* 1991;**44**:119–30.

347. North RB, Ewend MG, Lawton MT, Kidd DH, Piantadosi S. Failed back surgery syndrome: 5-year follow-up after spinal cord stimulator implantation. *Neurosurgery* 1991;**28**:692–9.

348. North RB, Long DM. Epidural 'dorsal column' stimulation for chronic intractable pain: eight-year follow-up. *Pain* 1984;S79.

349. Ohnmeiss DD, Rashbaum RF, Bogdanffy GM, Sonntag VKH, Hurlbert RJ. Prospective outcome evaluation of spinal cord stimulation in patients with intractable leg pain. *Spine*1996;**21**:1344–51.

350. Ohnmeiss DD, Rashbaum RF. Patient satisfaction with spinal cord stimulation for predominant complaints of chronic, intractable low back pain. *Spine J* 2001;358–363.

351. Pineda A. Dorsal column stimulation and its prospects. *Surg Neurol* 1975;**4**:157–63.

352. Probst C. Spinal cord stimulation in 112 patients with epi-/intradural fibrosis following operation for lumbar disc herniation. *Acta Neurochir (Wien)* 1990; **107**:147–51.

353. Rainov NG, Heidecke V, Burkert W. Short test-period spinal cord stimulation for failed back surgery syndrome. *Minim Invasive Neurosurg* 1996;**39**:41–4.

354. Ray CD. Control of pain by electrical stimulation. A clinical follow-up review. *Adv Neurosurg* 1975;**3**:216–24.

355. Ray DC, Burton CV, Lifson A. Neurostimulation as used in a large clinical practice. *Appl Neurophysiol* 1982;**45**:160–6.

356. Richardson RR, Siqueira EB, Cerullo LJ. Spinal epidural neurostimulation for treatment of acute and chronic intractable pain: initial and long term results. *Neurosurgery* 1979;**5**:344–8.

357. Robb MP. Practical considerations in spinal cord stimulation. *Pain* 1990;Suppl 5:S234.

358. Seijo F, Moreno JM, Jarodat M, Lopez JM. Spinal cord stimulation in post-surgical low back pain. *Pain Clin* 1993;**6**:241–7.

359. Shatin D, Mullett K, Hults G. Totally implantable spinal cord stimulation for chronic pain: design and efficacy. *Pace* 1986;**9**:577–83.

360. Shatin D. The efficacy of spinal cord stimulation for the control of chronic intractable pain. *Pain* 1990;Suppl 5:S115.

361. Sheldon, Paul F, Jacques DB, Pudenz RH. Electrical stimulation of the nervous system. *Surg Neurol* 1975:**4**:127–32.

362. Siegfried J, Lazorthes Y. Long-term follow-up of dorsal cord stimulation for chronic pain syndrome after multiple lumbar operations. *Appl Neurophysiol*. 1982;**45**:201–4.

363. Simpson BA. Spinal cord stimulation in 60 cases of intractable pain. *J Neurol Neurosurg Psychiatry* 1991; **54**:196–9.

364. Spiegelmann R, Friedman WA. Spinal cord stimulation: a contemporary series. *Neurosurgery* 1991;**28**:65–71.

365. Van Buyten J-P, Van Zundert J. Milbouw G. Treatment of failed back surgery syndrome patients with low back and leg pain: a pilot study of a new dual lead spinal cord stimulation system. *Neuromodulation* 1999;**2**:258–65.

366. Van de Kelft E, De La Porte C. Long-term pain relief during spinal cord stimulation. The effect of patients selection. *Qual Life Res* 1994;**3**: 21–7.

367. Vogel HP, Heppner B, Humbs N, Schramm J, Wagner C. Long-term effects of spinal cord stimulation in chronic pain syndromes. *J Neurol* 1986;**233**:16–18.

368. Waisbrod H, Gerbershagen HU. Spinal cord stimulation in patients with a battered root syndrome. *Arch Orthop Trauma Surg* 1985; **104**:62–4.

369. Wester K. Dorsal column stimulation in pain treatment. *Acta Neurol Scand* 1987;**75**: 151–5.

370. Winkelmuller, W. Experience with the control of low back pain by the dorsal column stimulation (DCS) system and by the peridural electrode system (PISCES). In Hosobuchi Y, Corbin T, editors. *Indications for spinal cord stimulation*. Amsterdam: Excerpta Medica; 1981. pp. 34–41.

Appendix 1

Data extraction forms for audit of NICE HTAs

Author, completion date	Hyde, Sept. 2001[40]
Title of review	Fludarabine as second line therapy for B-cell chronic lymphocytic leukaemia
Reference	*Health Technol Assess* 2002;**6**(2)
Review group	University of Birmingham
Type of review	Therapy
Number of included RCTs	2
Number and type of other included evidence	0
Number of included case series	7
Reason for including case series	Absence of RCTs
Criteria for case series inclusion/exclusions	Included studies with >50 patients
Reason for inclusion criteria	To limit work required
Methods used to assess case series quality	Cochrane framework • conducted prospectively • consecutive series • describe patient characteristics • loss to follow-up <10% • adequate follow-up • analysis of prognostic factors • relevance
Methods used to synthesise case series results	Descriptive/tabulation
Conclusions drawn from case series evidence	Confirm that cautious interpretation of RCT results is appropriate. No conclusions drawn from case series

Author, completion date	Forbes, July 2001[41]
Title of review	A rapid and systematic review of the clinical effectiveness and cost effectiveness of pegylated liposomal doxurubicin hydrochloride for ovarian cancer
Reference	*Health Technol Assess* 2002;**6**(23)
Review group	University of York
Type of review	Therapy
Number of included RCTs	2
Number and type of other included evidence	0
Number of included case series	6
Reason for including case series	Requested by NICE
Criteria for case series inclusion/exclusions	None stated
Reason for inclusion criteria	Not applicable

continued

Methods used to assess case series quality	Crombie, 1996,[54] Pocket guide to appraisal • participants described • clear aims • control group used • should there have been a control group? • was the best study design used? • sufficient follow-up • adequate sample size • valid outcome measures • compliance with treatment • relevant outcomes missed • adequate description of statistical methods • untoward events that may affect findings • appropriate use of survival analysis • all patients accounted for • basic data adequately described • statistical significance reported • confounders assessed • null findings appropriately interpreted • important effects overlooked
Methods used to synthesise case series results	Descriptive/tabulation
Conclusions drawn from case series evidence	No conclusions drawn from case series

Author, completion date	Vale, June 2002[42]
Title of review	Systematic review of the effectiveness and cost-effectiveness of metal-on-metal hip resurfacing arthroplasty for treatment of hip disease
Reference	*Health Technol Assess* 2002;**6**(15)
Review group	University of Aberdeen
Type of review	Therapy
Number of included RCTs	1 (comparator)
Number and type of other included evidence	Systematic review = 3 (comparator)
Number of included case series	20 (5 intervention, 15 comparator)
Reason for including case series	Lack of data from RCTs
Criteria for case series inclusion/exclusions	Studies included if they had a minimum follow-up of 2 years
Reason for inclusion criteria	Not stated
Methods used to assess case series quality	Morris and colleagues checklist (orthopaedic)[55] • clarity of question/definition of outcome • description of prosthesis and fixation • description of study sample • control of bias in study design • duration of follow-up • statistical and analytical (score out of 6)
Methods used to synthesise case series results	Descriptive. Pooled for use in economic model sensitivity analysis
Conclusions drawn from case series evidence	No conclusions drawn from case series. Pooled estimates used in economic model sensitivity analysis

Author, completion date	Bryant, Oct. 2001[53]
Title of review	Clinical effectiveness and cost-effectiveness of growth hormone in children
Reference	*Health Technol Assess* 2002;**6**(18)
Review group	University of Southampton
Type of review	Therapy
Number of included RCTs	21
Number and type of other included evidence	Undetermined
Number of included case series	11
Reason for including case series	Lack of data from RCTs on main outcome measure 'height'
Criteria for case series inclusion/exclusions	Studies included if they had >300 patients
Reason for inclusion criteria	In order to maximise generalisability
Methods used to assess case series quality	Modified Spitzer criteria • random • proper sampling • adequate sample size • objective outcome measurement • blind outcome measurement • eligibility criteria • attrition rates • comparable groups • generalisability
Methods used to synthesise case series results	Narrative review and tabulation
Conclusions drawn from case series evidence	Undetermined

Author, completion date	Peters, April 2002[43]
Title of review	The clinical effectiveness and cost-effectiveness of inhaler devices used in routine management of chronic asthma in older children
Reference	*Health Technol Assess* 2002;**6**(5)
Review group	University of Sheffield
Type of review	Therapy
Number of included RCTs	10
Number and type of other included evidence	4
Number of included case series	16
Reason for including case series	Patient preferences data
Criteria for case series inclusion/exclusions	None stated
Reason for inclusion criteria	Not applicable
Methods used to assess case series quality	User's Guide to EBM: how to use an article about harm • clearly identified and comparable groups • outcomes measured in same way for groups • sufficient follow-up • temporal relationship/dose – response gradient • strength of association and precision • generalisability
Methods used to synthesise case series results	Descriptive
Conclusions drawn from case series evidence	Provided information on instructions given, ease of use, patient compliance, adherence and preference. Recommended qualitative research to confirm

Author, completion date	Vardulaki, Dec. 2000[44]
Title of review	A systematic review of the effectiveness and cost-effectiveness of laparoscopic surgery for colorectal cancer
Reference	Report to NICE. URL: http://www,nice.org.uk/pdf/htareportonlapsurgcoloreccanc.pdf
Review group	External (Royal College of Surgeons)
Type of review	Therapy
Number of included RCTs	2
Number and type of other included evidence	Prospective cohort = 7 Retrospective cohort = 5 Historically controlled cohort = 2
Number of included case series	37
Reason for including case series	Long-term follow-up
Criteria for case series inclusion/exclusions	Studies with >10 patients were included
Reason for inclusion criteria	Studies with <10 patients likely to contain selected and hence unrepresentative patients
Methods used to assess case series quality	Vardulaki instrument, 2000 • aims of study clear • is case definition clear • data collected prospectively • patients consecutive • use of CIs or SE • outcomes stratified by disease stage • clear definition of outcomes (score out of 7)
Methods used to synthesise case series results	Medians, descriptive, pooled estimates and weighted regression to explore variation
Conclusions drawn from case series evidence	Make broad suggestions, but do not draw conclusions from case series

Author, completion date	Wake, March 2002[45]
Title of review	Rituximab as third-line treatment for refractory or recurrent stage III or IV follicular non-Hodgkin's lymphoma
Reference	*Health Technol Assess* 2002;**6**(3)
Review group	University of Birmingham
Type of review	Therapy
Number of included RCTs	0
Number and type of other included evidence	0
Number of included case series	4
Reason for including case series	Lack of RCTs or comparative evidence
Criteria for case series inclusion/exclusions	Included studies with >10 patients
Reason for inclusion criteria	Not stated
Methods used to assess case series quality	Young *et al.*[56] • conducted prospectively • consecutive patients • clear patient characteristics • loss to follow up <10%
Methods used to synthesise case series results	Qualitative
Conclusions drawn from case series evidence	No conclusions drawn from case series

Author, completion date	Ward, May 2001[46]
Title of review	A rapid and systematic review of the clinical effectiveness and cost-effectiveness of gemcitabine for the treatment of pancreatic cancer
Reference	*Health Technol Assess* 2001;**5**(24)
Review group	University of Sheffield
Type of review	Therapy
Number of included RCTs	7
Number and type of other included evidence	Other design? = 7
Number of included case series	57 (phase II studies)
Reason for including case series	Not stated
Criteria for case series inclusion/exclusions	None
Reason for inclusion criteria	Not applicable
Methods used to assess case series quality	Applied hierarchy of evidence
Methods used to synthesis case series results	Results tabulated
Conclusions drawn from case series evidence	No conclusions drawn from case series

Author, completion date	Dinnes, April 2001[47]
Title of review	The effectiveness and cost-effectiveness of temozolomide for the treatment of recurrent malignant glioma
Reference	*Health Technol Assess* 2001;**5**(13)
Review group	University of Southampton
Type of review	Therapy
Number of included RCTs	1
Number and type of other included evidence	0
Number of included case series	6
Reason for including case series	Anticipated lack of data
Criteria for case series inclusion/exclusions	Included studies with a minimum of 45 patients
Reason for inclusion criteria	Enabled the pool of larger studies to be included. Other studies had considerably less patients
Methods used to assess case series quality	Modified Spitzer checklist • proper random assignment • proper sampling • adequate sample size • objective outcomes • blind assessment • objective eligibility criteria • reported attrition • comparability of groups • generalisability
Methods used to synthesise case series results	Narrative synthesis/description
Conclusions drawn from case series evidence	Present speculation but do not draw conclusions based on case series

Author, completion date	Lewis, March 2002[48]
Title of review	The clinical effectiveness and cost-effectiveness of trastuzumab for breast cancer
Reference	*Health Technol Assess* 2002;**6**(3)
Review group	University of York
Type of review	Therapy
Number of included RCTs	2
Number and type of other included evidence	0
Number of included case series	2
Reason for including case series	Requested by NICE
Criteria for case series inclusion/exclusions	None stated
Reason for inclusion criteria	Not applicable
Methods used to assess case series quality	CRD Report No. 4 • representative sample • explicit inclusion criteria • individuals entered survey at similar time point • long enough follow-up • use of objective criteria and blinding to assess outcomes • description of subseries and distribution of prognostic factors
Methods used to synthesise case series results	Structured tables and narrative summary
Conclusions drawn from case series evidence	Conclusion made while highlighting limitations of case series evidence

Author, completion date	Fitzpatrick, June 1997[49]
Title of review	Primary total hip replacement surgery: a systematic review of outcomes and modelling of cost-effectiveness associated with different prostheses
Reference	*Health Technol Assess* 1998;**2**:(20)
Review group	Universities of Oxford and York
Type of review	Therapy
Number of included RCTs	11
Number and type of other included evidence	Non-randomised comparative studies = 18
Number of included case series	159
Reason for including case series	Longer follow-up data
Criteria for case series inclusion/exclusions	Studies included with >5 years follow-up
Reason for inclusion criteria	Review of survival analyses confirmed that cut-off was conservative and would omit only small numbers of adverse outcomes
Methods used to assess case series quality	Quality assessed for 15 studies only • clarity of study question and outcomes • description of prosthesis and method of fixation • description of study sample • control bias in study design • duration and completeness of follow-up • statistical and analytical considerations (added to give an overall score)
Methods used to synthesise case series results	Included in a meta-analysis
Conclusions drawn from case series evidence	Conclusions drawn with caution

Author, completion date	Jobanputra, Dec. 2000[50]
Title of review	Effectiveness of autologous chondrocyte transplantation for hyaline cartilage defects in knees
Reference	*Health Technol Assess* 2001;**5**:(11)
Review group	University of Birmingham
Type of review	Therapy
Number of included RCTs	0
Number and type of other included evidence	0
Number of included case series	20
Reason for including case series	Only available evidence
Criteria for case series inclusion/exclusions	None
Reason for inclusion criteria	Not applicable
Methods used to assess case series quality	Categorisation as follows: A. patient input to outcomes before and after surgery, adverse effects B. patient input into outcomes before and after surgery C. patient input to outcomes after surgery D. input from clinician or radiographic evaluation only
Methods used to synthesise case series results	Narrative summary
Conclusions drawn from case series evidence	No conclusions drawn

Author, completion date	Bagnall, March 2002[51]
Title of review	A rapid and systematic review of atypical antipsychotics in schizophrenia
Reference	*Health Technol Assess* 2003;**7**(13)
Review group	University of York
Type of review	Therapy
Number of included RCTs	70 (new RCTs, not including previous review that this updates)
Number and type of other included evidence	Cohort = 13 Case–control = 1
Number of included case series	27
Reason for including case series	Safety
Criteria for case series inclusion/exclusions	Included studies with at least 2 years follow-up or >2000 patients
Reason for inclusion criteria	Not stated
Methods used to assess case series quality	CRD Report No. 4 checklist
Methods used to synthesise case series results	Narrative
Conclusions drawn from case series evidence	Conclusions not based on case series

Author, completion date	Woolacott, March 2002[52]
Title of review	The clinical effectiveness and cost-effectiveness of bupropion SR and nicotine replacement therapy (NRT) for smoking cessation
Reference	*Health Technol Assess* 2002;**6**(16)
Review group	Universities of York and Birmingham
Type of review	Therapy
Number of included RCTs	18
Number and type of other included evidence	Systematic review = 2 Non-RCTs = 3 Uncontrolled = 19 Case–control = 1 Surveillance studies = 5
Number of included case series	17 (case series or case reports)
Reason for including case series	Safety
Criteria for case series inclusion/exclusions	None
Reason for inclusion criteria	Not applicable
Methods used to assess case series quality	CRD checklist for cohort studies
Methods used to synthesise case series results	Narrative
Conclusions drawn from case series evidence	Unable to determine

Appendix 2
Search strategies

Case series search strategies

Database and years searched	Search strategy	No. retrieved
Cochrane Methodology Register 2003, Issue 1	Saved strategy: 1. Case series – studies #1 ((single next arm) or (single next group) or non-comparative) #2 (phase next ii) #3 (observational or retrospective or (case next series)) #4 (#1 or #2 or #3) #5 ((evidence next base) or (intermethod next comparison) or (outcomes next research) or (research next design) or methodolog* or (critical next appraisal) or (epidemiologic next research next design)) #6 ((sample next size) or (effect next size)) #7 BIAS (EPIDEMIOLOGY) explode all trees (MeSH) #8 (research next design) #9 RESEARCH DESIGN explode all trees (MeSH) #10 (research:ti next design:ti) #11 #5 or #6 or #7 or #9 or #10 #12 #4 and #11 #13 #5 or #6 or #7 or #8 or #9 #14 #4 and #13	83 112
MEDLINE 1966–2003/Feb. week 4 (4 March 2003)	((single arm) or (single group) or (non-comparative) or (phase ii) or (phase 2) or (observational) or (case series)) and ((systematic* near (review* or overview*)) or (meta-analy* or metaanaly*) or (meta-analysis in pt)) not ((comment in pt) or (letter in pt) or (editorial in pt))	768
HTA database (12 February 2003)	single arm or single group or non-comparative or phase ii or phase 2 or observational or retrospective or case series/All fields AND evidence base or intermethod comparison or outcomes research or research design or methodolog* or critical appraisal or sample size or effect size or bias/All fields	19
Regard ESRC research register (13 February 2003)	single arm or single group or non-comparative or phase ii or phase 2 or observational or retrospective or case series AND evidence base or intermethod comparison or outcomes research or research design or methodolog* or critical appraisal or sample size or effect size or bias	2
	single arm or single group or non-comparative or phase ii or phase 2 or observational or retrospective or case series or evidence base or intermethod comparison or outcomes research or research design or methodolog* or critical appraisal or sample size or effect size or bias	207
NRR (National Research Register) 2002, Issue 4 13 February 2003 – checked 2003, Issue 1 on Internet – no extra references	(single arm or single group or non-comparative or observational or retrospective or case series or evidence base or intermethod comparison or outcomes research or research design or methodolog* or critical appraisal or sample size or effect size or bias) in pk	46
24 February 2003	("single arm" or "single group" or "non comparative" or observational or retrospective or phase) and ("evidence base" or "intermethod comparison" or "outcomes research" or "research design" or methodolog* or "critical appraisal" or sample or effect or bias)	9019

continued

Database and years searched	Search strategy	No. retrieved
Handsearching – *Int J Technol Assess Health Care*	Issues checked: 18(1) 17(1)ol 16(1)ol 15(1) 14(1) 13(1) 12(1) 10(1) 18(2) 17(2) 16(2) 15(2) 14(2) 13(2) 12(2) 10(2) 18(3) 17(3) 16(3) 15(3) 14(3) 13(3) 12(3) 10(3) 18(4)ol 17(4) 16(4) 15(4) 14(4) 13(4) 12(4) 10(4) 9(2) 19(1)	3

ol, on-line.

Search strategy – for case series – angina repeat search

Database and years searched	Search strategy	No. retrieved
MEDLINE (Webspirs) 1980–2003 February, week 1	#1 'Cost-Benefit-Analysis' / all subheadings in MIME,MJME (3443 records) #2 'Cost-Savings' / all subheadings in MIME,MJME (517 records) #3 'Cost-of-Illness' / all subheadings in MIME,MJME (1202 records) #4 'Economics-' / all subheadings in MIME,MJME (223 records) #5 cost benefit in ti,ab (485 records) #6 (cost effective* or cost utility) in ti,ab (4757 records) #7 cost in ti,ab (13,739 records) #8 (cost saving or cost minimization or cost minimisation) in ti,ab (327 records) #9 'Economics-Pharmaceutical' / all subheadings in MIME,MJME (211 records) #10 qaly in ti,ab (162 records) #11 quality adjusted life year* in ti,ab (284 records) #12 economic in ti,ab (6087 records) #13 (analysis or evaluation) in ti,ab (175,200 records) #14 #12 and #13 (1936 records) #15 benefit or (effective* in ti,ab) (80,812 records) #16 #7 and #15 (7199 records) #17 #4 and #16 (4 records) #18 (efficacy or response or sensitivity) in ti,ab (131,884 records) #19 #7 and #18 (2680 records) #20 (specificity or outcome) in ti,ab (57,075 records) #21 #7 and #20 (1746 records) #22 angina in ti,ab (2406 records) #23 'Angina-Pectoris' / all subheadings in MIME,MJME (822 records) #24 (angina in ti,ab) or ('Angina-Pectoris' / all subheadings in MIME,MJME) (2684 records) #25 'Nitrates-' / all subheadings in MIME,MJME (1330 records) #26 nitrate* in ti,ab (3001 records) #27 beta with blocker* (2033 records) #28 'Adrenergic-beta-Agonists' / all subheadings in MIME,MJME (1113 records) #29 calcium channel blocker* in ti,ab (788 records) #30 'Calcium-Channel-Blockers' / all subheadings in MIME,MJME (2143 records) #31 coronary artery bypass in ti,ab (2340 records) #32 cabg in ti,ab (1037 records) #33 percutaneous transluminal coronary angioplasty in ti,ab (554 records) #34 ptca in ti,ab (622 records) #35 angioplasty in ti,ab (2826 records) #36 atherectomy in ti,ab (161 records) #37 stents in ti,ab (1463 records) #38 'Myocardial-Revascularization' / all subheadings in MIME,MJME (606 records) #39 'Angioplasty-' / all subheadings in MIME,MJME (336 records) #40 'Balloon-Dilatation' / all subheadings in MIME,MJME (753 records) #41 'Angioplasty-Balloon' / all subheadings in MIME,MJME (700 records) #42 'Angioplasty-Laser' / all subheadings in MIME,MJME (29 records)	5339

continued

Database and years searched	Search strategy	No. retrieved
	#43 'Stents-' / all subheadings in MIME,MJME (3337 records) #44 ('Nitrates-' / all subheadings in MIME,MJME) or (nitrate* in ti,ab) or (beta with blocker*) or ('Adrenergic-beta-Agonists' / all subheadings in MIME,MJME) or (calcium channel blocker* in ti,ab) or ('Calcium-Channel-Blockers' / all subheadings in MIME,MJME) or (coronary artery bypass in ti,ab) or (cabg in ti,ab) or (percutaneous transluminal coronary angioplasty in ti,ab) or (ptca in ti,ab) or (angioplasty in ti,ab) or (atherectomy in ti,ab) or (stents in ti,ab) or ('Myocardial-Revascularization' / all subheadings in MIME,MJME) or ('Angioplasty-' / all subheadings in MIME,MJME) or ('Balloon-Dilatation' / all subheadings in MIME,MJME) or ('Angioplasty-Balloon' / all subheadings in MIME,MJME) or ('Angioplasty-Laser' / all subheadings in MIME,MJME) or ('Stents-' / all subheadings in MIME,MJME) (17,822 records) #45 'Evaluation-Studies' / all subheadings in MIME,MJME (1208 records) #46 randomized controlled trial in pt (0 records) #47 'Randomized-Controlled-Trials' / all subheadings in MIME,MJME (6466 records) #48 'Random-Allocation' / all subheadings in MIME,MJME (3188 records) #49 'Double-Blind-Method' / all subheadings in MIME,MJME (6785 records) #50 'Single-Blind-Method' / all subheadings in MIME,MJME (1090 records) #51 clinical trial in pt (0 records) #52 'Clinical-Trials' / all subheadings in MIME,MJME (7146 records) #53 clinical near5 trial (4557 records) #54 (singl* or doubl* or trebl* or tripl*) near5 (blind* or mask*) (10,589 records) #55 'Placebos-' / all subheadings in MIME,MJME (1138 records) #56 placebo* or (random* in ti,ab) (48,147 records) #57 'Research-Design' / all subheadings in MIME,MJME (3591 records) #58 'Follow-Up-Studies' / all subheadings in MIME,MJME (22,895 records) #59 'Prospective-Studies' / all subheadings in MIME,MJME (20,745 records) #60 control* or prospectiv* or (volunteer* in ti,ab) (229,710 records) #61 regist* in ti,ab (10,083 records) #62 'Registries-' / all subheadings in MIME,MJME (2505 records) #63 ('Evaluation-Studies' / all subheadings in MIME,MJME) or (randomized controlled trial in pt) or ('Randomized-Controlled-Trials' / all subheadings in MIME,MJME) or ('Random-Allocation' / all subheadings in MIME,MJME) or ('Double-Blind-Method' / all subheadings in MIME,MJME) or ('Single-Blind-Method' / all subheadings in MIME,MJME) or (clinical trial in pt) or ('Clinical-Trials' / all subheadings in MIME,MJME) or (clinical near5 trial) or ((singl* or doubl* or trebl* or tripl*) near5 (blind* or mask*)) or ('Placebos-' / all subheadings in MIME,MJME) or (placebo* or (random* in ti,ab)) or ('Research-Design' / all subheadings in MIME,MJME) or ('Follow-Up-Studies' / all subheadings in MIME,MJME) or ('Prospective-Studies' / all subheadings in MIME,MJME) or (control* or prospectiv* or (volunteer* in ti,ab)) or (regist* in ti,ab) or ('Registries-' / all subheadings in MIME,MJME) (279,738 records) #64 #19 or #21 (3777 records) #65 #45 and #64 (7 records) #66 #4 and #64 (3 records) #67 #7 and #24 (84 records) #68 #1 or #2 or #3 or #4 or #5 or #6 or #67 (8497 records) #69 #8 or #9 or #10 or #11 or #14 (2556 records) #70 #16 and #45 (26 records) #71 #17 or #70 (30 records) #72 #68 or #69 or #71 (10,038 records) #73 #65 or #66 or #72 (10,040 records) #74 #63 or #73 (285,329 records) #75 #24 and #74 (1416 records) #76 (benefit or effective*) in ti,ab (78,945 records) #77 #7 and #76 (6838 records) #78 #4 and #77 (4 records)	

continued

Database and years searched	Search strategy	No. retrieved
	#79 #45 and #77 (23 records) #80 #19 or #21 (3777 records) #81 #45 and #80 (7 records) #82 #4 and #80 (3 records) #83 #68 or #69 or #71 or #81 or #82 (10,040 records) #84 #63 or #83 (285,329 records) #85 #24 and #44 (995 records) #86 #84 and #85 (571 records) #87 (animal in tg) not ((animal in tg) and (human in tg)) (141,734 records) #88 #86 not #87 (565 records)	

Appendix 3

Papers identified at abstract stage

Reference	Included or excluded?
Cochrane Methodology Database – 112 hits, 11 papers requested	
Albert JM, Yun H. Statistical advances in AIDS therapy trials. *Stat Methods Med Res* 2001;**10**:85–100	Excluded, no information about quality
Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med* 2000;**342**:1878–86	Obtained – excluded at full text stage – observational studies have control groups
Castillo S, Marson A, Chadwick D, Hutton J. Systematic reviews of randomized controlled trials and observational studies of antiepileptic drugs: is there a systematic bias? *The Cochrane Methodology Register in The Cochrane Library*. Oxford: Update Software; 2001	Excluded – abstract only available. Assumed not case series but non-randomised comparison trials
Chen CI, Skingley P, Meyer RM. A comparison of elderly patients with aggressive histology lymphoma who were entered or not entered on to a randomized Phase II Trial. *Leuk Lymphoma* 2000;**38**:327–34	Obtained – excluded at full text stage – not empirical examination of case series characteristics and impact on validity
Guyatt GH, DiCenso A, Farewell V, Willan A, Griffith L. Randomized trials versus observational studies in adolescent pregnancy prevention. *J Clin Epidemiol* 2000; **53**:167–74	Excluded – compares effect (observational trials indicate greater effect than RCTs) but no quality assessment of trials
Linde K, Scholz M, Melchart D, Willich SN. Should systematic reviews include non-randomized and uncontrolled studies? The case of acupuncture for chronic headache. *J Clin Epidemiol* 2002;**55**:77–85	Obtained – excluded at full text stage – no empirical examination of case series characteristics and impact on validity
Meade MO, Cook DJ, Kernerman P, Bernard G. How to use articles about harm: the relationship between high tidal volumes, ventilating pressures, and ventilator-induced lung injury. *Crit Care Med* 1997;**25**:1915–22	Excluded – describes hierarchy of evidence using articles on harm as an example. Quality aspects beyond study design are not discussed
Petitti DB. Coronary heart disease and estrogen replacement therapy. Can compliance bias explain the results of observational studies? *Ann Epidemiol* 1994;**4**:115–18	Excluded – no discussion of quality of observational studies
Radford MJ, Foody JM. How do observational studies expand the evidence base for therapy? *JAMA* 2001;**286**:1228–30	Obtained – excluded at full text stage – observational studies have control groups
Skovlund E. A critical review of papers from clinical cancer research. *Acta Oncol* 1998; **37**:339–45	Obtained – excluded at full text stage – no empirical examination of case series characteristics and impact on validity
Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, *et al*. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000;**283**:2008–12	Obtained – excluded at full text stage – observational studies have control groups
HTA database, 19 hits, 3 papers requested	
CRD. Evaluating non-randomised interventions	Ongoing study
Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, *et al*. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;**7**(27).	Obtained – excluded at full text stage – used for background
Eastwood A. Indirect comparisons of competing interventions	Obtained – excluded at full text stage – no empirical examination of case series characteristics and impact on validity

continued

Reference	Included or excluded?
National Research Register: 32 hits, one paper requested	
Agency for Healthcare Research and Quality. *Systems to Rate the strength of scientific evidence*. Evidence Report/Technology Assessment No. 47. Rockville: Agency for Healthcare Research and Quality;2002	Excluded – observational studies assessed but assumed to have a control group
Handsearched journals – requested papers	
American Journal of Epidemiology	
Giovannucci E. Meta-analysis of coffee consumption and risk of colorectal cancer *Am J Epidemiol* 1998;**147**:1043–52	Excluded – cohort and case–control studies only
Korte JE, Brennan P, Henley SJ, Boffetta P. Dose-specific meta-analysis and sensitivity analysis of the relation between alcohol consumption and lung cancer risk. *Am J Epidemiol* 2002;**155**:496–506	Excluded – epidemiological data meta-analysis – cohorts with reference groups, case–controls
Pladevall-Vila M, Delclos GL, Varas C, Guyer H, Brugues-Tarradellas J, Anglada-Arisa A. Controversy of oral contraceptives and risk of rheumatoid arthritis: meta-analysis of conflicting studies and review of conflicting metaanalyses with special emphasis on analysis of heterogeneity. *Am J Epidemiol* 1996;**144**:1–14	Excluded – only included case–control and cohort studies
Stram DO, Huberman M, Wu AH. Is residual confounding a reasonable explanation for the apparent protective effectiveness of beta-carotene found in epidemiologic studies of lung cancer in smokers? *Am J Epidemiol* 2002;**155**:622–8	Excluded – epidemiological. Looks at β-carotene as a confounder. β-Carotene is actually a marker of actual smoking intake, rather than as offering protective effect against lung cancer
Controlled Clinical Trials	
Dunn D, Babiker A, Hooker M, Darbyshire J. The dangers of inferring treatment effects form observational data: a case study in HIV infection. *Control Clin Trials* 2002;**23**:106–10	Excluded – only controlled sample discussed
Gorkin L, Schron EB, Handshaw K, Shea S, Kinney MR, Branyon M, *et al*. Clinical trial enrollers vs. nonenrollers: the Cardiac Arrhythmia Suppression Trial (CAST) Recruitment and Enrollment Assessment in Clinical Trials (REACT) project. *Control Clin Trials* 1996;**17**:46–59	Obtained – excluded at full text – used for background
Heitjan DF. Causal inference in a clinical trial: a comparative example. *Control Clin Trials* 1999;**20**:309–18	Excluded – no case series data
Omoigui NA, Topol EJ.Observational versus randomised medical device testing before and after market approval – the atherectomy-versus-angioplasty controversy. *Control Clin Trials* 1995;**16**:143–9	Excluded – discussion only, no mention of quality of RCTs or observational studies
International Journal of Health Technology Assessment	
Clarke M, Clarke T. A study of the references used in Cochrane protocols and reviews. *Int J Health Technol Assess* 2000;**16**:907–9	Excluded – only distinguishes 'journal articles', not type of research
Granados A. Health technology assessment and clinical decision making: which is the best evidence? *Int J Health Technol Assess* 1999;**15**:585–614	Excluded – opinion paper
Hyde CJ. Using the evidence: a need for quantity, not quality? *Int J Health Technol Assess* 1996;**12**:280–7	Excluded – opinion paper
Journal of Epidemiology and Community Health	
Freemantle N, Wood J, Crawford F. Evidence into practice, experimentation and quasi experimentation: are the methods up to the task? *J Epidemiol Community Health* 1998;**52**:75–81	Excluded – discussion only, no new data
Hotopf M, Lewis G, Normand C. Putting trials on trial – the costs and consequences of small trials in depression: a systematic review of methodology. *J Epidemiol Community Health* 1997;**51**:354–8	Excluded – no case series studies included
Jefferson T, Demicheli V. Relation between experimental and non-experimental study designs. HB vaccines: a case study. *J Epidemiol Community Health* 1999;**53**:51–4	Excluded – no assessment of study quality

Appendix 4

Checklists for evaluating quality of case series studies

DuRant (1994).[36] Checklist for the evaluation of research articles

I. Introduction

(a) Is the review of the previous research appropriate and sufficient? Have the relevant studies been cited and discussed?
(b) Is the problem to be studied clearly stated?
(c) Is the significance of the problem established?
(d) Have the authors established a theoretical framework for their study?
(e) Are the theoretical terms or concepts clearly described and defined?
(f) Are the objectives or the hypotheses clearly stated?
(g) Does the literature review provide a justification for the hypotheses (do the hypotheses logically flow from the literature review)?
(h) Do the hypotheses logically flow from the theoretical model?

II. Methods and procedures

(a) Are the methods that were selected appropriate to test the hypotheses adequately?
(b) Is there evidence of protection of human subjects in terms of the study being approved by an institutional review board?
(c) Is the study design:
    1. Experimental or quasi-experimental (go to III).
    2. Survey or cross-sectional (go to V).
    3. Retrospective chart (medical record) reviews and retrospective study (go to VII).
    4. Case–control study (go to VIII).

III. Experimental or quasi-experimental designs

(a) Has the study sample been clearly described in terms of sample size and demographic characteristics such as age, gender, location, socioeconomic status, etc.?
(b) Do the authors describe how the subjects were selected? Were they selected randomly, haphazardly, convenience sample, clinic population, etc.?
(c) What were the selection-eligibility criteria?
(d) Were the selection-eligibility criteria applied without knowledge of the specific treatment regimens to which the patients were being assigned?
(e) Did the selection criteria have an impact upon the subject's response to the treatment? For example, were subjects selected because they scored either very high or very low on a particular scale or were patients at low risk or high risk of contracting a particular disease selected for study?
(f) How were subjects assigned to experimental groups? (Any method besides random indicates that the study is a quasi-experimental design.)
(g) If subjects were randomly assigned to treatment and control groups, how was randomisation accomplished? Was a random numbers table used? (Methods such as alternating assignments, coin tossing, picking numbers out of a hat are not random.)
(h) Were individual subjects randomly assigned to treatment and control groups or were subjects assigned to treatment and control groups in blocks or groups?
(i) If subjects were randomly assigned to experimental groups on an individual basis, is it possible that subjects within treatment and control groups may have interacted, leading to a contamination of the treatment effect?
(j) If subjects were assigned to experimental groups *en bloc*, were a sufficient number of blocks included in each treatment group to ensure adequate statistical power?
(k) Were subjects blinded as to which experimental group they were assigned?
(l) Was the individual measuring the outcome variable(s) blinded to the experimental group that the subject was assigned?
(m) If the subjects had knowledge of which experimental group they were in, did this knowledge influence the subjects' responses to either the treatment or control interventions?
(n) If the investigator measuring the outcome variable was not blinded, was the outcome

variable measured in such a way that such knowledge could bias this measurement?

(o) Do the investigators clearly describe the treatment effect or intervention? Are the outcome, independent and control variables measured with appropriate and accurate methods? Do the operational definitions of the variables match the theoretical definitions?

(p) Have the laboratory tests, instruments and/or questionnaires used to measure the variables undergone validity and reliability testing?

(q) Have the procedures or methods used to measure each of the variables undergone standardisation for the particular population that is being studied?

(r) Did the subjects in the control or comparison group receive the exact same experimental procedures and measurements as the subjects in the treatment group, except for the treatment intervention?

(s) Was there strict adherence to the protocol?

(t) Were the side-effects from the treatment and control interventions clearly described?

(u) Was compliance with the treatment and control intervention clearly described and was compliance measured with an appropriate method?

(v) Was compliance different in the treatment and control groups?

(w) Was subject attrition discussed adequately?

(x) Was attrition kept to less than 10% in both groups?

(y) If a multi-centre trial was used, what methods were used to ensure that the experiment was conducted the same at all centres?

(z) Do the investigators compare the results from the different centres prior to pooling the data for final analysis?

IV. Statistical analysis for experimental designs

(a) Were between-group comparisons made at the pretest period and then at the post-test, or do the investigators assess the results using within-group comparisons, assessing pretest–post-test differences within each experimental group?

(b) Do the investigators demonstrate a lack of statistical differences in the pretest measurements between the control and the treatment groups? If not, was a covariance analysis used?

(c) If the investigators indicate that a *t*-test for two independent means was used to analyse the data, were the following assumptions met:
   1. two and only two groups are compared
   2. that the outcome variable is measured on an interval, ratio or continuous level scale.
   3. that the variances of the measurement of the outcome variable are similar for both the treatment and control group.
   4. that the measurement of the outcome variable is normally distributed (in a bell-shaped curve), or was the sample size large enough to invoke the central limit theorem?

(d) If more than two groups are compared do the investigators use an analysis of variance test? Note: the assumptions of the analysis of variance test are the same as the *t*-test except that three or more groups can be compared simultaneously.

(e) If an analysis of variance test is used is it followed by an appropriate multiple comparison test? (Go to IX.)

V. Survey designs and cross-sectional studies

(a) Are the criteria for inclusion of subjects described?

(b) Has the study sample been clearly described in terms of sample size and demographic characteristics such as age, race, gender, location, socioeconomic status, etc.?

(c) Is the study sample appropriate to the problem being studied or the hypotheses being tested?

(d) Is the study sample large enough to test the hypotheses?

(e) How was the study sample selected (random, haphazard, consecutive patients presenting with a particular disease, all subjects in a particular group, etc.)?

(f) Is the design of the study clearly described?

(g) Does the design of the study adequately test the hypotheses?

(h) How was random selection of subjects achieved? Was any other method besides the use of a random numbers table used?

(i) Have the measurement of the outcome, independent and control variables been clearly described?

(j) Are the variables measured with appropriate and accurate methods? Do the operational definitions match the theoretical variables?

(k) Have the laboratory tests, instruments and/or questionnaires used to measure the variables undergone validity and reliability testing?

(l) Have the procedures or methods used to measure each of the variables undergone standardisation for the particular population that is being studied?

(m) Were the outcome variables measured using appropriate 'blinded' methods?
(n) Have the number of non-respondents, refusals and subjects lost to the follow-up been kept reasonably small (less than 10%)?
(o) Was there strict adherence to the protocol?

VI. Statistical analysis for survey designs and cross-sectional studies

(a) Were the statistical tests used to analyse the data clearly described?
(b) Were the statistical tests chosen to analyse the data appropriate in terms of
  – adequately testing the hypotheses?
  – matching the study or research design?
  – meeting the statistical assumptions of the distribution of the data and the types of scales that were used to measure the outcome, independent and control variables?
  – the manner in which the sample was selected (random vs other)?
  – sample size?
(c) In most cases, survey designs require multivariate statistical tests to test the hypotheses adequately. Examples of such tests are multiple regression analysis, multivariate analysis of variance, discriminative function analysis, logistic regression analysis and factor analysis. Were any of these tests used and were they used appropriately? (Go to IX.)

VII. Retrospective chart (medical record) reviews and retrospective studies

(a) Was this study designed as a pilot study to assess the feasibility of doing a prospective study or was it designed as a definitive test of a hypothesis?
(b) What method was used to identify patients and their medical records? Was the total targeted population identified and measured?
(c) Over what time period was the record review conducted?
(d) Were there changes in procedures, diagnostic tests, medical technology and treatments, etc., during the time period? How were these changes handled?
(e) Did secular trends occur in cause and effect relationships during the time period (i.e. changes in diet and its relationship to heart disease)?
(f) Were information and data collected in a standardised manner?
(g) Were the definitions of disease and other variables exact, specific and clearly defined?
(h) How many people reviewed the medical records? Was interobserver or reviewer reliability assessed?
(i) Was the information in the medical records complete?
(j) How were missing data handled?

Many of the same questions asked concerning survey designs are appropriate for chart review. First answer the questions in Sections V and VI and then go to IX.

VIII. Case–control studies

(a) Case–control studies use a retrospective design and often require the review of the cases' and controls' medical records. If the study includes collecting data from the medical record first go to Section VII and answer questions a to j.
(b) How does the investigator control for recall bias? Are multiple methods used to measure important variables that could be influenced by recall bias?
(c) Does the problem or disease being studied suggest that recall bias may differ for cases and controls?
(d) Are the list of factors found to be significantly associated with the disease or outcome specific to that disease? If several non-specific factors are associated with the disease, does this suggest a differential recall bias for cases and controls?
(e) How were the comparison subjects selected?
  – one control per case selected in a non-random fashion
  – one control per case selected randomly from a matched pool of subjects
  – several controls per case selected randomly from a pool of subjects
  – several controls per case selected randomly from two or more pools of subjects.
(f) Were the controls appropriate for the hypothesis that was tested? Do they represent people like the general population or like people who have filtered through the health care system?
(g) Were controls matched to cases?
(h) Were the variables chosen to match controls or cases adequate to reduce competing explanations for the outcome or disease in question?
(i) Did wasted matching occur, i.e. did the investigator match cases and controls on variables that have no relationship to the study?
(j) Did overmatching occur? Did the investigator match on possible aetiological agents?

(k) What kind of a population do the cases represent? Are they a heterogeneous representation of the disease or outcome in question or a highly selected population for whom responses have limited generalisation?
(l) Are other biases evident? Do we know more about cases because they have been under closer surveillance, volunteered more information or been subject to more extensive testing than control subjects?

IX. Results section

(a) Are the findings presented clearly, objectively and in sufficient detail to enable the reader to judge the results for himself/herself?
(b) Are the findings internally consistent, i.e. do the numbers add up properly, can the tables be reconciled, etc.?
(c) Is there sufficient analysis to determine whether significant differences may in fact be due to the lack of comparability in sex or age distribution in clinical characteristics, or in other relevant variables?
(d) Were the appropriate variables or factors controlled for or blocked during the analysis?
(e) Were other potentially confounding variables handled appropriately?
(f) Was the number of subjects studied sufficiently large to avoid concluding that no relationship exists when in fact a significant relationship may have existed?
(g) Was the sample size so large that clinically insignificant results were declared statistically significant?
(h) Do the investigators present sufficient data in the tables and in the text to evaluate the results adequately?
(i) Are adequate summary data presented in the tables (i.e. are continuous level data presented as means ± SDs?)?
(j) Were appropriate probability levels (p-values) used to determine statistical significance?
(k) Do the investigators avoid retrospective hypothesis testing?

X. Discussion section

(a) Do the investigators consider all possible logical interpretations of their results?
(b) Are the conclusions clearly stated?
(c) Are the conclusions substantiated by the data that are presented in the results section?
(d) Do the investigators avoid introducing new results in the discussion?
(e) Are the results adequately compared to the previous studies in this area?
(f) Are the results adequately discussed in relation to the theoretical model chosen to develop the hypotheses?
(g) Are generalisations confined to the population from which the sample was drawn?
(h) Are the limitations of the study considered and are they taken into consideration when conclusions are drawn?
(i) Are recommendations for future research made?

Littenberg and colleagues (1998).[65] Closed fractures of the tibial shaft. Quality assessment of all types of study

1. Were reviewers of outcomes blinded to treatment?
3. Were more than 85% of the patients in each group followed up?
4. Were subjective (patient-reported) outcomes described?
5. Was follow-up active (meaning that patients were checked at prespecified intervals regardless of whether they had complaints) rather than passive (meaning that a complaint triggered an assessment)

Score 3 points for yes, 2 for probably yes, 1 for probably no, 0 for no.

Finally: RCT = 3 points, non-randomised comparative study = 2 points, case series = 0 points.

Maximum score = 15.

McAweeney and colleagues (1997).[66] Psychosocial interventions in the rehabilitation of people with spinal cord injury: a comprehensive methodological enquiry

Each paper was given a score of 0–5 for each of the elements (0 = no criteria met, 5 = all met).

Research evaluation criteria

Adequate statistical power
Effect size calculated
Confidence intervals stated
Type 1 error reported
Adequate description of non-participants
Clinical limitations stated

Hypothesis stated
Reported the validity of measures
Measurement limitations stated
Control of error attempted
Reported the reliability of measures
Method limitations stated
Control of environment
Adequate criteria for entry
Discuss generalisations of the results
Adequate selection of subjects
Discuss implications of the results
Variables defined
Contribution to spinal cord injury literature
Completeness of study
Hypothesis matches the design
Overall
Adequate review of the literature
Reported the study measures used
Practical significance of the results
Cutting edge of available research
Descriptive statistics are provided
Conceptualisation
Administration of measures
Design matches hypothesis
Concise review
Methods are clear
Clear problem stated
Purpose stated
Appropriate statistics
Use of *p*-values

Range of scores
Grand mean
Median
Standard deviation

Nielsen and Reilly (1985).[67] A guide to understanding and evaluating research articles

Was the research based upon sound, current theory in gifted education?

- An author's review of the literature, justifying the need for a particular study, should reflect the theory upon which that study was founded.
- Articles cited should include those by noted authorities in gifted education and be up-to-date.
- The literature review should demonstrate the need for this research and its value to the field of gifted education.

Is the problem to be investigated clearly stated?

- The problem statement should be free of highly specialised vocabulary and be easily understood by the average reader.
- The problem may be stated in narrative form, such as 'Students in a pull-out gifted programme will increase their ability to think critically'. This statement becomes the research hypothesis.
- The problem also may be stated using statistical notation, and then be termed the statistical hypothesis. For example, the statement 'the mean score of the experimental group will be greater than the mean score of the control group' would be written in statistical notation as 'H_1: $\mu_e > \mu_c$.' H_1 is the hypothesis to be tested; μ_e is the mean or average score for the population from which the experimental group was taken, μ_c is the mean or average score of the population from which the control group was taken.

Are the variables clearly defined or recognisable?

- An independent X variable is a factor that may be manipulated or varied by the experimenter (i.e. the amount of time spent by a student in gifted classes, the type of programme or the instructional material used).
- A dependent or Y variable is one that is expected to change as the result of the manipulations of the independent variable (i.e. a student's score on an achievement test).
- The researcher should state how the variables were measured (i.e. using achievement tests or teacher rating scales).

Is the population to be studied adequately described?

- Ages, gender, socioeconomic status and other information relevant to the study must be presented. This will assist in assuring that the research results were not confused or confounded by any of these variables.
- The method used to select the sample from the population must be explained.

Was the study carefully designed?

- The procedure used to set up the study and collect the data should be logical, clearly stated, and capable of future replication by others.
- The type of research used (experimental, descriptive, correlational, etc.) should be evident. Each type has its own limitations and requires that appropriate test statistics be applied to the data.

Did the research establish the reliability and validity of test instruments that were used?

- The reliability of a measuring instrument is the degree to which that instrument is dependable, consistent and stable over time. Reliability statements provide information regarding the precision and accuracy of a measurement. They define the magnitude of discrepancies between true ability and a measurement of ability. A reliability coefficient of 1.00 would indicate perfect reliability, although in educational research this situation never exists. However, the nearer to 1.00 the reliability coefficient is, the more reliable the measure.
- The validity of a measuring instrument states the degree to which that instrument measures what the researcher thinks it is measuring. There are three major types of validity: content validity, criterion-related validity and construct validity. Content validity attempts to determine whether the instrument has been constructed adequately in order to be representative of the substance and topics to be examined. Criterion-related validity is determined by comparing the instrument's results with an outside measure, predictor or criterion, such as grade-point average. The higher the correlation between the two, the more valid is the instrument. Construct validity is established when the researcher simultaneously defines some concept, construct or variable and develops the instrument to measure it. As with reliability, the nearer to 1.00 a validity coefficient is, the more valid the measure.
- If a new test is developed by a researcher, reliability and validity should be reported as determined from the data.
- If a standardised test is used, reliability and validity results from the test manual should be reported. Additionally, the researcher should determine the reliability of the test when applied to the subject population in the reported study.

Are the stated results consistent with the statistical data?

- When interpreting the results, statistics are used to estimate to what degree, if any, the results are likely to have occurred by chance alone. In educational research, a 0.05 or 0.01 'level of significance' is customarily chosen. If a 0.05 level was selected, the probability of obtaining the results by chance alone would only be 5% or less ($p \leq 0.05$).
- Researchers also report confidence levels, again commonly 0.05 or 0.01, when interpreting data. Confidence levels indicate the probability that the actual mean score of a population lies within a given range (confidence interval). Using the example μ (mean score) = 75.0 ± 2.5 and a 0.05 level of confidence, one could predict that the true mean score of the population would fall between 77.5 and 72.5, with a 95% probability of being correct.

Were the stated conclusions consistent with the results?

- The conclusions should be easily understood by the reader.
- Any conclusions which state that one variable caused an effect upon another variable must be consistent with and demonstrated by the statistical results.
- The author must be cautious in forming any generalisations or in suggesting that the same results will be found in groups not included in the study.

Did the author suggest areas of further research?

- The researcher should describe those areas that need further examination.
- The author should suggest practical recommendations as to how the results or conclusions of the study could be implemented.

Shay and colleagues (1972).[68] The factorial validity of a rating scale for the evaluation of research articles

Characteristics (scales) on the evaluation form	Factors						
	I	II	III	IV	V	VI	VII
1. Problem is clearly stated							
2. Hypotheses are clearly stated							
3. Problem is significant							
4. Assumptions are clearly stated							
5. Limitations of the study are stated							
6. Important terms are defined							
7. Relationship of the problem to previous research is made clear							
8. Research design is described fully							
9. Research design is appropriate for the solution of the problem							
10. Research design is free of specific weaknesses							
11. Population and sample are described							
12. Method of sampling is appropriate							
13. Data-gathering methods or procedures are described							
14. Data-gathering methods or procedures are appropriate to the solution of the problem							
15. Data-gathering methods or procedures are utilised correctly							
16. Validity and reliability of the evidence gathered are established							
17. Appropriate methods are selected to analyse the data							
18. Methods utilised in analysing the data are applied correctly							
19. Results of the analysis are presented clearly							
20. Conclusions are clearly stated							
21. Conclusions are substantiated by the evidence presented							
22. Generalisations are confined to the population from which the sample was drawn							
23. Report is clearly written							
24. Report is logically organised							
25. Tone of the report displays an unbiased, impartial scientific attitude							

Related factor loadings above 0.40 derived from principal components solution of intercorrelations of judges' ratings on each of 25 characteristics of 125 research articles.

The descriptive expressions of each of the five steps for each of the 25 rating scales were as follows: (1) completely incompetent, (2) poor, (3) mediocre, (4) good and (5) excellent. The factors were tentatively identified as follows:

I. Method of analysis
II. Design
III. Sampling
IV. Rigour
V. Significance
VI. Hypothesis
VII. Exposition
VIII. Objectivity

Each expert rated his/her article on the 25 characteristics. A product–moment correlation matrix was calculated for the ratings assigned by the 125 judges to the 25 characteristics. Principal components were extracted using the squared multiple R_1 for each characteristic with the other characteristics as the estimate of its communality. A comparison of the factors derived from the factor analysis with the *a priori* categorisation made by the Committee revealed one major difference, i.e. 'data gathering' did not emerge as a factor, although three scales (13, 14 and 15) had been specifically designed to evaluate this aspect of the research articles. It would appear that the judges did not view data gathering as a separate entity but subsumed such procedures under Research Design and, to a lesser degree, under Analysis of Data and Sampling.

Sheldon and colleagues (1993).[69] Critical appraisal of the medical literature: how to assess whether health-care interventions do more harm than good

A. Appropriateness of study design to objectives?					
	1. Good?	2. Fair?	3. Poor?		
B. Study population and sample					
B.1. Study population					
	1. Good?	2. Fair?	3. Poor?	4. ?	
B.2. Description of sample:					
	1. Good?	2. Fair?	3. Poor?	4. ?	
B.3. Entry criteria and exclusions:					
	1. Good?	2. Fair?	3. Poor?	4. N/A	5. ?
B.4. Sample method:					
	1. Good?	2. Fair?	3. Poor?	4. ?	
B.5. *A priori* estimate of required sample size?					
	1. Good?	2. Fair?	3. Poor?	4. N/A	
B.6. Sample size:					
Adequate?	1. Good?	2. Fair?	3. Poor?	4. N/A	5. ?
B.7. Sample representative of study population?					
	1. Good?	2. Fair?	3. Poor?	4. N/A	5. ?
B.8. Sample representative of target population					
	1. Good?	2. Fair?	3. Poor?	4. N/A	5. ?
C. Control group					
C.1. Description of controls:					
	1. Good?	2. Fair?	3. Poor?	4. N/A	5. ?
C.2. Adequacy of controls:					
	1. Good?	2. Fair?	3. Poor?	4. N/A	5. ?
C.3. Random treatment allocation:					
	1. Good?	2. Fair?	3. Poor?	4. N/A	
C.4. Randomisation tested?					
	1. Good?	2. Fair?	3. Poor?	4. N/A	
C.5. Matching to control confounding?					
	1. Good?	2. Fair?	3. Poor?		
C.6. Adequacy of matching?					
	1. Good?	2. Fair?	3. Poor?	4. N/A	5. ?
C.7. Group comparability?					
	1. Good?	2. Fair?	3. Poor?	4. N/A	5. ?

D. Interventions					
D.1. Therapeutic intervention:					
Standardisation:	1. Good?	2. Fair?	3. Poor?	4. N/A	5. ?
D.2. Use of placebo?					
	1. Good?	2. Fair?	3. Poor?		
D.3. Placebo intervention:					
Standardisation:	1. Good?	2. Fair?	3. Poor?	4. N/A	5. ?
D.4. Adequacy of placebo?					
	1. Good?	2. Fair?	3. Poor?	4. N/A	5. ?
E. Measurements/outcomes					
E.1. Measurements/outcomes used:					
Validity:	1. Good?	2. Fair?	3. Poor?	4. N/A	5. ?
E.2. Reproducibility of measures/outcomes:					
	1. Good?	2. Fair?	3. Poor?	4. N/A	5. ?
E.3. Blinding of subjects:					
	1. Absolute?	2. Partial?	3. No?	4. N/A	
E.4. Blinding of observers:					
	1. Absolute?	2. Fair?	3. No?	4. N/A	
F. Completeness					
F.1. Compliance (%):					
Adequacy?	1. Good?	2. Fair?	3. Poor?	4. N/A	5.
F.2. Withdrawal (%):					
Adequacy?	1. Good?	2. Fair?	3. Poor?	4. N/A	5.
F.3. Non respondents (%):					
Adequacy?	1. Good?	2. Fair?	3. Poor?	4. N/A	5.
F.4. Extent of missing data:					
	1. Minimal	2. Moderate	3. Extensive	4. N/A	
F.5. Analysis performed on basis of intention-to-treat:					
	1. Good?	2. Fair?	3. Poor?	4. No?	5. N/A
	6. ?				
G. Statistical analysis					
G.1. Presentation of descriptive statistics:					
	1. Good?	2. Fair?	3. Poor?	4. ?	
G.2. Stratification (control confounding):					
Appropriate?	1. Good?	2. Fair?	3. Poor?	4. N/A	
Adequacy?	1. Good?	2. Fair?	3. Poor?	4. N/A	5. ?
G.3. Multivariate method (control confounding):					
Appropriate?	1. Good?	2. Fair?	3. Poor?	4. N/A	
Adequacy?	1. Good?	2. Fair?	3. Poor?	4. N/A	5. ?

G.4. Statistical analysis for point estimate:					
Appropriate?	1. Good?	2. Fair?	3. Poor?	4. N/A	
Adequacy?	1. Good?	2. Fair?	3. Poor?	4. N/A	5. ?
G.5. Statistical analysis for interval estimate:					
Appropriate?	1. Good?	2. Fair?	3. Poor?	4. N/A	
Adequacy?	1. Good?	2. Fair?	3. Poor?	4. N/A	5. ?
G.6. Statistical power analysis for negative results:					
	1. Good?	2. Partial?	3. Poor?	4. N/A	5. ?
H. Others					

Spitzer and colleagues (1990).[33] Links between passive smoking and disease: a best-evidence synthesis

	Yes	Uncertain/incomplete/substandard	No	Don't know/Not reported	N/A	N/C	Comments
1. Random assignment, properly done							
2. Suitable choice of reference group							
3. Similar methods of data collection for all groups							
4. Proper sampling or suitable assembly of comparison group							
5. Sample size a. enables adequately precise estimates of priority variables found to be significant b. enables adequate precision in secondary variables reported (confounding variables or incidental findings) c. power reported for non-significant findings d. power declared *a priori* e. clinical or practical significance of statistically significant differences set forth or justified							
6. Criteria for definition or measurement of the outcomes are objective or verifiable							
7. Definition of exposure; unambiguous and measurable							

continued

	Yes	Uncertain/incomplete/substandard	No	Don't know/Not reported	N/A	N/C	Comments
8. Measurement of exposure; accurate and verifiable							
9. Blind assessment							
10. Observation bias minimised by design or accounted for in analysis							
11. Selection bias accounted for							
12. Objective criteria for eligibility of subjects (inclusion and exclusion)							
13. Attrition rates (%) a. response rate b. losses to follow-up c. other							
14. Known confounders accounted for a. by design b. by analysis							
15. Any methods to attempt comparability between groups, other than randomisation 16. Comparability of groups under comparison demonstrated 17. Appropriate statistical analytic plan a. evidence that *a priori* hypotheses being tested b. correct method used c. adjustment made for – multiple comparisons – simultaneous multiple range testing d. display of raw data permits assessment of actual measures and adjustments or transformations made							
18. Conclusions supported by data presented							
19. Reproducibility of method(s)							
20. Generalisability of results a. from sample(s) to parent population b. from sample(s) to any relevant population							
21. Other, specify							

Quality assessment used by Taylor and colleagues (2002).[58] Spinal cord stimulation for chronic low back pain

Design of study

RCT
Non-randomised trial
Prospective cohort
Retrospective cohort
Case–control
Case series
Unclear

Case series – detailed quality assessment

1. Selection bias

Were the patients consecutive cases? Yes No Can't tell
If not, was it a representative sample? Yes No Can't tell

Sampling method: _____

If not, were those included shown to be the same as the total treated? Yes No Can't tell

How? _____

2. Sampling bias

In addition to SCS did the patient receive any co-interventions? Yes No Can't tell

List: _____

3. Detection bias

Were the cases prospective? Yes No Can't tell

Detail: _____

If not, was there assessment of outcome before and after the intervention?
Yes No Can't tell

Detail: _____

Was there assessment of outcome made by an independent or blinded assessor?
Yes No Can't tell

Detail: _____

If not, was the assessment of outcomes carried out by a blinded assessor?
Yes No Can't tell

Detail: _____

Were outcomes assessed using objective/validated measures? Yes No Can't tell

Detail: _____

4. Attrition bias

Was there loss to follow up for the patient series? Yes No Can't tell

If yes, what was the level of loss to follow-up? _____ %

Items included in quality checklists and scales that can be used for case series studies

Cowley[35] and Coleridge Smith[71] also produce separate checklists for other types of study design. Only those for case series are shown here.

	DuRant, 1994[36]	**McAweeney *et al.*, 1997[66]**	**Nielsen and Reilly, 1985[67]**	**Shay *et al.*, 1972[68]**	**Sheldon *et al.*, 1993[69]**	**Spitzer *et al.*, 1990[33]**	**Taylor *et al.*, 2002[58]**	**Littenberg *et al.*, 1998[65]**	**Boulware *et al.*, 2002[34]**	**Cowley, 1995[35]**	**Coleridge Smith, 1999[37]**	**CRD Report 4**	**Avis, 1994[16]**	**Bass *et al.*, 1993[72]**
Number of items	32, +32 for ex, +23 for CS, +33 for chart, +12 for CC	36	24	25	35	33	7	5	14	17	20	6	24	7
Type of studies appropriate for	Exp, quasi-exp, CS, CC	Exp, quasi-exp, correlation studies	All	All	RCTs, quasi-exp, cohort, CC, CS	All	All	All	RCTs CS, CC	CS	CS	CS	All	All
Previous research reviewed/ comprehensive bibliography	✓	✓	✓	✓	–	–	–	–	–	–	–	–	–	–
Clear aims/question	✓	✓	✓	✓	–	–	–	–	–	–	–	–	✓	–
Significance of problem established	✓	–	–	✓	–	–	–	–	–	–	–	–	✓	–
Contributes to existing literature	–	✓	–	–	–	–	–	–	–	–	–	–	–	–
Relevant to your clinical practice	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Is the author an authority?	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Theoretical terms described/defined	✓	–	✓	–	–	–	–	–	–	–	–	–	–	–
Hypothesis stated	✓	✓	✓	✓	–	✓	–	–	–	–	–	–	–	–
Hypotheses flow from literature review	✓	–	✓	–	–	–	–	–	–	–	–	–	–	–
Hypotheses flow from theoretical model	✓	–	–	–	–	–	–	–	–	–	–	–	–	–
Clinical/study limitations stated	–	✓	–	✓	–	–	–	–	–	–	–	–	–	–
Methods clear/reproducible	–	✓	✓		–	✓	–	–	–	–	–	–	–	–
Use of control group	Different questions for different study designs	–	–		✓	–	–	–	–	–	–	–	–	✓

continued

	DuRant, 1994[36]	McAweeney *et al.*, 1997[66]	Nielsen and Reilly, 1985[67]	Shay *et al.*, 1972[68]	Sheldon *et al.*, 1993[70]	Spitzer *et al.*, 1990[33]	Taylor *et al.*, 2002[58]	Littenberg *et al.*, 1998[65]	Boulware *et al.*, 2002[34]	Cowley, 1995[35]	Coleridge Smith, 1999[37]	CRD Report 4	Avis, 1994[16]	Bass *et al.*, 1993[72]
Appropriate study design used	✓	✓	✓	✓	✓	–	✓	–	–	–	–	–	✓	–
Application of hierarchy of evidence	Different questions for different study designs	–	–	–	–	–	–	✓	–	–	–	–	–	✓
Ethics approval gained/ subjects' rights not harmed	✓	–	–	–	–	–	–	–	–	–	–	–	✓	–
Study design adequately described	✓	–	–	✓	–	–	–	–	–	–	–	–	–	–
Protocol described/adhered to	✓	–	–	–	–	–	–	–	–	–	–	–	–	–
Methods to ensure multi-centre trials are same at each centre	✓	–	–	–	–	–	–	–	–	–	–	–	–	–
Each centre analysed separately as well as pooled	✓	–	–	–	–	–	–	–	–	–	–	–	–	–
Control of environment	–	✓	–	–	–	–	–	–	–	–	–	–	–	–
Study size	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Prospective enrolment	–[a]	–	–	–	–	–	✓	–	–	–	–	–	–	✓
Consecutive cases	–	–	–	–	–	–	✓	–	–	–	–	–	–	–
Number of subjects approached, recruited and lost to follow-up given	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Specification of condition using recognised criteria	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Clear description of method of data collection	–	–	–	✓	–	–	–	–	–	–	–	–	–	–
Data collection methods appropriate/same for both groups	–	–	–	✓	–	✓	–	–	–	–	–	–	✓	–
Data collection methods used correctly	–	–	–	✓	–	–	–	–	–	–	–	–	–	–
Interviewers trained	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Participants all assessed at same point	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Variables clearly defined	–	✓	✓	–	–	–	–	–	–	–	–	–	–	–

continued

	DuRant, 1994[36]	McAweeney *et al.*, 1997[66]	Nielsen and Reilly, 1985[67]	Shay *et al.*, 1972[68]	Sheldon *et al.*, 1993[70]	Spitzer *et al.*, 1990[33]	Taylor *et al.*, 2002[58]	Littenberg *et al.*, 1998[65]	Boulware *et al.*, 2002[34]	Cowley, 1995[35]	Coleridge Smith, 1999[37]	CRD Report 4	Avis, 1994[16]	Bass *et al.*, 1993[72]
Variable measures clearly described	–	–	✓	–	–	–	–	–	–	–	✓	–	–	–
Specific types of variables scored Variable	–	–	–	–	–	–	–	–	–	–	–	–	–	✓
Selection bias accounted for	–	–	–	–	–	✓	–	–	–	–	–	–	–	–
Sampling fully described	✓	–	✓	✓	✓	–	–	–	✓	✓	–	–	–	–
Appropriate sampling/ representative sample	✓	✓	✓	✓	✓	–	✓	–	–	✓	–	✓	✓	✓
Adequate sample size/ statistical power/Type I error reported	✓	✓	–	–	✓	✓	–	–	✓	–	–	–	✓	–
Power defined *a priori*	–	–	–	–	–	✓	–	–	–	–	–	–	–	–
Clinical significance of statistical significance justified	–	–	–	–	–	✓	–	–	–	–	–	–	–	–
Explicit inclusion/exclusion criteria	✓	✓	–	–	✓	✓	–	–	✓	–	✓	✓		–
Patients entered study at similar point in disease	–	–	–	–	–	–	–	–	–	–	–	✓	–	–
Method of allocation/ randomisation	✓	–	–	–	✓	✓	–	–	✓	–	–	–	–	✓
Block or individual assignment	✓	–	–	–	–	–	–	–	–	–	–	–	–	–
Randomisation tested	–	–	–	–	✓	–	–	–	–	–	–	–	–	–
Subjects blind to treatment group	✓	–	–	–	✓	–	–	–	✓	–	–	–	–	–
Control group described	–	–	–	–	✓	–	–	–	–	–	–	–	–	–
Rationale for type of control groups used	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Adequacy of controls	–	–	–	–	✓	✓	–	–	✓	–	–	–	✓	–
Comparability of groups	–	–	–	–	✓	✓	–	–	–	–	✓	–	✓	–
Matched to control confounding?	–	–	–	–	✓	–	–	–	–	–	–	–	✓	–
Adequacy of matching	–	–	–	–	✓	–	–	–	–	–	–	–	–	–
Description of patients/cases	✓	–	✓	✓	–	–	–	–	–	✓	✓	–	–	–
Actual patients used (not volunteers/students)	–	–	–	–	–	–	–	–	–	–	–	–	–	–

continued

	DuRant, 1994[36]	McAweeney *et al.*, 1997[66]	Nielsen and Reilly, 1985[67]	Shay *et al.*, 1972[68]	Sheldon *et al.*, 1993[70]	Spitzer *et al.*, 1990[33]	Taylor *et al.*, 2002[58]	Littenberg *et al.*, 1998[65]	Boulware *et al.*, 2002[34]	Cowley, 1995[35]	Coleridge Smith, 1999[37]	CRD Report 4	Avis, 1994[16]	Bass *et al.*, 1993[72]
Characteristics analysed at baseline	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Non-participants described	–	✓	–	–	–	–	✓	–	✓	–	–	–	–	–
Description of intervention	–	–	–	–	–	–	–	–	✓	✓	✓	–	–	–
Description of operators' skill/experience	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Intervention standardised	–	–	–	–	✓	✓	✓	–	–	–	–	–	–	–
Intervention executed as designed	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Use of placebo	–	–	–	–	✓	–	–	–	–	–	–	–	–	–
Placebo standardised	–	–	–	–	✓	–	–	–	–	–	–	–	–	–
Placebo adequate	–	–	–	–	✓	–	–	–	–	–	–	–	–	–
Appropriate run-in period	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Treatment compliance	✓	–	–	–	✓	–	–	–	–	–	–	–	–	–
Comparability of treatment received and method of effect measurement in both groups	✓	–	–	–	–	–	–	–	✓	–	–	–	–	–
Loss to follow-up described/sufficient	✓	–	–	–	✓	✓	✓	✓	✓	✓	✓	✓	✓	–
Drop-outs compared with drop-outs?	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Active follow-up	–	–	–	–	–	–	–	✓	–	–	–	✓	–	–
Intention-to-treat analysis	–	–	–	–	✓	–	–	–	–	–	–	–	–	–
Effect size calculated	–	✓	–	–	✓		–	–	–	–	–	–	–	–
Definition of outcomes	✓	✓	–	–	–		–	–	–	–	✓	–	–	–
Outcomes valid/blind/ objective/reliable/independent	✓	✓	✓	✓	✓	✓	✓	✓	–	✓	✓	✓	✓	✓
Observation bias minimised	–	–	–	–	–	✓	–	–	–	–	–	–	–	–
Measures/outcomes reproducible/verifiable/accurate	–	–	–	–	✓	✓	✓	–	–	–	–	–	✓	–
Limitations of outcomes	–	✓	–	–	–	–	–	–	–	–	–	–	–	–
All relevant outcomes included (e.g. adverse effects)	✓	–	–	–	–	–	–	–	–	–	–	–	–	–
Multiple change indices	–	–	–	–	–	–	–	–	–	–	–	–	–	–

continued

	DuRant, 1994[36]	McAweeney *et al.*, 1997[66]	Nielsen and Reilly, 1985[67]	Shay *et al.*, 1972[68]	Sheldon *et al.*, 1993[70]	Spitzer *et al.*, 1990[33]	Taylor *et al.*, 2002[58]	Littenberg *et al.*, 1998[65]	Boulware *et al.*, 2002[34]	Cowley, 1995[35]	Coleridge Smith, 1999[37]	CRD Report 4	Avis, 1994[16]	Bass *et al.*, 1993[72]
Assessment pre- and post-intervention	✓	–	–	–	–	–	✓	–	–	✓	–	–	–	–
Patient-relevant outcomes	–	–	–	–	–	–	–	✓	–	–	–	–	–	–
Multiple vantage points to assess outcome	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Dose-response relationship/ temporal relationship	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Prognostic factors or confounders assessed/analysis stratified/multivariate methods	✓	–	–	–	✓	✓	–	–	✓	–	–	✓	✓	–
Appropriate statistical methods	✓	✓	–	✓	✓	✓	–	–	✓	✓	✓	–	–	–
Statistical methods applied correctly	–	–	–	✓	✓	✓	–	–	–	–	✓	–	–	–
Results presented clearly	–	–	–	✓	✓	–	–	–	–	–	–	–	–	–
Appropriate interpretation of results	✓	–	–	–	–	–	–	–	–	–	–	–	–	–
CI/p values presented	✓	✓	✓	–	✓	–	–	–	✓	–	–	–	–	–
Results clinically and statistically significant	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Adjustment made for multiple comparisons/multiple range testing	–	–	–	–	–	✓	–	–	–	–	–	–	–	–
Sufficient result detail presented to assess	✓	–	–	–	–	✓	–	–	–	–	–	–	–	–
Data grouped at natural cut-off points/not forced	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Results internally consistent	✓	–	–	–	–	–	–	–	–	–	–	–	–	–
Statistical associations distinguished from causal relations?	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Could additional analyses be done?	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Retrospective hypotheses avoided	✓	–	–	–	–	–	–	–	–	–	–	–	–	–
Conclusions clear/all possible considered	✓	✓	✓	✓	–	–	–	–	–	–	–	–	–	–

continued

	DuRant, 1994[36]	McAweeney *et al.*, 1997[66]	Nielsen and Reilly, 1985[67]	Shay *et al.*, 1972[68]	Sheldon *et al.*, 1993[70]	Spitzer *et al.*, 1990[33]	Taylor *et al.*, 2002[58]	Littenberg *et al.*, 1998[65]	Boulware *et al.*, 2002[34]	Cowley, 1995[35]	Coleridge Smith, 1999[37]	CRD Report 4	Avis, 1994[16]	Bass *et al.*, 1993[72]
Conclusions contextualised with other studies	✓	✓	–	–	–	–	–	–	–	–	–	–	–	–
Conclusions discussed in relation to theoretical model	✓	–	–	–	–	–	–	–	–	–	–	–	–	–
Conclusions reasonable on basis of results	–	–	✓	✓	–	✓	–	–	–	–	–	–	✓	–
Conclusions made about drug tested	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Relevance/generalisability assessed	✓	✓	✓	✓	–	✓	–	–	–	–	–	–	–	–
Limitations of study considered	✓	✓	–	–	–	–	–	–	–	–	–	–	–	–
Recommendations on how to implement results/ conclusions/policy implications	–	–	✓	–	–	–	–	–	–	–	–	–	–	–
Recommendations for future research	✓	–	✓	–	–	–	–	–	–	–	–	–	–	–
Report clearly written	–	–	–	✓	–	–	–	–	–	–	–	–	–	–
Report logically organised	–	–	–	✓	–	–	–	–	–	–	–	–	–	–
Unbiased, impartial scientific attitude	–	–	–	✓	–	–	–	–	–	✓	–	–	–	–
Will the results affect your practice?	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Type of publication	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Public funding	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Summary score?	No	Yes	No	Yes	No	No	No	Yes	Yes	No	No	No	No	Yes

	Brown, 1991[73]	Bracken, 1989[74]	Brown et *al.*, 1996[75]	Campos-Outcalt et *al.*, 1995[76]	Cho and Bero, 1994[77]	Cox and Merkel, 1989[78]	Cuijpers, 1998[79]	Downs and Black, 1998[17]	Fowkes and Fulton, 1991[80]	Garber et *al.*, 1996[81]	Haynes et *al.*, 1975 [82]	Kreulen et *al.*, 1998[83]
Number of items	6	36	6	7	16	14	At least 6 – not all stated	26[b]	23	6	6	16
Type of studies appropriate for	RCT, non-RCT, pre- and post-test	Obs	RCTs and non-RCTs	All	All	All	RCTs pre- and post-test	RCT, coh, CC	All	All	All	All
Previous research reviewed/ comprehensive bibliography	–	✓	–	–	–	–	–		–	–	–	–
Clear aims/question	–	–	–	–	–	–	–		–	–	–	✓
Significance of problem established	–	–	–	–	–	–	–		–	–	–	–
Contributes to existing literature	–	–	–	–	–	–	–		–	–	–	–
Relevant to your clinical practice	–	–	–	–	–	–	–		–	–	–	–
Is the author an authority?	–	–	–	–	–	–	–		–	–	–	–
Theoretical terms described/defined	–	–	–	✓	–	–	–		–	–	–	–
Hypothesis stated	–	✓	–	–	–	–	–		–	–	–	–
Hypotheses flow from literature review	–	–	–	–	–	–	–		–	–	–	–
Hypotheses flow from theoretical model	–	–	–	–	–	–	–		–	–	–	–
Clinical/study limitations stated	–	–	–	–	–	–	–		–	–	–	–
Methods clear/reproducible	–	–		–	–	–	–		–	–	–	✓
Use of control group	✓	–	–	–	–	–	✓		–	–	–	–
Appropriate study design used	–	✓	–	–	✓	–			✓	✓	–	✓
Application of hierarchy of evidence	–	–	✓	✓	–	–	–		–	–	✓	–
Ethics approval gained/subjects' rights not harmed	–	✓	–	–	–	–	–		–	–	–	–
Study design adequately described	–	✓	–	–	–	–	–		–	–	–	–
Protocol described/adhered to	–	–	–	–	–	–	–		–	–	–	✓
Methods to ensure multi-centre trials are same at each centre	–	–	–	–	–	–	–		–	–	–	–
Each centre analysed separately as well as pooled	–	–	–	–	–	–	–		–	–	–	–

continued

	Brown, 1991[73]	Bracken, 1989[74]	Brown et al., 1996[75]	Campos-Outcalt et al., 1995[76]	Cho and Bero, 1994[77]	Cox and Merkel, 1989[78]	Cuijpers, 1998[79]	Downs and Black, 1998[17]	Fowkes and Fulton, 1991[80]	Garber et al., 1996[81]	Haynes et al., 1975 [82]	Kreulen et al., 1998[83]
Control of environment	–	–	–	–	–	–	–		–	–	–	–
Study size	–	–	–	✓	–	–	–		–	–	–	–
Prospective enrolment	–	–	–	–	–	–	–		–	–	–	–
Consecutive cases	–	–	–	–	–	–	–		–	–	–	–
Number of subjects approached, recruited and lost to follow-up given	–	✓	–	✓	✓	–	✓		✓	–	–	–
Specification of condition using recognised criteria	–	✓	–	–	–	–	✓		–	–	✓	–
Clear description of method of data collection	–	–	–	–	–	–	–		–	–	–	–
Data collection methods appropriate/same for both groups	–	✓	–	✓	–	–	–		–	–	–	–
Data collection methods used correctly	–	–	–	–	–	–	–		–	–	–	–
Interviewers trained	–	✓	–	–	–	–	–		–	–	–	–
Participants all assessed at same point	–	✓	–	–	–	✓	–		–	–	–	–
Variables clearly defined	–	✓	–	–	–	–	–		–	–	–	–
Variable measures clearly described	–	✓	–	–	–	–	–		–	–	–	–
Specific types of variables scored Variable	–	–	–	–	–	–	–		–	–	–	–
Selection bias accounted for	–	–	–	–	–	–	–		–	–	–	–
Sampling fully described	–	✓	✓	–	–	–	–		✓	–	–	✓
Appropriate sampling/ representative sample	✓	✓	–	–	✓	–	–		✓	–	✓	–
Adequate sample size/statistical power/Type I error reported	–	✓	–	–	✓	✓	–		✓	–	–	✓
Power defined *a priori*	–	✓	–	–	✓	–	–		–	–	–	–
Clinical significance of statistical significance justified	–	✓	–	–	–	–	–		–	–	–	–
Explicit inclusion/exclusion criteria	–	–	–	–	✓	–	–		✓	–	–	–
Patients entered study at similar point in disease	–	–	–	–	–	–	–		–	–	–	–

continued

	Brown, 1991[73]	Bracken, 1989[74]	Brown et *al.*, 1996[75]	Campos-Outcalt et *al.*, 1995[76]	Cho and Bero, 1994[77]	Cox and Merkel, 1989[78]	Cuijpers, 1998[79]	Downs and Black, 1998[17]	Fowkes and Fulton, 1991[80]	Garber et *al.*, 1996[81]	Haynes et *al.*, 1975 [82]	Kreulen et *al.*, 1998[83]
Method of allocation/ randomisation	✓	–	–	–	–	✓	✓		✓	–	–	✓
Block or individual assignment	–	–	–	–	–	–	–		–	–	–	–
Randomisation tested	–	–	–	–	–	–	–		–	–	–	–
Subjects blind to treatment group	–	–	–	–	✓	–	–		–	–	–	–
Control group described	–	✓	–	–	–	–	–		✓	–	–	✓
Rationale for type of control groups used	–	✓	–	–	–	–	–		✓	–	–	–
Adequacy of controls	–	✓	–	–	✓	–	–		–	–	–	–
Comparability of groups	–	–	–	–	–	–	–		✓	–	–	–
Matched to control confounding?	–	✓	–	–	–	–	–		–	✓	–	–
Adequacy of matching	–	–	–	–	–	–	–		–	–	–	–
Description of patients/cases	–	–	–	–	–	–	–		–	–	–	✓
Actual patients used (not volunteers/students)	–		–	–	–	–	–		–	–	–	–
Characteristics analysed at baseline	–	–	–	–	–	–	–		–	–	–	–
Non-participants described	–	–	–	–	–	–	–		–	–	–	–
Description of intervention	✓	–	✓	–	–	–	–		–	–	✓	✓
Description of operators' skill/experience	–	–	–	–	–	–	–		–	–	–	✓
Intervention standardised	–	–	–	–	–	✓	–		–	–	–	–
Intervention executed as designed	–	–	–	–	–	✓	–		–	–	–	–
Use of placebo	–	–	–	–	–	–	–		–	–	–	–
Placebo standardised	–	–	–	–	–	–	–		–	–	–	–
Placebo adequate	–	–	–	–	–	–	–		–	–	–	–
Appropriate run-in period	–	–	–	–	–	–	–		–	–	–	–
Treatment compliance	–	–	–	–	–	–	–		–	–	✓	–
Comparability of treatment received and method of effect measurement in both groups	–	–	–	–	–	✓	–		✓	–	–	–
Loss to follow-up described/sufficient	–	✓	–	✓	–	–	✓		✓	–	–	✓

continued

	Brown, 1991[73]	Bracken, 1989[74]	Brown et al., 1996[75]	Campos-Outcalt et al., 1995[76]	Cho and Bero, 1994[77]	Cox and Merkel, 1989[78]	Cuijpers, 1998[79]	Downs and Black, 1998[17]	Fowkes and Fulton, 1991[80]	Garber et al., 1996[81]	Haynes et al., 1975 [82]	Kreulen et al., 1998[83]
Drop-outs compared with drop-outs?	–	–	–	–	–	–	–		–	–	–	–
Active follow-up	–	✓	–	✓	–	✓	✓		–	–	–	✓
Intention-to-treat analysis	–	–	–	–	–	–	–		–	–	–	–
Effect size calculated	–	–	–	–	–	–	–		–	–	–	–
Definition of outcomes	✓	–	–	–	–	–	–		–	–	–	✓
Outcomes valid/blind/objective/reliable/independent	✓	–	–	–	✓	✓	✓		✓	–	–	✓
Observation bias minimised	–	–	–	–	✓	–	–		✓	–	–	–
Measures/outcomes reproducible/verifiable/accurate	✓	–	✓	–	–	–	–		✓	–	–	✓
Limitations of outcomes	–	–	–	–	–	–	–		–	–	–	–
All relevant outcomes included (e.g. adverse effects)	–	–	–	–	–	✓	–		–	–	–	–
Multiple change indices	–	–	–	–	–	✓	–		–	–	–	–
Assessment pre- and post-intervention	–	–	–	–	–	✓	–		–	–	–	–
Patient-relevant outcomes	–	–	–	–	–	–	–		–	–	–	✓
Multiple vantage points to assess outcome	–	–	–	–	–	✓	–		–	–	–	–
Dose–response relationship/temporal relationship	–	–	–	–	–	–	–		✓	✓	–	✓
Prognostic factors or confounders assessed/analysis stratified/multivariate methods	–	✓	–	–	✓	–	–		✓	–	–	–
Appropriate statistical methods	–	–	–	✓	✓	✓	–		–	–	–	–
Statistical methods applied correctly	–	–	–	–	–	–	–		–	–	–	–
Results presented clearly	–	✓	–	–	–	–	–		–	–	–	✓
Appropriate interpretation of results	–	✓	–	–	–	–	–		–	–	–	–
CI/*p*-values presented	–	✓	–	–	–	–	–		–	–	–	✓
Results clinically and statistically significant	–	–	–	–	–	–	–		–	✓	–	–

continued

	Brown, 1991[73]	Bracken, 1989[74]	Brown et al., 1996[75]	Campos-Outcalt et al., 1995[76]	Cho and Bero, 1994[77]	Cox and Merkel, 1989[78]	Cuijpers, 1998[79]	Downs and Black, 1998[17]	Fowkes and Fulton, 1991[80]	Garber et al., 1996[81]	Haynes et al., 1975 [82]	Kreulen et al., 1998[83]
Adjustment made for multiple comparisons/multiple range testing	–	✓	–	–	–	–	–		–	–	–	–
Sufficient result detail presented to assess	–	✓	–	–	–	–	–		–	–	–	–
Data grouped at natural cut-off points/not forced	–	✓	–	–	–	–	–		–	–	–	–
Results internally consistent	–	–	–	–	–	–	–		–	–	–	–
Statistical associations distinguished from causal relations?	–	✓	–	–	–	–	–		–	–	–	–
Could additional analyses be done?	–	✓	–	–	✓	–	–		–	–	–	–
Retrospective hypotheses avoided	–	✓	–	–	–	–	–		–	–	–	–
Conclusions clear/all possible considered	–	–	–	–	–	–	–		–	–	–	–
Conclusions contextualised with other studies	–	✓	–	–	–	–	–		–	–	–	–
Conclusions discussed in relation to theoretical model	–	–	–	–	–	–	–		–	✓	–	–
Conclusions reasonable on basis of results	–	✓	–	–	✓	–	–		–	–	–	–
Conclusions made about drug tested	–	–	–	–	✓	–	–		–	–	–	–
Relevance/generalisability assessed	–	–	–	–	✓	–	–		–	–	–	–
Limitations of study considered	–	–	–	–	–	–	–		–	–	–	–
Recommendations on how to implement results/conclusions/ policy implications	–	✓	–	–	–	–	–		–	–	–	–
Recommendations for future research	–	–	–	–	–	–	–		–	–	–	–
Report clearly written	–	–	–	–	–	–	–		–	–	–	–
Report logically organised	–	–	–	–	–	–	–		–	–	–	–
Unbiased, impartial scientific attitude	–	–	–	–	–	–	–		–	–	–	–

continued

	Brown, 1991[73]	Bracken, 1989[74]	Brown et al., 1996[75]	Campos-Outcalt et al., 1995[76]	Cho and Bero, 1994[77]	Cox and Merkel, 1989[78]	Cuijpers, 1998[79]	Downs and Black, 1998[17]	Fowkes and Fulton, 1991[80]	Garber et al., 1996[81]	Haynes et al., 1975 [82]	Kreulen et al., 1998[83]
Will the results affect your practice?	–	–	–	–	–	–	–		–	–	–	–
Type of publication	–	–	–	–	–	–	–		–	–	–	–
Public funding	–	–	–	–	–	–	–		–	–	–	–
Summary score?	Yes	No	No	Yes	No	Yes	No		No	Yes	No	Yes

	Krogh, 1985[84]	Massy et al., 1995[85]	Meijman and Melker, 1995[86]	Morris et al., 1988[87]	Rey and Walter, 1997[88]	Rowe et al., 1997[89]	Stock, 1991[90]
Number of items	11	15	11	23	7	Not all stated	7
Type of studies appropriate for	All	Controlled and non-controlled	All	All	All	All	All
Previous research reviewed/comprehensive bibliography	✓	–	–	–	–	–	–
Clear aims/question	–	–	✓	–	–	–	–
Significance of problem established	✓	–	–	–	–	–	–
Contributes to existing literature	–	–	–	–	–	–	–
Relevant to your clinical practice	✓	–	–	–	–	–	–
Is the author an authority?	✓	–	–	–	–	–	–
Theoretical terms described/defined	–	–	–	–	–	–	–
Hypothesis stated	–	–	✓	–	–	–	–
Hypotheses flow from literature review	–	–	–	–	–	–	–
Hypotheses flow from theoretical model	–	–	–	–	–	–	–
Clinical/study limitations stated	–	–	–	–	–	–	–
Methods clear/reproducible	–	–	–	–	–	–	–
Use of control group	–	✓	–	✓	–	✓	–
Appropriate study design used	–	–	✓	–	–	–	–
Application of hierarchy of evidence	–	–	✓	–	–	–	–
Ethics approval gained/subjects' rights not harmed	–	–	–	–	–	–	–
Study design adequately described	–	–	–	–	–	–	–
Protocol described/adhered to	–	–	–	–	–	–	–

continued

	Krogh, 1985[84]	Massy *et al.*, 1995[85]	Meijman and Melker, 1995[86]	Morris *et al.*, 1988[87]	Rey and Walter, 1997[88]	Rowe *et al.*, 1997[89]	Stock, 1991[90]
Methods to ensure multi-centre trials are same at each centre	–	–	–	–	–	–	–
Each centre analysed separately as well as pooled	–	–	–	–	–	–	–
Control of environment	–	–	–	–	–	–	–
Study size	–	–	–	–	–	–	–
Prospective enrolment	–	–	–	–	–	–	–
Consecutive cases	–	–	–	–	–	–	–
Number of subjects approached, recruited and lost to follow-up given	–	–	–	–	–	–	✓
Specification of condition using recognised criteria	–	–	–	–	✓	–	–
Clear description of method of data collection	–	–	–	–	–	–	–
Data collection methods appropriate/same for both groups	–	–	–	–	–	–	–
Data collection methods used correctly	–	–	–	–	–	–	–
Interviewers trained	–	–	–	–	–	–	–
Participants all assessed at same point	–	–	–	–	–	–	–
Variables clearly defined	–	–	–	–	–	–	–
Variable measures clearly described	–	–	–	–	–	✓	–
Specific types of variables scored Variable	–	–	–	–	–	–	–
Selection bias accounted for	–	–	–	–	–	–	✓
Sampling fully described	–	–	✓	–	–	–	–
Appropriate sampling/representative sample	–	–	–	–	–	–	–
Adequate sample size/statistical power/Type I error reported	✓	–	–	✓	–	–	–
Power defined *a priori*	–	–	–	–	–	–	–
Clinical significance of statistical significance justified	–	–	–	–	–	–	–
Explicit inclusion/exclusion criteria	–	✓	✓	–	–	✓	–
Patients entered study at similar point in disease	–	–	–	–	–	–	–
Method of allocation/randomisation	–	✓	–	✓	–	✓	–
Block or individual assignment	–	–	–	–	–	–	–
Randomisation tested	–	–	–	–	–	–	–

continued

	Krogh, 1985[84]	Massy et *al.*, 1995[85]	Meijman and Melker, 1995[86]	Morris et *al.*, 1988[87]	Rey and Walter, 1997[88]	Rowe et *al.*, 1997[89]	Stock, 1991[90]
Subjects blind to treatment group	–	✓	–	–	–	–	–
Control group described	–	–	✓	–	–	–	–
Rationale for type of control groups used	–	–	–	–	–	–	–
Adequacy of controls	✓	–	✓	–	–	–	–
Comparability of groups	–	–	✓	–	–	–	✓
Matched to control confounding?	–	–	–	–	–	–	✓
Adequacy of matching	–	–	–	–	–	–	–
Description of patients/cases	✓	–	✓	✓	✓	–	–
Actual patients used (not volunteers/students)	–	–	–	✓	–	–	–
Characteristics analysed at baseline	–	–	–	✓	–	–	–
Non-participants described	–	–	✓	–	–	–	✓
Description of intervention	–	–	–	–	✓	✓	–
Description of operators' skill/experience	–	–	–	✓	–	–	–
Intervention standardised	–	–	–	✓	–	–	–
Intervention executed as designed	–	–	–	–	–	–	–
Use of placebo	–	✓	–	–	–	–	–
Placebo standardised	–	–	–	–	–	–	–
Placebo adequate	–	✓	–	–	–	–	–
Appropriate run-in period	–	✓	–	–	✓	–	–
Treatment compliance	–	–	✓	–	–	✓	–
Comparability of treatment received and method of effect measurement in both groups	–	–	–	–	–	–	✓
Loss to follow-up described/sufficient	–	–	✓	✓	–	–	–
Drop-outs compared with drop-outs?	–	–	–	✓	–	–	–
Active follow-up	–	–	–	–	–	✓	–
Intention-to-treat analysis	–	–	–	–	–	–	–
Effect size calculated	–	–	–	–	–	–	–
Definition of outcomes	–	–	–	–	–	–	–
Outcomes valid/blind/objective/reliable/independent	–	✓	✓	✓	–	✓	✓
Observation bias minimised	–	–	–	–	–	–	–
Measures/outcomes reproducible/verifiable/accurate	–	–	✓	✓	–	–	–
Limitations of outcomes	–	–	–	–	–	–	–

continued

	Krogh, 1985[84]	Massy *et al.*, 1995[85]	Meijman and Melker, 1995[86]	Morris *et al.*, 1988[87]	Rey and Walter, 1997[88]	Rowe *et al.*, 1997[89]	Stock, 1991[90]
All relevant outcomes included (e.g. adverse effects)	–	–	–	–	✓	–	–
Multiple change indices	–	–	–	–	–	–	–
Assessment pre- and post-intervention	–	✓	–	–	–	–	–
Patient-relevant outcomes	–	–	–	–	–	–	–
Multiple vantage points to assess outcome	–	–	–	✓	–	✓	–
Dose-response relationship/temporal relationship	–	–	–	–	✓	–	–
Prognostic factors or confounders assessed/ analysis stratified/multivariate methods	–	–	✓	–	–	–	–
Appropriate statistical methods	–	–	✓	–	–	–	–
Statistical methods applied correctly	–	–	–	–	–	–	–
Results presented clearly	–	–	–	–	–	–	–
Appropriate interpretation of results	–	–	–	–	–	–	–
CI/*p*-values presented	–	–	–	–	–	✓	–
Results clinically and statistically significant	–	–	–	–	–	–	–
Adjustment made for multiple comparisons/ multiple range testing	–	–	–	–	–	–	–
Sufficient result detail presented to assess	–	–	–	–	–	–	–
Data grouped at natural cut-off points/not forced	–	–	–	–	–	–	–
Results internally consistent	–	–	–	–	–	–	–
Statistical associations distinguished from causal relations?	–	–	–	–	–	–	–
Could additional analyses be done?	–	–	–	–	–	–	–
Retrospective hypotheses avoided	–	–	–	–	–	–	–
Conclusions clear/all possible considered	✓	–	–	–	–	–	–
Conclusions contextualised with other studies	–	–	–	–	–	–	–
Conclusions discussed in relation to theoretical model	✓	–	✓	–	–	–	–
Conclusions reasonable on basis of results	✓	–	✓	–	–	–	–
Conclusions made about drug tested	–	–	–	–	–	–	–
Relevance/generalisability assessed	✓	–	–	–	–	–	–
Limitations of study considered	–	–	–	–	–	–	–

continued

	Krogh, 1985[84]	Massy et *al.*, 1995[85]	Meijman and Melker, 1995[86]	Morris et *al.*, 1988[87]	Rey and Walter, 1997[88]	Rowe et *al.*, 1997[89]	Stock, 1991[90]
Recommendations on how to implement results/conclusions/policy implications	–	–	–	–	–	–	–
Recommendations for future research	–	–	–	–	–	–	–
Report clearly written	✓	–	–	–	–	–	–
Report logically organised	–	–	–	–	–	–	–
Unbiased, impartial scientific attitude	–	–	–	–	–	–	–
Will the results affect your practice?	✓	–	–	–	–	–	–
Type of publication	✓		–	–	–	–	–
Public funding	✓		–	–	–	–	–
Summary score?	No	Yes	No	No	Yes	Yes	No

CC, case–control; CS, case series; coh, cohort; exp, experimental design; quasi-exp, quasi-experimental design.

[a] DuRant[36] also contains additional questions for retrospective record reviews that are not included here. Slightly different items are included for each type of included study design with some anomalies – for example, use of a sample size calculation is required for surveys but not for RCTs.

[b] Downs and Black[17] do not give details of the actual items included but cover the following main areas: reporting (nine items), external validity (three items), bias (seven items), confounding (six items) and power (one item).

Appendix 5

Papers excluded from angina search

Paper identified through search	Reason for exclusion
Bertelsen CA, Kjoller M, Hoier-Madsen K, Folke K, Fritz-Hansen P. Influence of complete revascularization on long-term survival after coronary artery bypass surgery. *Scand Cardiovasc J* 1997;**31**:271–4	Excludes in-hospital deaths
Diegeler A, Spyrantis N, Matin M, Falk V, Hambrecht R, Autschbach R, *et al*. The revival of surgical treatment for isolated proximal high grade LAD lesions by minimally invasive coronary artery bypass grafting. *Eur J Cardiothorac Surg* 2000;**17**:501–4	Minimally invasive CABG technology
Gould BL, Clayton PD, Jensen RL, Liddle HV. Association between early graft patency and late outcome for patients undergoing artery bypass graft surgery. *Circulation* 1984;**69**:569–76	Does not report appropriate outcomes
Hirzel HO, Eichhorn P, Kappenberger L, Gander MP, Schlumpf M, Gruentzig AR. Percutaneous transluminal coronary angioplasty: late results at 5 years following intervention. *Am Heart J* 1985;**109**:575–81	Only includes patients with successful PTCA
Kiebzak GM, Pierson LM, Campbell M, Cook JW. Use of the SF36 general health status survey to document health-related quality of life in patients with coronary artery disease: effect of disease and response to coronary artery bypass graft surgery. *Heart Lung* 2002; **31**:207–13	Does not report appropriate outcomes
Laarman G, Luijten HE, van Zeyl LG, Beatt KJ, Tijssen JG, Serruys PW, *et al*. Assessment of 'silent' restenosis and long-term follow-up after successful angioplasty in single vessel coronary artery disease: the value of quantitative exercise electrocardiography and quantitative coronary angiography. *J Am Coll Cardiol* 1990;**16**:578–85	Does not report appropriate outcomes
Lawrie GM, Morris GCJ. Survival after coronary artery bypass surgery in specific patient groups. *Circulation* 1982;**65**:43–8	Duplicate publication: Lawrie *et al*., 1982[91]
Pijls NH, Bech GJ, el Gamal MI, Bonnier HJ, De Bruyne B, Van Gelder B, *et al*. Quantification of recruitable coronary collateral blood flow in conscious humans and its potential to predict future ischemic events. *J Am Coll Cardiol* 1995;**25**:1522–8	Does not report appropriate outcomes
Rubin DA, Nieminski KE, Monteferrante JC, Magee T, Reed GE, Merman MV. Ventricular arrhythmias after coronary artery bypass graft surgery: incidence, risk factors and long term prognosis. *J Am Coll Cardiol* 1985;**6**:307–10	Does not report appropriate outcomes
Slagboom T, Kiemeneij F, Laarman GJ, van der Wieken R, Odekerken D. Actual outpatient PTCA: results of the OUTCLAS pilot study. *Catheter Cardiovasc Interv* 2001;**53**:204–8	Only 24-hour follow-up
Staudacher RA, Hess KR, Harris SL, Abu-Khalil J, Heibig J. Percutaneous transluminal coronary angioplasty utilizing prolonged balloon inflations: initial results and six-month follow-up. *Catheter Cardiovasc Diagn* 1991;**23**:239–44	Does not report appropriate outcomes
ten Berg JM, Bal ET, Gin TJ, Ernst JM, Mast EG, Ascoop CA, *et al*. Initial and long-term results of percutaneous transluminal coronary angioplasty in patients 75 years of age and older. *Catheter Cardiovasc Diagn* 1992;**26**:165–70	Duplicate publication: ten Berg *et al*., 1996[92]
Tsang J, Sheppard R, Mak KH, Brown D, Huynh T, Schechter D, *et al*. Six-month outcomes of percutaneous transluminal coronary angioplasty in hypertensive patients: results from the ROSETTA registry. Routine Versus Selective Exercise Treadmill Testing After Angioplasty. *Am Heart J* 2002;**143**:124–9	Only patients with successful PTCA included
Yli-Mayry S, Huikuri HV. Clinical and angiographic prediction of myocardial infarction and recurrence of severe angina during a five-year follow-up after coronary artery bypass grafting. *Am J Cardiol* 1993;**72**:1371–5	Patients who died of had angina excluded from the study

Appendix 6

Data extraction for exploration of study characteristics

Functional endoscopic surgery for nasal polyps

Study	Study design	Country	Publication date	Journal	No. of patients	With polyps (%)	Median age (years)	Male (%)	Preventive surgery (%)
Kurent and Zargi[93]	RCT	Slovenia	1998	ERS & ISIAN Meeting	20	100			
Penttila *et al.*[94]	RCT	Finland	1997	*Acta Otolaryngol Suppl*	75	69	47	40	24
Venkatachalam *et al.*[95]	RCT	India	1998	*JK Practitioner*	25	100		70	
Jankowski *et al.*[96]	Comparative	France	1997	*Acta Otolaryngol*	37	100	44	69	52
Unlu *et al.*[97]	Comparative	Turkey	1994	*J Otolaryngol*	50	28	36	50	
Harkness *et al.*[98]	Comparative	UK	1997	*Clin Otolaryngol*	1064	37	45	56	
Danielsen[99]	Case series	Norway	1992	*Clin Otolaryngol*	100	42		57	12
Danielsen and Olofsson[100]	Case series	Norway	1996	*Acta Otolaryngol*	230	40		19	20
Davis *et al.*[101]	Case series	USA	1991	*Otolaryngol Head Neck Surg*	200	74			
Delank and Stoll[102]	Case series	Germany	1998	*Rhinology*	115	77	44	55	0
Fortune and Duncavage[103]	Case series	USA	1998	*Ann Otol Rhinol Laryngol*	100	28	39	46	
Franzen and Klausen[104]	Case series	Norway	1994	*Clin Otolaryngol*	50	54	47	60	52
Friedman *et al.*[105]	Case series	USA	2000	*Am J Rhinol*	200	34	41		20
Friedman *et al.* [106]	Case series	USA	2000	*Otolaryngol Head Neck Surg*	500	27			0
Frisch *et al.*[107]	Case series	Denmark	1994	*Rhinology*	85	56	45	63	
Jacobs[108]	Case series	USA	1997	*Laryngoscope*	112	45			29
Jakobsen and Svenstrup[109]	Case series	Denmark	2000	*Acta Otolaryngol Suppl*	237	62	46	59	10
Jiang and Hsu[110]	Case series	Taiwan	2001	*Ear Nose Throat J*	1112	40		62	
Katsantonis *et al.*[111]	Case series	USA	1994	*Otolaryngol Head Neck Surg*	972	Unknown			
Kennedy[112]	Case series	USA	1992	*Laryngoscope*	120	59			71
Klossek *et al.*[113]	Case series	USA, France, Canada	1997	*Otolaryngol Head Neck Surg*	50	100	47	54	
Lawson[114]	Case series	USA	1991	*Laryngoscope*	90	59	49	67	40
Levine[115]	Case series	USA	1990	*Laryngoscope*	250	52			5
Lund and Mackay[116]	Case series	UK	1994	*J R Soc Med*	650	47			
Massegur *et al.*[117]	Case series	Spain	1995	*Rhinology*	250	81			23
Moses *et al.*[118]	Case series	USA	1998	*Ear Nose Throat J*	90	56	42	40	100
Nishioka *et al.*[119]	Case series	USA	1994	*Otolaryngol Head Neck Surg*	283	49	44	57	

continued

Study	Study design	Country	Publication date	Journal	No. of patients	With polyps (%)	Median age (years)	Male (%)	Preventive surgery (%)
Park *et al.*[120]	Case series	USA	1998	*J Otolaryngol*	79	73	50		56
Rice[121]	Case series	USA	1989	*Otolaryngol Head Neck Surg*	100	25		62	63
Roth *et al.*[122]	Case series	Israel	1995	*Int Surg*	100	42	35	66	19
Ryan *et al.*[123]	Case series	UK	1994	*Rhinology*	137	14	45		
Sato and Nakashima[124]	Case series	Japan	2000	*Laryngoscope*	10	100			
Schaefer[125]	Case series	USA	1998	*Laryngoscope*	509	27	42	51	50
Schaefer *et al.*[126]	Case series	USA	1989	*Laryngoscope*	100	41	39	50	
Schaitkin *et al.*[127]	Case series	USA	1993	*Laryngoscope*	100	42			
Shapshay *et al.*[128]	Case series	USA	1992	*Laryngoscope*	17	88	50	35	47
Sipila *et al.*[129]	Case series	Finland	1996	*Eur Arch Otorhinolaryngol*	51	8	47	43	
Sobol *et al.*[130]	Case series	Canada	1992	*J Otolaryngol*	393	47	45	50	32
Stammberger and Posawetz[131]	Case series	Austria	1990	*Eur Arch Otorhinolaryngol*	500	49			
Stoop et al.[132]	Case series	The Netherlands	1992	*Eur Arch Otorhinolaryngol*	72	100	44		
Venkatachalam and Bhat[133]	Case series	India	1999	*Ind J Otolaryngol Head Neck Surg*	210	31		58	19
Vleming and de Vries[134]	Case series	The Netherlands	1991	*Rhinology*	5	100		80	60
Weber *et al.*[135]	Case series	Germany, India	1997	*Am J Otolaryngol*	325	100			
Wigand and Hosemann[136]	Case series	Germany	1989	*Rhinology suppl*	220	100			
Wigand *et al.*[137]	Case series	Germany	1978	*Endoscopy*	315	Unknown			
Wolf *et al.*[138]	Case series	Austria	1995	*Rhinology*	124	42	12	48	

Study	Consecutive enrolment	Prospective	Random allocation?	Blind allocation?	ITT?	Independently measured outcome?	Loss to follow-up (%)	Average follow-up (months)	Multi-centre?
Kurent and Zargi[93]	Uncertain	Uncertain	No	No	No	No	15	36	No
Penttila *et al.*[94]	Uncertain	Uncertain	Uncertain	No	No	No			No
Venkatachalam *et al.* [95]	Uncertain	Uncertain	Uncertain	No	Yes	Uncertain	4	17	No
Jankowski *et al.*[96]	Uncertain	No	No	No	No	Yes	18	24	No
Unlu *et al.*[97]	Uncertain	No	No	No	No	No	23		No
Harkness *et al.*[98]	Uncertain	Uncertain	No	No	N/A	No			Yes
Danielsen[99]	Uncertain	Uncertain	N/A	N/A	Yes	Yes	0	14	No
Danielsen and Olofsson[100]	Uncertain	Yes	N/A	N/A	No	Yes	6	41	No
Davis *et al.*[101]	Yes	Yes	N/A	N/A	No	Yes	42	36	Yes
Delank and Stoll[102]	Uncertain	Yes	N/A	N/A	N/A	Yes	0		No
Fortune and Duncavage[103]	Yes	No	N/A	N/A	N/A	Uncertain			No
Franzen and Klausen[104]	Yes	Yes	N/A	N/A	Yes	Uncertain	0	24	No
Friedman *et al.*[105]	Yes	No	N/A	N/A	N/A	Uncertain	0	12	No
Friedman *et al.*[106]	Uncertain	Yes	N/A	N/A	N/A	No	0	10	No
Frisch *et al.*[107]	Yes	Uncertain	N/A	N/A	No	Uncertain	27		No
Jacobs[108]	Uncertain	No	N/A	N/A	No	Uncertain	10	16	No
Jakobsen and Svenstrup[109]	Yes	Yes	N/A	N/A	No	Uncertain	3	12	No
Jiang and Hsu [110]	No	No	N/A	N/A	No	Yes	39		No
Katsantonis *et al.*[111]	No	No	N/A	N/A	Uncertain	No		14	Yes
Kennedy[112]	Uncertain	Uncertain	N/A	N/A	Uncertain	Yes		18	Yes
Klossek *et al.*[113]	Uncertain	Yes	N/A	N/A	N/A	Yes	0		Yes
Lawson[114]	No	Yes	N/A	N/A	Uncertain	Uncertain		42	No
Levine[115]	Uncertain	No	N/A	N/A	No	Uncertain	12	17	No
Lund and Mackay[116]	Uncertain	Yes	N/A	N/A	N/A	Uncertain	0	6	No
Massegur *et al.*[117]	Uncertain	No	N/A	N/A	N/A	Yes	0		No
Moses *et al.*[118]	No	No	N/A	N/A	N/A	No	0	23	No
Nishioka *et al.*[119]	Uncertain	Yes	N/A	N/A	N/A	No	0	15	Yes
Park *et al.*[120]	No	No	N/A	N/A	N/A	Yes	0	19	No
Rice[121]	Yes	Uncertain	N/A	N/A	Yes	Uncertain	0	24	No
Roth *et al.*[122]	Yes	No	N/A	N/A	N/A	Uncertain		10	No

continued

Study	Consecutive enrolment	Prospective	Random allocation?	Blind allocation?	ITT?	Independently measured outcome?	Loss to follow-up (%)	Average follow-up (month)	Multi-centre?
Ryan *et al*[123]	Yes	No	N/A	N/A	No	Uncertain	12	3	No
Sato and Nakashima[124]	Uncertain	Uncertain	N/A	N/A	N/A	Yes		23	No
Schaefer[125]	No	No	N/A	N/A	N/A	No	0		No
Schaefer *et al.*[126]	Yes	Uncertain	N/A	N/A	N/A	Uncertain		5	No
Schaitkin *et al.*[127]	Yes	Uncertain	N/A	N/A	No	Uncertain	9		No
Shapshay *et al.*[128]	Uncertain	Uncertain	N/A	N/A	No	Yes	0	6	No
Sipila *et al.*[129]	Uncertain	Uncertain	N/A	N/A	Yes	Yes	0	3	No
Sobol *et al.*[130]	Uncertain	No	N/A	N/A	No	No	32	12	No
Stammberger and Posawetz[131]	No	No	N/A	N/A	Uncertain	Yes			No
Stoop *et al.*[132]	Uncertain	Yes	N/A	N/A	N/A	Uncertain	0		No
Venkatachalam and Bhat[133]	Uncertain	Uncertain	N/A	N/A	No	Uncertain	4	18	No
Vleming and de Vries[134]	Uncertain	Uncertain	N/A	N/A	Yes	Uncertain	0	4	No
Weber *et al.*[135]	Yes	No	N/A	N/A	No	No	52		Yes
Wigand and Hosemann[136]	Uncertain	Uncertain	N/A	Uncertain	Uncertain	No			No
Wigand *et al.*[137]	Yes	No	N/A	N/A	N/A	Uncertain	0		No
Wolf *et al.*[138]	Uncertain	No	N/A	N/A	No	Yes	43		No

Study	Symptomatic improvement (%)	Symptomatic improvement – ITT (%)	Polyp/disease recurrence (%)	Polyp/disease recurrence – ITT (%)	Revision surgery (%)	Revision surgery – ITT (%)	Patency (%)	Patency – ITT (%)
Kurent and Zargi[93]			30	30				
Penttila *et al.*[94]	78	78			21	19	57	
Venkatachalam *et al.* [95]	88	88						
Jankowski *et al.*[96]	72	57			14	11		
Unlu *et al.*[97]	85	68	8	6			87	
Harkness *et al.*[98]	82	82						
Danielsen[99]	95	95	4	4	8	8		
Danielsen and Olofsson[100]	40	39	27		6			
Davis *et al.*[101]	96	33					88	
Delank and Stoll[102]								
Fortune and Duncavage[103]								
Franzen and Klausen[104]	90	90			6	6		
Friedman *et al.*[105]	92				6		90	
Friedman *et al.* [106]	92							
Frisch *et al.*[107]	82	56			19	14		
Jacobs[108]	84	76					93	
Jakobsen and Svenstrup[109]	89	86			9			
Jiang and Hsu[110]								
Katsantonis *et al.*[111]								
Kennedy[112]	98		4				97	
Klossek *et al.*[113]	96						100	
Lawson[114]	73		27					
Levine[115]	83	85			4			
Lund and Mackay[116]	87				3			
Massegur *et al.*[117]	90		16		6			
Moses *et al.*[118]			33					
Nishioka *et al.*[119]					4		93	
Park *et al.*[120]	86							
Rice[121]	83	83	7	7	6	6		
Roth *et al.*[122]	79						73	

continued

Study	Symptomatic improvement (%)	Symptomatic improvement – ITT (%)	Polyp/disease recurrence (%)	Polyp/disease recurrence – ITT (%)	Revision surgery (%)	Revision surgery – ITT (%)	Patency (%)	Patency – ITT (%)
Ryan *et al.*[123]	78	69						
Sato and Nakashima[124]					0			
Schaefer[125]					3			
Schaefer *et al.*[126]	93				4	4		
Schaitkin *et al.*[127]	91	83			25	23		
Shapshay *et al.*[128]	59	59						
Sipila *et al.*[129]	76	76	24	24	6	6	71	71
Sobol *et al.*[130]	70	47			4			
Stammberger and Posawetz[131]	95							
Stoop *et al.*[132]			56					
Venkatachalam and Bhat[133]	89		7					
Vleming and de Vries[134]			0	0				
Weber *et al.*[135]	89	46						
Wigand and Hosemann[136]	82							
Wigand *et al.*[137]	76				5			
Wolf *et al.*[138]			16		34			

Data extraction–angina

Study	Study design	Country	Publication date	Intervention	Journal	No. of patients	Prospective	Registry	Consecutive enrolment	Lost to follow-up (%)	Average follow-up (years)	Objective/ blinded outcomes	Multi-centre
BARI[139]	RCT	USA	1997	CABG	*JAMA*	914	Yes			2	5.4		Yes
CABRI[140]	RCT	Europe	1995	CABG	*Lancet*	513	Yes		Yes		1		Yes
ERACI[141]	RCT	Argentina	1996	CABG	*J Am Coll Cardiol*	64	Yes			5			No
Goy *et al.*[142]	RCT	Switzerland	1994	CABG	*Lancet*	66	Yes		Yes	1.50	2		No
Hamm *et al.*[143]	RCT	Germany	1994	CABG	*N Engl J Med*	177	Yes			9			Yes
Hueb *et al.*[144]	RCT	Brazil	1995	CABG	*J Am Coll Cardiol*	70	Yes				3.5		No
Jones and Weintraub[145]	RCT	USA	1996	CABG	*J Thorac Cardiovasc Surg*	3890	Yes				5.3		No
King *et al.*[146]	RCT	USA	1994	CABG	*N Engl J Med*	194	Yes				3		No
Puel *et al.*[147]	RCT	France	1992	CABG	*Circulation*	52	Yes	No			2.8		No
RITA[148]	RCT	UK	1993	CABG	*Lancet*	501	Yes			2	2.5		Yes
Bonnier *et al.*[149]	Comparative	The Netherlands	1993	CABG	*Br Heart J*	81				0	8.3	Yes	No
Mick *et al.*[150]	Comparative	USA	1991	CABG	*Am J Cardiol*	142	No		Yes	0	3.3	No	No
Tyras *et al.*[151]	Comparative	USA	1980	CABG	*J Thorac Cardiovasc Surg*	184	No		Yes	4	4	Yes	No
Ullyot *et al.*[152]	Comparative	USA	1975	CABG	*J Thorac Cardiovasc Surg*	149			Yes	0	1.5		No
Acar *et al.*[153]	Case series	France	1998	CABG	*J Thorac Cardiovasc Surg*	102	No		Yes	0	5.3	No	
Acinapura *et al.*[154]	Case series	USA	1989	CABG	*Ann Thorac Surg*	3853	Yes				1.6		Yes
Acinapura *et al.*[155]	Case series	USA	1992	CABG	*J Cardiovasc Surg*	7470					8.5		Yes
Arnold *et al.*[156]	Case series	USA	1979	CABG	*Ann Thorac Surg*	282			Yes	0	5	Yes	
Ashor *et al.*[157]	Case series	USA	1973	CABG	*Arch Surg*	100	No		Yes	17	0.9	No	
Azariades *et al.*[158]	Case series	USA	1990	CABG	*Ann Thorac Surg*	1081	Yes	No		9		Yes	
Azariades *et al.*[158]	Case series	USA	1990	CABG	*Ann Thorac Surg*	1081	No	No		8	1.8	Yes	No
Baldwin *et al.*[159]	Case series	USA	1998	CABG	*Chest*	100	No		Yes	13	4	No	No
Barner *et al.*[160]	Case series	USA	1985	CABG	*J Thorac Cardiovasc Surg*	1000	Yes			8	6.3		
Bathgate and Irving[161]	Case series	UK	1997	CABG	*Heart*	102	No		Yes	5	10	Yes	No
Bell *et al.*[162]	Case series	USA	1992	CABG	*Circulation*	3372	No	Yes			4.9	Yes	Yes
Beretta *et al.*[163]	Case series	Italy	1990	CABG	*Eur J Cardiothorac Surg*	20			Yes	0	0.67	No	
Bergsma *et al.*[164]	Case series	The Netherlands	1998	CABG	*Circulation*	256	No			0	4.25	Yes	No

continued

Study	Study design	Country	Publication date	Intervention	Journal	No. of patients	Prospective	Registry	Consecutive enrolment	Lost to follow-up (%)	Average follow-up (years)	Objective/ blinded outcomes	Multi-centre
Brandup-Wognsen *et al.*[165]	Case series	Sweden	1995	CABG	*Thorac Cardiovasc Surg*	2000	Yes				2		Yes
Cameron *et al.*[166]	Case series	USA	1995	CABG	*J Am Coll Cardiol*	5289	No	Yes			4-8		
Canver *et al.*[167]	Case series	USA	1996	CABG	*Ann Thoracic Surg*	1689					10		
Carter[168]	Case series	Australia	1987	CABG	*Aust N Z J Surg*	30				0	0.45	No	No
Christakis *et al.*[169]	Case series	Canada	1993	CABG	*J Cardiac Surg*	1228	Yes	No		3	4.2		
Christenson and Schmuziger[170]	Case series	Switzerland	1997	CABG	*Ann Thorac Surg*	92			Yes	1		Yes	No
Egstrup[171]	Case series	Denmark	1988	CABG	*Am J Cardiol*	36	No		Yes		0.75	No	No
Farrer *et al.*[60]	Case series	UK	1997	CABG	*QJM*	353	Yes		Yes	10	1	Yes	No
Fitzgibbon *et al.*[172]	Case series	Canada	1996	CABG	*J Am Coll Cardiol*	1388							
French *et al.*[173]	Case series	New Zealand	1995	CABG	*Circulation*	221	No		Yes	5	10	Yes	No
Gale *et al.*[174]	Case series	Australia	1977	CABG	*Med J Aust*	543			Yes			No	No
Gelbfish *et al.*[175]	Case series	USA	1986	CABG	*Ann Thorac Surg*	28				7	1	No	No
Gelfand *et al.*[176]	Case series	Canada	1983	CABG	*Can J Surg*	92	No			33	3.5	No	No
Green *et al.*[177]	Case series	USA	1979	CABG	*J Thorac Cardiovasc Surg*	140			Yes	0	5	No	No
Higginbothom *et al.*[178]	Case series	Australia	1981	CABG	*Med J Aust*	35	No		Yes		2	Yes	No
Horgan *et al.*[179]	Case series	Eire	1981	CABG	*Ir Med J*	79			Yes		2	No	No
Ivert *et al.*[180]	Case series	Sweden	1989	CABG	*Eur J Cardiothorac Surg*	94	No		Yes	0	2.5	No	No
Jenkins *et al.*[181]	Case series	USA	1983	CABG	*JAMA*	318	Yes		No		0.5	No	No
Jones and Weintraub[145]	Case series	USA	1996	CABG	*J Thorac Cardiovasc Surg*	2860		No		1	12		Yes
Killen *et al.*[182]	Case series	USA	1982	CABG	*South Med J*	2628	Yes	No			5.3		
Killen *et al.*[183]	Case series	USA	1989	CABG	*Ann Thorac Surg*	266	No		Yes	0	12	Yes	
Killen *et al.*[184]	Case series	USA	1998	CABG	*Texas Heart Inst J*	648					8.9	No	No
Kornfeld *et al.*[185]	Case series	USA	1982	CABG	*Circulation*	100			Yes	8	4.5	No	No
Laird-Meeter *et al.*[186]	Case series	The Netherlands	1987	CABG	*Br Heart J*	1041	Yes	No			7.5		
Laks *et al.*[187]	Case series	USA	1978	CABG	*Am J Cardiol*	77			Yes		2.2	No	No
Lawrie *et al.*[91]	Case series	USA	1982	CABG	*Circulation*	500	Yes		Yes	9	10	No	No
Liao *et al.*[188]	Case series	USA	1992	CABG	*JAMA*	1719	No	No			4		
Lytle *et al.*[189]	Case series	USA	1984	CABG	*J Am Coll cardiol*	107	No		No	5	9.6	Yes	No

continued

Study	Study design	Country	Publication date	Intervention	Journal	No. of patients	Prospective	Registry	Consecutive enrolment	Lost to follow-up (%)	Average follow-up (years)	Objective/ blinded outcomes	Multi-centre
Maddern *et al.*[190]	Case series	Australia	1984	CABG	*Med J Aust*	4001				1.20			Yes
MacDonald *et al.*[191]	Case series	Canada	1998	CABG	*Can J Cardiol*	100	Yes		Yes		1	Yes	No
Morin *et al.*[192]	Case series	Canada	1992	CABG	*Ann Thorac Surg*	67	No		Yes	3	3	Yes	No
Morris *et al.*[193]	Case series	USA	1990	CABG	*Circulation*	1063	Yes	No			4		
Mullany *et al*[194]	Case series	USA	1990	CABG	*Circulation*	159	No			2	2.4	Yes	No
Nicholson and Paterson[195]	Case series	Australia	1997	CABG	*Ann Thorac Surg*	75	Yes		No				No
Ochsner *et al.*[196]	Case series	USA	1977	CABG	*Ann Thorac Surg*	100	No			2		No	No
Palatianos *et al.*[197]	Case series	USA	1993	CABG	*Ann Thorac Surg*	145	No			1	2.6	No	No
Patel *et al.*[198]	Case series	UK	1993	CABG	*Br Heart J*	76					5.8	No	No
Peterson *et al.*[199]	Case series	USA	1995	CABG	*Circulation*	172283	No	Yes			3		
Pinna-Pintor *et al.*[200]	Case series	Italy	1992	CABG	*Qual Life Res*	626			Yes	21	4.8		No
Rahimtoola *et al.*[201]	Case series	USA	1993	CABG	*J Am Coll Cardiol*	7529	Yes	No			20	Yes	
Rahimtoola *et al.*[202]	Case series	USA	1993	CABG	*Circulation*	8906	Yes			11			
Richardson and Cyrus[203]	Case series	USA	1986	CABG	*Ann Thorac Surg*	1089	Yes	No			5		
Risum *et al.*[204]	Case series	Norway	1995	CABG	*Cardiovasc Surg*	1025	Yes				6.5		
Risum *et al.*[205]	Case series	Norway	1996	CABG	*Scand J Thorac Cardiovasc Surg*	1025	Yes	No			7.2		
Ruygrok *et al.*[206]	Case series	New Zealand	1993	CABG	*Aust N Z J Med*	96			Yes	2	6.1	Yes	No
Saatvedt *et al.*[207]	Case series	Norway	1997	CABG	*Scand Cardiovasc J*	10	Yes			0	0.25	Yes	No
Salomon *et al.*[208]	Case series	USA	1990	CABG	*J Thorac Cardiovasc Surg*	7059	Yes	No					
Schaff *et al.*[209]	Case series	USA	1983	CABG	*Circulation*	500			Yes	0.20		No	No
Schmuziger *et al.*[210]	Case series	Switzerland	1994	CABG	*Cardiovasc Surg*	3103	Yes	No		19	1.4	Yes	Yes
Sheldon and Loop[211]	Case series	USA	1984	CABG	*Postgrad Med*	29373	No	No					No
Simmons *et al.*[212]	Case series	USA	1987	CABG	*Am J Cardiol*	73			Yes	0	1	No	No
Sterling *et al.*[213]	Case series	USA	1984	CABG	*Am Heart J*	54	No		Yes				No
Tector *et al.*[214]	Case series	USA	1986	CABG	*J Thorac Cardiovasc Surg*	100	No		Yes				No
Tschan *et al.*[215]	Case series	Switzerland	1985	CABG	*Chest*	218			No		4.5	Yes	No
Tyras *et al.*[151]	Case series	USA	1980	CABG	*J Thorac Cardiovasc Surg*	1459	No	No			3.7		
Ullyot *et al.*[216]	Case series	USA	1977	CABG	*J Thorac Cardiovasc Surg*	200	Yes		Yes	2	2.25	Yes	No

continued

Study	Study design	Country	Publication date	Intervention	Journal	No. of patients	Prospective	Registry	Consecutive enrolment	Lost to follow-up (%)	Average follow-up (years)	Objective/ blinded outcomes	Multi-centre
Verhiest *et al.*[217]	Case series	Belgium	1977	CABG	*Acta Cardiol*	200			Yes		1.8	No	No
Weintraub *et al.*[218]	Case series	USA	1995	CABG	*Circulation*	2030	Yes	No		1	4.3		
Wright[219]	Case series	Australia	1979	CABG	*Med J Aust*	122			Yes	3	6		
Cohen *et al*[220]	Case–control	USA	1986	CABG	*Chest*	123	No		No	7	5	Yes	Yes
Horneffer *et al.* [221]	Comparative	USA	1987	CABG	*Circulation*	228	Yes	No	Yes	5	2.5	No	No
Horneffer *et al.*[221]	Comparative	USA	1987	CABG	*Circulation*		Yes	No	Yes	2	2.5	No	No
Horneffer *et al.*[221]	Comparative	USA	1987	CABG	*Circulation*	228	Yes	No	Yes	2	2.5	No	No

PTCA for chronic stable angina – sample characteristics

Study	Study design	Country	Publication date	Intervention	Journal	No. of patients	Prospective	Registry	Consecutive enrolment	Lost to follow-up (%)	Average follow-up (years)	Objective/ blinded outcomes	Multi-centre
BARI[139]	RCT	America	1997	PTCA	*JAMA*	915	Yes			2	5.4		Yes
CABRI[140]	RCT	Europe	1995	PTCA	*Lancet*	541	Yes				1		Yes
ERACI[141]	RCT	Argentina	1993	PTCA	*J Am Coll Cardiol*	63	Yes			2	3		No
Goy *et al.*[142]	RCT	Switzerland	1994	PTCA	*Lancet*	68	Yes			0	2		No
Hamm *et al.*[143]	RCT	Germany	1994	PTCA	*N Engl J Med*	182	Yes			3	1		Yes
Hueb *et al.*[144]	RCT	Brazil	1995	PTCA	*J Am Coll Cardiol*	72	Yes				3.5		No
Jones and Weintraub *et al.*[145]	RCT	USA	1996	PTCA	*J Thorac Cardiovasc Surg*	2924	Yes			3	5.3		No
King *et al.*[146]	RCT	USA	1994	PTCA	*N Engl J Med*	198	Yes				3		Yes
Puel *et al.*[147]	RCT	France	1992	PTCA	*Circulation*	57	Yes				2.8	Yes	No
RITA[148]	RCT	UK	1993	PTCA	*Lancet*	510	Yes			3	2.5		Yes
Bonnier *et al.*[149]	Comparative	The Netherlands	1993	PTCA	*Br Heart J*	93				0	8.2	Yes	No
Cavallini *et al.*[222]	Comparative	Italy	1994	PTCA	*Am Heart J*	152			Yes		1.2	No	No
Jeroudi *et al.*[223]	Comparative	USA	1990	PTCA	*Ann Intern Med*	54	No		Yes	0	1.6	Yes	No
Kamp *et al.*[224]	Comparative	The Netherlands	1989	PTCA	*Am Heart J*	840	No		Yes	4	2	No	No
Meyer *et al.*[225]	Comparative	Germany	1983	PTCA	*Am Heart J*	100				37	0.5	Yes	No

continued

Study	Study design	Country	Publication date	Intervention	Journal	No. of patients	Prospective	Registry	Consecutive enrolment	Lost to follow-up (%)	Average follow-up (years)	Objective/ blinded outcomes	Multi-centre
Mick *et al.*[150]	Comparative	USA	1991	PTCA	*Am J Cardiol*	53	No		Yes		2.2	No	No
Perry *et al.*[226]	Comparative	UK	1988	PTCA	*Eur Heart J*	224	No		Yes			No	No
Simpfendorfer *et al.*[227]	Comparative	USA	1988	PTCA	*Am J Cardiol*	336			Yes	0	2.4	No	No
Thomas *et al.*[228]	Comparative	USA	1988	PTCA	*Am Heart J*	281	No		No		1	No	No
Thompson *et al.*[229]	Comparative	USA	1991	PTCA	*J Am Coll Candiol*	326	No		Yes		2	No	No
Thompson *et al.* prt [229]	Comparative	USA	1991	PTCA	*J Am Coll Candiol*	233	No		Yes		2	No	No
Thompson *et al.* prt [229]	Comparative	USA	1991	PTCA	*J Am Coll Candiol*	193	No		Yes		2	No	No
Anderson and Ward[230]	Case series	UK	1991	PTCA	*Br Heart J*	11				0		Yes	No
Arnold *et al.*[231]	Case series	USA	1994	PTCA	*Am J Cardiol*	5000	Yes			3	4		No
Bell *et al.*[232]	Case series	USA	1995	PTCA	*Circulation*	3027	No			0	5.5		No
Bentivoglio *et al.*[233]	Case series	USA	1991	PTCA	*Catheter Cardiovasc Diagn*	1720	No	Yes		1	2		No
Berger *et al.* [234]	Case series	USA	1986	PTCA	*Am Heart J*	186	Yes			2	1.2	No	No
Buffet *et al.*[235]	Case series	France	1992	PTCA	*Int J Cardiol*	102			Yes	0	1.9	Yes	No
Buffet *et al.*[236]	Case series	France	1994	PTCA	*Am Heart J*	140			Yes	2	6	No	No
Burton *et al.*[237]	Case series	Canada	1990	PTCA	*Cardiovasc Drug Ther*	100				4	1	Yes	No
Ciampricotti *et al.*[238]	Case series	The Netherlands	1992	PTCA	*Catheter Cardiovasc Diagn*	67			Yes	1	2.2	No	No
Cowley *et al.*[239]	Case series	USA	1985	PTCA	*Circulation*	100			No	56	2.2	Yes	No
Cowley *et al.*[240]	Case series	USA	1993	PTCA	*J Am Coll Candiol*	370	No		Yes	1	2.3	Yes	Yes
de Jaegere *et al.*[241]	Case series	The Netherlands	1992	PTCA	*Br Heart J*	166					1.8	No	No
Deligonul *et al.*[242]	Case series	USA	1988	PTCA	*J Am Coll Candiol*	470			Yes	21	2.3	Yes	No
Dorros and Janke[243]	Case series	USA	1985	PTCA	*Herz*	235			Yes			No	No
Dorros *et al.*[244]	Case series	USA	1988	PTCA	*Am J Cardiol*	76				5	2.6	Yes	No
Dorros *et al.* [244]	Case series	USA	1988	PTCA	*Clin Cardiol*	752	Yes		Yes	8	2.6	Yes	No
Ellis *et al.*[245]	Case series	USA	1991	PTCA	*Circulation*	350			Yes	1	1.8	Yes	Yes
Ellis *et al.*[246]	Case series	New Zealand	1998	PTCA	*Am J Cardiol*	86	Yes		Yes	3	4	Yes	No
Ernst *et al.*[247]	Case series	The Netherlands	1987	PTCA	*Br Heart J*	1352	No						No

continued

Study	Study design	Country	Publication date	Intervention	Journal	No. of patients	Prospective	Registry	Consecutive enrolment	Lost to follow-up (%)	Average follow-up (years)	Objective/ blinded outcomes	Multi-centre
el Gaylani *et al.*[248]	Case series	Ireland	1996	PTCA	*Irish Med J*	129	No		No	16	0.4	No	No
Glazier *et al.*[249]	Case series	London	1990	PTCA	*J R Coll Physicians*	100	No				1	Yes	No
Grigg *et al.*[250]	Case series	Australia	1988	PTCA	*Aust N Z J Med*	42					2	No	No
Gurbel *et al.*[251]	Case series	The Netherlands	1997	PTCA	*Catheter Cardiovasc Diagn*	12	Yes		Yes	0	0.7		No
Henderson *et al.*[252]	Case series	UK	1991	PTCA	*Eur Heart J*	295				1	2.9	Yes	No
Holmes *et al.*[253]	Case series	USA	1984	PTCA	*Am J Cardiol*	665	Yes			16	0.5	Yes	Yes
Ilsley *et al.*[254]	Case series	New Zealand	1985	PTCA	*N Z Med J*	50			Yes			No	No
Ivanhoe *et al.*[255]	Case series	USA	1992	PTCA	*Circulation*	480			Yes	11	2.2	No	No
Jost *et al.*[256]	Case series	Germany	1991	PTCA	*Am Heart J*	90				42	3	Yes	No
Kelsey *et al.*[257]	Case series	USA	1993	PTCA	*Circulation*	2136	No	Yes		5	4	Yes	No
King and Schlumpf[258]	Case series	Switzerland	1993	PTCA	*J Am Coll Candiol*	169			Yes	22	10	Yes	No
Kofflard *et al.*[259]	Case series	The Netherlands	1995	PTCA	*Br Heart J*	57	Yes				4.7	Yes	No
Krajcer *et al.*[260]	Case series	USA	1982	PTCA	*Catheter Cardiovasc Diagn*	33				48	0.5	No	No
Leisch *et al.*[261]	Case series	Austria	1986	PTCA	*Br Heart J*	22					2	No	No
Maiello *et al.*[262]	Case series	Italy	1992	PTCA	*Int J Cardiol*	92	No		Yes		1.1	No	No
Mata *et al.*[263]	Case series	Canada	1985	PTCA	*J Am Coll Candiol*	74			Yes	18		Yes	No
Melchior *et al.*[264]	Case series	Switzerland	1987	PTCA	*Am J Cardiol*	100			Yes	51	0.7	Yes	No
Morton *et al.*[265]	Case series	Canada	1989	PTCA	*Can J Cardiol*	145			Yes	1	3.4	No	No
Myler *et al.*[266]	Case series	USA	1987	PTCA	*Catheter Cardiovasc Diagn*	494			Yes	0	1.4		
Piovaccari *et al.*[267]	Case series	Italy	1991	PTCA	*Int J Cardiol*	206			Yes		2.1	No	No
Richardson *et al.*[268]	Case series	Australia	1994	PTCA	*Aust N Z J Med*	2571	Yes				1.7		No
Ruygrok *et al.*[269]	Case series	The Netherlands	1996	PTCA	*J Am Coll Candiol*	856			Yes	2	9.6	Yes	No
Ruygrok *et al.*[270]	Case series	New Zealand	1998	PTCA	*Catheter Cardiovasc Diagn*	126	Yes		Yes	1%	0.1	No	Yes
Safian *et al.*[271]	Case series	USA	1994	PTCA	*Circulation*	146			Yes	8	0.5	No	
Sahni *et al.*[272]	Case series	USA	1989	PTCA	*Clin Cardiol*	124			Yes		1.6	No	
Scott *et al.*[273]	Case series	USA	1994	PTCA	*Am J Cardiol*	2015	No	Yes		11	5		No
Simpfendorfer *et al.*[274]	Case series	USA	1989	PTCA	*Cleve Clin J Med*	33	No		Yes	0	2.5	No	No
Skinner *et al.*[61]	Case series	UK	1999	PTCA	*Heart*	353	Yes		Yes				

continued

Study	Study design	Country	Publication date	Intervention	Journal	No. of patients	Prospective	Registry	Consecutive enrolment	Lost to follow-up (%)	Average follow-up (years)	Objective/ blinded outcomes	Multi-centre
Stammen *et al.*[275]	Case series	Belgium	1991	PTCA	*Am J Cardiol*	507	No		Yes	2	0.5	Yes	No
Stein *et al.*[276]	Case series	USA	1995	PTCA	*Circulation*	10433	No			4	4		No
Suryapranata *et al.*[277]	Case series	The Netherlands	1993	PTCA	*Cor Vasa*	2183		No			1.8	No	No
Talley *et al.*[278]	Case series	USA	1988	PTCA	*Circulation*	427	Yes		Yes	0	5	No	No
Tan *et al.*[279]	Case series	UK	1995	PTCA	*Br Heart J*	163			Yes	0	2.9	Yes	No
Thompson *et al.*[280]	Case series	USA	1993	PTCA	*Circulation*	982	Yes				2.1	No	No
Urban *et al.*[281]	Case series	UK	1987	PTCA	*Br Heart J*	51	No			0	0.3	Yes	No
Valentine and Manolas[282]	Case series	Australia	1984	PTCA	*Med J Aust*	110			Yes			No	No
Vandormael *et al.*[283]	Case series	USA	1991	PTCA	*Am J Cardiol*	637			Yes	5		No	No
Voudris *et al.*[284]	Case series	Greece	1993	PTCA	*Angiology*	37					1.8	Yes	No
Webb *et al.*[285]	Case series	USA	1990	PTCA	*J Am Coll Candiol*	148			Yes	3	3.7	No	No
Weintraub *et al.*[286]	Case series	USA	1993	PTCA	*Circulation*	3363	Yes			3	3.8		No
Weintraub *et al.*[287]	Case series	USA	1994	PTCA	*J Am Coll Cardiol*	10785	Yes			8	3.5		No
Weintraub *et al.*[288]	Case series	USA	1995	PTCA	*J Am Coll Cardiol*	10783	No	Yes			3.5		No
Wilson & Stone[289]	Case series	USA	1994	PTCA	*Am J Cardiol*	161			Yes	1	3.3	No	No
Yamaguchi[290]	Case series	Japan	1990	PTCA	*Jpn J Med*	1174	No			1	2.6		No
ten Berg *et al.*[92]	Case control	The Netherlands	1996	PTCA	*Am J Cardiol*	192			Yes	0	1.9	No	No
ten Berg *et al.*[92]	Case control	The Netherlands	1996	PTCA	*Am J Cardiol*	192			Yes	0	4.8	No	No

CABG – sample characteristics

Study	Single-vessel disease (%)	Median age (years)	Male (%)	Left ventricular dysfunction (%)	Ejection fraction (%)	NYHA angina grade 3 or 4 (%)	Left main artery disease (%)	Proximal LAD stenosis (%)	Diabetes (%)	Unstable angina (%)	Hypertension (%)
BARI[139]	0	61	74		58	16				65	
CABRI[140]	0	60	78		63	49			12	15	
ERACI[141]	0	55	89		62						
Goy *et al.*[142]	100	54	80			78			12		
Hamm *et al.*[143]	0		80						15	15	
Hueb *et al.*[144]	100	58	83		74			88	18		
Jones and Weintraub[145]											
King *et al.*[146]	0		73		62	83			21		52
Puel *et al.*[147]	0										
RITA[148]	44	57	79			61					
Bonnier *et al.*[149]	4	53	65			98			4		49
Mick *et al.*[150]	2	82	61	63		67			18	60	53
Tyras *et al.*[151]	100	51	73	47		83			13		29
Ullyot *et al.*[152]										29	
Acar *et al.*[153]		67							30		52
Acinapura *et al.*[154]		63	67				20		19		
Acinapura *et al.*[155]		66	67				20				
Arnold *et al.*[156]	12		85	59.10		81				58.40	
Ashor *et al.*[157]		68	82						5	7	
Azariades *et al.*[158]	8	75	68	61					18	8	
Azariades *et al.*[158]	8	75	68	61		75	15	14	18	38	53
Baldwin *et al.*[159]	100	59	100		56						
Barner *et al.*[160]		52	86								
Bathgate and Irving[161]	5	54	87	52		78					
Bell *et al.*[162]	2	57	88		59	73					
Beretta *et al.*[163]	0	59	80		58	90	40			20	
Bergsma *et al.*[164]	0								5		
Brandup-Wognsen *et al.*[165]	7	64	81			85	20		12		37
Cameron *et al.*[166]		54	82						12		
Canver *et al.*[167]			100		58						
Carter[168]			70								

continued

Study	Single-vessel disease (%)	Median age (years)	Male (%)	Left ventricular dysfunction (%)	Ejection fraction (%)	NYHA angina grade 3 or 4 (%)	Left main artery disease (%)	Proximal LAD stenosis (%)	Diabetes (%)	Unstable angina (%)	Hypertension (%)
Christakis *et al.*14[169]	7		84			69				72	
Christenson and Schmuziger[170]	0	56	89							3	49
Cohen *et al.*[220]	27				64		10		11		41
Egstrup[171]		53			63						
Farrer *et al.*[60]	22	57	84	59		83	10		6		25
Fitzgibbon *et al.*[172]			99								
French *et al.*[173]		36	84		58	68		12	5		17
Gale *et al.*[174]		53	86	51						15	
Gelbfish *et al.*[175]		66	50								
Gelfand *et al.*[176]	32	36	85	43					11		15
Green *et al.*[177]	15		81	39			11			44	
Higginbothom *et al.*[178]	9	67	77	74			20			40	
Horgan *et al.*[179]	20	51	87								
Horneffer *et al.*[221]		48	78		61				17	27	47
Horneffer *et al.*[221]		61	79		62				18	31	56
Horneffer *et al.*[221]		73	66		63				19	48	59
Ivert *et al.*[180]	19	72	68				14		11	7	26
Jenkins *et al.*[181]		54	84								
Jones and Weintraub[145]	11	57	84		59	53	11		15		
Killen *et al.*[182]	17		85								
Killen *et al.*[183]		55	74			80		100		73	
Killen *et al.*[184]		55	77					100			
Kornfeld *et al.*[185]		52	93								
Laird-Meeter *et al.*[186]	19	53	88				8				
Laks *et al.*[187]	39		91	65		64			21		55
Lawrie *et al.*[91]	19		89	30							
Liao *et al.*[188]	30	57	45		61		7		28	55	
Lytle *et al.*[189]	37	<35	90	30		29	3		6		22
Maddern *et al.*[190]			85								
MacDonald *et al.*[191]	4	79	66		61	93	18		22	67	51
Morin *et al.*[192]		59	87	62					18		36
Morris *et al.*[193]		60			50					42	

continued

Study	Single-vessel disease (%)	Median age (years)	Male (%)	Left ventricular dysfunction (%)	Ejection fraction (%)	NYHA angina grade 3 or 4 (%)	Left main artery disease (%)	Proximal LAD stenosis (%)	Diabetes (%)	Unstable angina (%)	Hypertension (%)
Mullany *et al.*[194]	1	82	67		55	97	41		13	89	
Nicholson and Paterson[195]		55	84			33			20	15	
Ochsner *et al.*[196]	42	51	90								
Palatianos *et al.*[197]	1	60	85		53				29		65
Patel *et al.*[198]	4	57	88				20				
Peterson *et al.*[199]		69	69						15		
Pinna-Pintor *et al.*[200]		61	86								
Rahimtoola *et al.*[201]	13	61	78	53					12	26	42
Rahimtoola *et al.*[202]	15	62	78	50			12		14	31	
Richardson and Cyrus[203]	6	56	76	28			14		16		
Risum *et al.*[204]	8.60		89			91			4.40	19	
Risum *et al.*[205]	8		89		62				4	19	19
Ruygrok *et al.*[206]		72	73		60				6		
Saatvedt *et al.*[207]		56	90			100		100			
Salomon *et al.*[208]		60	80		63	92					
Schaff *et al.*[209]	16	52	91			88	15				
Schmuziger *et al.*[210]	4	61	83		59	58	27		9	31	
Sheldon and Loop[211]											
Simmons *et al.*[212]		58	55	34	57				23		69
Sterling *et al.*[213]	7	49	100		61	91	17		19	9	56
Tector *et al.*[214]	0		93								
Tschan *et al.*[215]	22	54	91			73	11				21
Tyras *et al.*[151]	8.50	53	85	51				13		13	
Ullyot *et al.*[216]		52	99	19	64					19	
Verhiest *et al.*[217]		50	92	25						19	
Weintraub *et al.*[218]	6	61	84		51	76	19		22		
Wright[219]											

PTCA studies – sample characteristics

Study	Single-vessel disease (%)	Median age (years)	Male (%)	Left ventricular dysfunction (%)	Ejection fraction (%)	NYHA angina grade 3 or 4 (%)	Left main artery disease (%)	Proximal LAD stenosis (%)	Diabetes	Unstable angina (%)	Hypertension (%)
BARI[139]	0	62	73		57	14			24	63	
CABRI[140]	0	60	78		63	45			12	14	
ERACI[141]	0	59	81		59						
Goy *et al.*[142]	100	57	80			80	100		12		
Hamm *et al.*[143]	0		79						10	13	
Hueb *et al.*12[144]	100	54	81		77				15	0	
Jones and Weintraub[145]											
King *et al.*[146]	0		75		61	77			25		54
Puel *et al.*[147]	0										
RITA[148]	46	57	83		59	57					
Bonnier *et al.*[149]	4	80	58			100			5		51
Cavallini *et al.*[222]		62	84		55				13	45	
Jeroudi *et al.*[223]	19	82	54			91			20	59	63
Kamp *et al.*[224]	64	55	81		60					40	
Meyer *et al.*[225]	91	49	89			66	2	71		50	
Mick *et al.*[150]	21	82	53	53		92			21	57	58
Perry *et al.*[226]		50	81					58			
Simpfendorer *et al.*[227]	73	74	58	22				59		63	
Thomas *et al.*[228]	67	56	77						8	81	29
Thompson *et al.*[229]	56	67	67		62	58				70	
Thompson *et al.*[229]	55	72	59		61	63				76	
Thompson *et al.*[229]	37	79	56		59	65				80	
Anderson and Ward[230]	64	58	91			18		18			
Arnold *et al.*[231]	84	58	75	11		46			14		
Bell *et al.*[232]	33	62	73		61	67			13	69	
Bentivoglio *et al.*[233]	80		74						13	55	
Berger *et al.*[234]	23	54	77								
Buffet *et al.*[235]	39	78	56		60	86				76	
Buffet *et al.*[236]	75	34	94		64				3	28	16
Burton *et al.*[237]	27	57	85		56				12		30
Ciampricotti *et al.*[238]		55	75					87		48	

continued

Study	Single-vessel disease (%)	Median age (years)	Male (%)	Left ventricular dysfunction (%)	Ejection fraction (%)	NYHA angina grade 3 or 4 (%)	Left main artery disease (%)	Proximal LAD stenosis (%)	Diabetes	Unstable angina (%)	Hypertension (%)
Cowley *et al.*[239]	14	55	71			87		75		66	
Cowley *et al.*[240]	0	58	72		58			34	17		49
de Jaegere *et al.*[241]	52	73	52		59		3	57		49	
Deligonul *et al.*[242]	0		76			61		28		49	
Dorros *et al.*[243]	19	58	79			35			15		39
Dorros *et al.*[244]		58	82			62			22		49
Dorros *et al.* [244]	0	58	79			44			15		43
Ellis *et al.*[245]		58	71		58	72			19		52
Ellis *et al.*[246]	52	37	83		68				7	29	
Ernst *et al.*[247]	70		80		55	72					
el Gaylani *et al.*[248]	62	55	79							48	
Glazier *et al.*[249]	0	58	90			67				16	29
Grigg *et al.*[250]	100	52	86					83			
Gurbel *et al.*[251]		58							0	75	33
Henderson *et al.*[252]	100	53	78			77		100			
Holmes *et al.*[253]	84	51	81			63	1	71			
Ilsley *et al.*[254]	66	55	78					50		26	
Ivanhoe *et al.*[255]	62	55	81		55	63			12		38
Jost *et al.*[256]		55	91			36		50			
Kelsey *et al.*[257]	49	58	77			23			13		
King and Schlumpf[258]	58	50	85								
Kofflard *et al.*[259]	75	33	84			86		63	9	40	42
Krajcer *et al.*[260]		55	88			55		58			27
Leisch *et al.*[261]	100	52	68								
Maiello *et al.*[262]	41	74	67		51					79	
Mata *et al.*[263]	0	52	78		61			95	8	25	23
Melchior *et al.*[264]		54	87		57	61		44			
Morton *et al.*[265]		52	77					66			
Myler *et al.*[266]	0	58	82			54		79			
Piovaccari *et al.*[267]		56	83					11		30	
Richardson *et al.*[268]		58	60			79					
Ruygrok *et al.*[269]	63	56	80							38	

continued

Study	Single-vessel disease (%)	Median age (years)	Male (%)	Left ventricular dysfunction (%)	Ejection fraction (%)	NYHA angina grade 3 or 4 (%)	Left main artery disease (%)	Proximal LAD stenosis (%)	Diabetes	Unstable angina (%)	Hypertension (%)
Ruygrok *et al.*[270]		61	74	33			3	46	11	37	27
Safian *et al.*[271]	1	65	78		47	92				54	
Sahni *et al.*[272]	44	61	66			57					
Scott *et al.*[273]	49		75			53			14		
Simpfendorfer *et al.*[274]	48	32	85	42		73		33	15	52	30
Skinner *et al.*[61]											
Stammen *et al.*[275]	56	59	77				1.0	51	9		40
Stein *et al.*[276]	72	58	74		58	62			11		
Suryapranata *et al.*[277]	66	58	78							41	
Talley *et al.*[278]	86	54	77	8			1	58		61	
Tan *et al.*[279]	29	73	63	40		72			7	17	18
ten Berg *et al.*[92]	51	78	65	16		88					
ten Berg *et al.*[92]	51	55	65	16		88					
Thompson *et al.*[280]		72	62			66					
Urban *et al.*[281]	32	69	74	31		89					
Valentine and Manolas[282]	100	51	88								
Vandormael *et al.*[283]	0	59	74						19	47	44
Voudris *et al.*[284]	59		73		53					54	
Webb *et al.*[285]	51	36	82			59			10	50	35
Weintraub *et al.*[286]	74	56	77			55			12		
Weintraub *et al.*[287]	70	58	74			61			14		
Weintraub *et al.*[288]	71	58	74		58	62					
Wilson and Stone[289]	0	65	66						23	54	
Yamaguchi[290]	54		81			29					

CABG for chronic stable angina – study outcomes

Study	Mortality (%)	Angina recurs (%)	No. of grafts	5-Year survival (from curve) (%)	10-Year survival (from curve) (%)	7-Year survival (from curve) (%)
BARI[139]	12	15				
CABRI[140]	2.70	14				
ERACI[141]	4.70	21				
Goy *et al.*[142]	2	5				
Hamm *et al.*[143]	5.10	26				
Hueb *et al.*[144]	1	2				
Jones and Weintraub[145]						
King *et al.*[146]	6	12				
Puel *et al.*[147]	12					
RITA[148]	3.60	21				
Bonnier *et al.*[149]	17			96	84	
Mick *et al.*[150]	25	7	3.1			
Tyras *et al.*[151]	2	27		97.9		
Ullyot *et al.*[152]	3	27				
Acar *et al.*[153]	8.40	11.30	2.8	92		
Acinapura *et al.*[154]	15.40					
Acinapura *et al.*[155]	23.10	20				
Arnold *et al.*[156]	13	37	2.2	87		
Ashor *et al.*[157]	7	24	2			
Azariades *et al.*[158]	18.4			81.6		
Azariades *et al.*[158]	6	30		82		
Baldwin *et al.*[159]	8	45				
Barner *et al.*[160]	23	37		93	84	
Bathgate and Irving[161]	42	56	2.95	90	65	
Bell *et al.*[162]	13.40					
Beretta *et al.*[163]	0	0	3.8			
Bergsma *et al.*[164]	4.70	14.60		96		91.10
Brandrup-Wognsen *et al.*[165]	5.60					
Cameron *et al.*[166]						
Canver *et al.*[167]	35				59	
Carter[168]	3.30	17	3.5			
Christakis *et al.*[169]	10			90		
Christenson and Schmuziger[170]	18			90	74	
Cohen *et al.*[220]	16					
Egstrup[171]	3	19	3.3			
Farrer *et al.*[60]	8	26	2.7			
Fitzgibbon *et al.*[172]	62		3.8	94	81	
French *et al.*[173]	26		2.3	91	74	
Gale *et al.*[174]	4	54	2.3			
Gelbfish *et al.*[175]	21	39	2.6			
Gelfand *et al.*[176]	3	21	2.3			
Green *et al.*[177]	6	27				
Higginbothom *et al.*[178]	14	29	2.4			

continued

Study	Mortality (%)	Angina recurs (%)	No. of grafts	5-Year survival (from curve) (%)	10-Year survival (from curve) (%)	7-Year survival (from curve) (%)
Horgan *et al.*[179]	13	27				
Horneffer *et al.*[221]	5	28				
Horneffer *et al.*[221]	10	25				
Horneffer *et al.*[221]	14	26				
Ivert *et al.*[180]	16	17	2.6			
Jenkins *et al.*[181]						
Jones and Weintraub[145]	25			90	75	
Killen *et al.*[182]	11	63		90	71	
Killen *et al.*[183]	24	40	1.1	96	86	
Killen *et al.*[184]	28		1.2	95	87	
Kornfeld *et al.*[185]	23	46				
Laird-Meeter *et al.*[186]	14			92	79	
Laks *et al.*[187]	4	33	1.9	92		
Lawrie *et al.*[91]	7	52				
Liao *et al.*[188]	21.40					
Lytle *et al.*[189]	15	35		94	85	
Maddern *et al.*[190]	9	7		91		
MacDonald *et al.*[191]	16					
Morin *et al.*[192]	10	36	4			
Morris *et al.*[193]	9					
Mullany *et al.*[194]	25	21	3.2	71		
Nicholson and Paterson[195]	0	7	1			
Ochsner *et al.*[196]	28	46				72
Palatianos *et al.*[197]	5	11				
Patel *et al.*[198]	9		3.3			
Peterson *et al.*[1995]	15.30					
Pinna-Pintor *et al.*[200]	6					
Rahimtoola *et al.*[201]	62				73	
Rahimtoola *et al.*[202]	59	44		88	72	
Richardson and Cyrus[203]	11					
Risum *et al.*[204]	16			89		
Risum *et al.*[205]	26			89	75	
Ruygrok *et al.*[206]	41	23	2.8	77	50	
Saatvedt *et al.*[207]	0	0				
Salomon *et al.*[208]				89	74	
Schaff *et al.*[209]	23	53		92	71	
Schmuziger *et al.*[210]	1.80					
Sheldon and Loop[211]				89	77	
Simmons *et al.*[212]	8	51	2.8			
Sterling *et al.*[213]	6		2.9			
Tector *et al.*[214]	1	0	1.7			
Tschan *et al.*[215]	7		2	90.2		88.4
Tyras *et al.*[151]	12			88		
Ullyot *et al.*[216]	4.50	33	2.4			
Verhiest *et al.*[217]			2			
Weintraub *et al.*[218]	45	41		76	55	
Wright[219]	11	50				

PTCA for chronic stable angina – outcomes

Study	Mortality (%)	Angina recurs (%)	No. of grafts	5-Year Survival (from curve) (%)	10-Year survival (from curve) (%)	7-Year survival (from curve) (%)
BARI[139]	14	21				
CABRI[140]	4	10				
ERACI[141]	9.50	43				
Goy *et al.*[142]	4					
Hamm *et al.*[143]	2	29				
Hueb *et al.*[144]	1					
Jones and Weintraub[145]						
King *et al.*[146]	7	20				
Puel *et al.*[147]	10					
Bonnier *et al.*[149]	11	4		93	92	
Cavallini *et al.*[222]	0	34				
Jeroudi *et al.*[223]	13	13				
Kamp *et al.*[224]	5					
Meyer *et al.*[225]	1	30				
Mick *et al.*[150]	25	28				
Perry *et al.*[226]	0.50					
Simpfendorfer *et al.*[227]	7	23				
Thomas *et al.*[228]	0.40	27				
Thompson *et al.*[229]				92		
Thompson *et al.*[229]				72		
Thompson *et al.*[229]				72		
Anderson and Ward[220]	0	64				
Arnold *et al.*[231]	7					
Bell *et al.*[232]		36			76	
Bentivoglio *et al.*[233]	5					
Berger *et al.*[234]	2	16				
Buffet *et al.*[235]	14	27				
Buffet *et al.*[236]	3.50	12		96	96	
Burton *et al.*[237]	1	40				
Ciampricotti *et al.*[238]		18				
Cowley *et al.*[239]	0	34				
Cowley *et al.*[240]	5	26				
de Jaegere *et al.*[241]	10	27				
Deligonul *et al.*[242]	7	20				
Dorros *et al.*[243]	2	30				
Dorros *et al.*[244]	11	54				
Dorros *et al.*[244]	7			90		
Ellis *et al.*[245]	4.50					
Ellis *et al.*[246]	4	57		95	91	
Ernst *et al.*[247]	2	26				
Gaylani *et al.*[248]	1	34				
Glazier *et al.*[249]	1					
Grigg *et al.*[250]	0	12				
Gurbel *et al.*[251]	0	8				

continued

Study	Mortality (%)	Angina recurs (%)	No. of grafts	5-Year Survival (from curve) (%)	10-Year survival (from curve) (%)	7-Year survival (from curve) (%)
Henderson *et al*.[252]	6	26		96		
Holmes *et al*.[253]		45				
Ilsley *et al*.[254]	2	32				
Ivanhoe *et al*.[255]	2	33		98		
Jost *et al*.[256]	3	56				
Kelsey *et al*.[257]	8					
King and Schlumpf[258]	8	29			90	
Kofflard *et al*.[259]	12	20		87		
Krajcer *et al*.[260]		42				
Leisch *et al*.[261]		18				
Maiello *et al*.[262]	9	45				
Mata *et al*.[263]	0	64				
Melchior *et al*.[264]	0	66				
Morton *et al*.[265]	3	61				
Myler *et al*.[266]	1	74				
Piovaccari *et al*.[267]	2	16				
Richardson *et al*.[268]	10					
RITA[148]	3	31				
Ruygrok *et al*.[269]	23	47		90	78	
Ruygrok *et al*.[270]	0	12				
Safian *et al*.[271]	7	37				
Sahni *et al*.[272]	5	15				
Scott *et al*.[273]	10	20				
Simpfendorfer *et al*.[274]	6	26				
Skinner *et al*.[61]						
Stammen *et al*.[275]	1	25				
Stein *et al*.[276]				92		
Suryapranata *et al*.[277]	2					
Talley *et al*.[278]	4	15		96		
Tan *et al*.[279]	12	48		83		
ten Berg *et al*.[92]	23	45		77		
ten Berg *et al*.[92]	2	25		98		
Thompson *et al*.[280]	9	27				
Urban *et al*.[281]	8	67				
Valentine and Manolas[282]	1					
Vandormael *et al*.[283]	11			86		
Voudris *et al*.[284]	3	46				
Webb *et al*.[285]	6	21		95		
Weintraub *et al*.[286]	6	54				
Weintraub *et al*.[287]		30		94		
Weintraub *et al*.[288]		30		91	83	
Wilson and Stone[289]	9					
Yamaguchi[290]	2	16				

Spinal cord stimulation – included studies

Study	Study design	Country	Publication date	Journal	No. of patients implanted
Dario[294]	Cohort	Italy	2001	*Neuromodulation*	49
North[295]	RCT	USA	1995	*Acta Neurochir Suppl*	51
Barolat[296]	Case series	USA	1999	*Neuromodulation*	80
Barolat[297]	Case series	USA	2001	*Neuromodulation*	41
Batier[298]	Case series	France	1989	*Agressologie*	14
Bel[299]	Case series	Germany	1991	*Acta Neurochir Suppl*	14
Blond[300]	Case series	France	1991	*Neurochirurgie*	59
Blond[301]	Case series	France	1998	*Douleur et Analgesie*	250
Blume[302]	Case series	USA	1992	*Neurosurgery*	28
Burchiel[303]	Case series	USA	1993	*IASP 7th World Congress on Pain*	42
Burchiel[304]	Case series	USA	1995		40
Clark[305]	Case series	USA	1975	*Surg Neurol*	6
Dam Hieu[306]	Case series	France	1994	*Rev Rhum (Engl Edn)*	77
De La Porte[307]	Case series	Switzerland	1983	*Spine*	38
De La Porte[308]	Case series	Belgium	1993	*Pain*	64
Demirel[309]	Case series	Germany	1984	*Neurochirurgie*	7
Devulder[310]	Case series	Belgium	1990	*Clin J Pain*	23
Devulder[311]	Case series	Belgium	1991	*Clin J Pain*	43
Devulder[312]	Case series	Belgium	1997	*J Pain Symptom Manage*	69
Fassio[313]	Case series	France	1988?	*Rev Chir Orthop*	20
Gonzalez-Darder[314]	Case series	Spain	1992	*Rev Esp Anestesiol Reanim*	13
Hassenbusch[315]	Case series	USA	1995	*Acta Neurochir*	9
Heidecke[316]	Case series	Germany	2000	*Neuromodulation*	42
Hoppenstein[317]	Case series	USA	1975	*Surg Neurol*	13
Hunt[318]	Case series	USA	1975	*Surg Neurol*	5
Kalin[319]	Case series		1984	*Pain*	77
Kavar[320]	Case series	Australia	2000	*J Clin Neurosc*	19
Kay[321]	Case series	UK	2001	*Br J Neurosurg*	36
Kim[322]	Case series	Japan and France	1994	*Jpn J Neurosurg*	58
Kim[323]	Case series	Korea and USA	2001	*Neurosurg*	19
Kumar[324]	Case series	Canada	1986	*Pain Clinic*	38
Kumar[325]	Case series	Canada	1991	*J Neurosurg*	57
Kumar[326]	Case series	USA and Canada	1996	*Surg Neurol*	101
Kumar[327]	Case series		1998	*Curr Rev Pain*	
Kumpulainen[328]	Case series	Finland	1986	*Ann Clin Res*	4
Law[329]	Case series	USA	1991	*Pain Manage*	115
Law[330]	Case series	USA	1992	*Stereotact Funct Neurosurg*	117
Lazorthes[331]	Case series	France	1995	*Neurochirurgie*	304
Leclercq[332]	Case series		1981	*Neurochirurgie*	
Leclercq[333]	Case series	USA	1982	*R I Med J*	20
LeDoux[334]	Case series	USA	1993	*Spine*	26
Leibock[335]	Case series	USA	1984	*Nebr Med J*	11

continued

Study	Study design	Country	Publication date	Journal	No. of patients implanted
LeRoy[336]	Case series	USA	1981	*App Neuropsychol*	49
Leveque[337]	Case series	USA	2001	*Neuromodulation*	16
Long[338]	Case series	USA	1975	*Surg Neurol*	55
Long[339]	Case series	USA	1981	*Indications for Spinal Cord Stimulation*	24
Meglio[340]	Case series	Italy	1994	*Stereotact Funct Neurosurg*	21
Meilman[341]	Case series	USA	1989	*Clin J Pain*	20
Mittal[342]	Case series	UK	1987	*Ann R Coll Surg Engl*	22
Mundinger[343]	Case series	Germany	1982	*Appl Neuropsychol*	1
Neilson[344]	Case series	USA	1975	*Surg Neurol*	81
North[345]	Case series	USA	1977	*Appl Neuropsychol*	24
North[346]	Case series	USA	1991	*Pain*	50
North[347]	Case series	USA	1991	*Neurosurgery*	102
North[348]	Case series	USA	1984	*Pain*	24
Ohnmeiss[349]	Case series	USA	1996	*Spine*	40
Ohnmeiss[350]	Case series	USA	2001	*Spine J*	36
Pineda[351]	Case series	USA	1975	*Surg Neurol*	56
Probst[352]	Case series	Switzerland	1990	*Acta Neurochir (Wien)*	112
Rainov[353]	Case series	Germany	1996	*Minim Invasive Neurosurg*	29
Ray[354]	Case series	USA	1975	*Adv Neurosurg*	95
Ray[355]	Case series	USA	1982	*Appl Neuropsychol*	50
Richardson[356]	Case series	USA	1979	*Neurosurgery*	9
Robb[357]	Case series		1990	*Pain*	13
Seijo[358]	Case series	Spain	1993	*Pain Clin*	34
Shatin[359]	Case series	USA	1986	*Pace*	86
Shatin[360]	Case series	USA	1990	*Pain*	77
Sheldon[361]	Case series	USA	1975	*Surg Neurol*	3
Siegfried[362]	Case series	France/ Switzerland	1982	*Appl Neuropsychol*	89
Simpson[363]	Case series	UK	1991	*J Neurol Neurosurg Psychiatry*	7
Spiegelmann[364]	Case series	USA	1991	*Neurosurgery*	12
Van Buyten[365]	Case series	Belgium	1999	*Neuromodulation*	20
Van de Kelft[366]	Case series	Belgium	1994	*Qual Life Res*	64
Vogel[367]	Case series	Germany	1986	*J Neurol*	29
Waisbrod[368]	Case series	Germany	1985	*Arch Orthop Trauma Surg*	16
Wester[369]	Case series	Norway	1987	*Acta Neurol Scand*	10
Winkelmuller[370]	Case series	Germany	1981	*Indications for Spinal Cord Stimulation*	56

Spinal cord stimulation: study design

Study	Consecutive enrolment	Absence of co-interventions	Prospective	Objective/ blinded outcomes	Validated outcomes	Loss to follow-up (%)	Quality score	Average follow-up (months)
Dario[294]					Yes	0		42
North[295]			Yes	No	Yes	0		6
Barolat[296]			Yes		Yes	0	3	24
Barolat[297]			Yes		Yes		2	12
Batier[298]						0	1	12.7
Bel[299]		No	No			0	2	24
Blond[300]	Yes	No	No		Yes	2	3	37
Blond[301]			No		Yes		1	75
Blume[302]			No				0	
Burchiel[303]					Yes		1	6
Burchiel[304]	Yes	No	Yes		Yes	0	5	3
Clark[305]		No	No		No	0	2	24–60
Dam Hieu[306]		No			Yes	0	2	42
De La Porte[307]		No	Yes		Yes	0	2	3
De La Porte[308]	Yes	No	No	No	Yes		2	48
Demirel[309]	Yes	No	No		Yes	0	4	24
Devulder[310]		No	No		Yes	0	2	29
Devulder[311]		No	No		Yes	0	2	2–96
Devulder[312]		No	No		Yes	0	2	
Fassio[313]		No			Yes		1	24
Gonzalez-Darder[314]						0	1	5–27
Hassenbusch[315]		No	No		Yes	0	2	18
Heidecke[316]			No			0	1	46
Hoppenstein[317]		No	No			0	2	5–24
Hunt[318]		No	No			1	1	9–51
Kalin[319]		No	No				0	
Kavar[320]	Yes	No	Yes		Yes	0	5	18.5
Kay[321]		No	No	No		32	0	65
Kim[322]			Yes		Yes	0	2	Up to 48
Kim[323]			Yes			17	1	47
Kumar[324]		No	No			15	1	6–60
Kumar[325]			Yes		Yes	0	3	40
Kumar[326]				Yes	Yes	0	3	66
Kumar[327]				Yes	Yes	6	3	
Kumpulainen[328]						0	1	10
Law[329]		No	No		Yes		1	40
Law[330]	Yes	No	No			0	3	>30
Lazorthes[331]			No			0	1	120
Leclercq[332]								
Leclercq[333]	Yes	No	No			0	3	1–24
LeDoux[334]		No	No		Yes	8	1	12
Leibock[335]		No	No			0	1	
LeRoy[336]			No		Yes		1	30.7

continued

Study	Consecutive enrolment	Absence of co-interventions	Prospective	Objective/ blinded outcomes	Validated outcomes	Loss to follow-up (%)	Quality score	Average follow-up (months)
Leveque[337]		No	Yes	Yes	Yes	0	3	34
Long[338]			No			0	0	24
Long[339]		No			Yes		0	
Meglio[340]		No			Yes		1	45.5
Meilman[341]		No	Yes		Yes	0	3	1
Mittal[342]		No				0	1	Up to 96
Mundinger[343]			No				0	
Nielson[344]			No			5	1	0–35
North[345]		No				11	1	0–13
North[346]	Yes	No		Yes	Yes	8	4	26
North[347]	Yes	No		Yes	Yes	8	4	69.6
North[348]	Yes			Yes	Yes	31	3	96
Ohnmeiss[349]	Yes		Yes	Yes	Yes	5	6	24
Ohnmeiss[350]	Yes		No	Yes		0	4	5.5–19
Pineda[351]						0	1	
Probst[352]		No				0	0	54
Rainov[353]					Yes	0	2	24–42
Ray[354]			No			56	0	18
Ray[355]						32	0	19.4
Richardson[356]		No	No		Yes		1	12
Robb[357]			No		Yes	18	1	6
Seijo[358]	Yes	No	Yes	Yes	Yes	12	6	36
Shatin[359]		No	No			45	0	14.5
Shatin[360]		No	No		Yes	57	1	6
Sheldon[361]		No	No			0	0	
Siegfried[362]					Yes		1	48–96
Simpson[363]	Yes	No	No		Yes	3	4	29
Spiegelmann[364]			No		Yes		1	13
Van Buyten[365]	Yes	No	No		Yes	15	4	28
Van de Kelft[366]	Yes	No	No		Yes	0	4	47
Vogel[367]								Up to 48
Waisbrod[368]								16.2
Wester[369]			Yes		Yes	30	1	15
Winkelmuller[370]		No			Yes		0	4–84§

Spinal cord stimulation – sample details and outcomes

Study	Multi-centre	Median age (years)	Male (%)	Mean pain duration (years)	Pain relief >50% (%)	VAS: difference in means
Dario[294]	No	54	53			0
North[295]	No					
Barolat[296]	No					
Barolat[297]		48.8	75	7.0	66.7	
Batier[298]		53			64.3	
Bel[299]			57		35.7	
Blond[300]		44.7	57		89.7	
Blond[301]		44	65		76.0	
Blume[302]						
Burchiel[303]					59.5	
Burchiel[304]		51.6	50	5.6	44.1	2
Clark[305]		41–48	80		33.3	
Dam Hieu[306]		47	64		63.6	
De La Porte[307]		47.2	52	11.2	94.7	
De La Porte[308]		54	67	6.5	54.7	
Demirel[309]						
Devulder[310]					73.9	
Devulder[311]			51		58.5	
Devulder[312]		43	51			
Fassio[313]		35–50	50		75.0	
Gonzalez-Darder[314]			67	1.5		
Hassenbusch[315]		56	22			
Heidecke[316]		52	66	2		
Hoppenstein[317]				3.8	76.9	
Hunt[318]		46.4		5.6	60.0	
Kalin[319]		47	61			
Kavar[320]		52.5	41	9.9	35.3	
Kay[321]		47	59			
Kim[322]		45	44	5.5	89.6	
Kim[323]		48	49	6.5	83.3	
Kumar[324]		23–74	73		52.6	
Kumar[325]		49	65		64.9	
Kumar[326]		51.4	64		51.5	
Kumar[327]					58.8	
Kumpulainen[328]		45.2	75		75.0	
Law[329]					70.2	
Law[330]		45–52		7	27.4	
Lazorthes[331]						
Leclercq[332]					55.0	
Leclercq[333]						
LeDoux[334]					76.9	
Leibock[335]		49	71	5.5		
LeRoy[336]		20–59	45		49.0	
Leveque[337]		50	23	6.4	50.0	

continued

Study	Multi-centre	Median age (years)	Male (%)	Mean pain duration (years)	Pain relief >50% (%)	VAS: difference in means
Long[338]					83.6	
Long[339]					75.0	
Meglio[340]		51.5	42	6.3	61.9	
Meilman[341]		41.4–50.5	65	7.9	55.0	
Mittal[342]		48	59	9.7	45.5	
Mundinger[343]					100.0	
Nielson[344]		19–72	55	8.4	45.6	
North[345]		48	42		79.2	
North[346]		43.8	55	10.6	50.0	
North[347]		49.1	44	7.8	32.4	
North[348]		48	42		33.3	
Ohnmeiss[349]		48	42	5.5		
Ohnmeiss[350]		47.9	51	6.9		
Pineda[351]		28–76	46		42.9	
Probst[352]		48	57	1–33	63.4	
Rainov[353]		46.5	41	1.4	12.8	
Ray[354]					48.5	
Ray[355]		74	49		54.0	
Richardson[356]					85.7	
Robb[357]		17–87			46.2	
Seijo[358]		30–64	39			
Shatin[359]		50	55	9.1	73.8	
Shatin[360]					69.8	
Sheldon[361]					66.7	
Siegfried[362]		28–76	58		65.2	
Simpson[363]		55	56	7	57.1	
Spiegelmann[364]		52.9	58	5.7	75.0	
Van Buyten[365]		47	30			0.86
Van de Kelft[366]		29–73	54	6.5	56.3	
Vogel[367]		50			13.8	
Waisbrod[368]		43	75		75.0	
Wester[369]						
Winkelmuller[370]					75.5	

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

Professor Shah Ebrahim, Professor in Epidemiology of Ageing, University of Bristol

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 Shah Ebrahim,
Professor in Epidemiology of Ageing, Department of Social Medicine, University of Bristol

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 Ann Bowling, Professor of Health Services Research, Primary Care and Population Studies, University College London

Professor Andrew Bradbury, Professor of Vascular Surgery, Department of Vascular Surgery, Birmingham Heartlands Hospital

Professor John Brazier, Director of Health Economics, Sheffield Health Economics Group, School of Health & Related Research, University of Sheffield

Dr Andrew Briggs, Public Health Career Scientist, Health Economics Research Centre, University of Oxford

Professor Nicky Cullum, Director of Centre for Evidence Based Nursing, Department of Health Sciences, University of York

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, Research Professor in Physiotherapy/Co-Director, Interdisciplinary Research Centre in Health, Coventry University

Professor Julian Little, Professor of Epidemiology, Department of Medicine and Therapeutics, University of Aberdeen

Professor Stuart Logan, Director of Health & Social Care Research, The Peninsula Medical School, Universities of Exeter & Plymouth

Professor Tim Peters, Professor of Primary Care Health Services Research, Division of Primary Health Care, University of Bristol

Professor Ian Roberts, Professor of Epidemiology & Public Health, Intervention Research Unit, London School of Hygiene and Tropical Medicine

Professor Peter Sandercock, Professor of Medical Neurology, Department of Clinical Neurosciences, University of Edinburgh

Professor Mark Sculpher, Professor of Health Economics, Centre for Health Economics, Institute for Research in the Social Services, University of York

Professor Martin Severs, Professor in Elderly Health Care, Portsmouth Institute of Medicine

Dr Jonathan Shapiro, Senior Fellow, Health Services Management Centre, Birmingham

Ms Kate Thomas, Deputy Director, Medical Care Research Unit, University of Sheffield

Professor Simon G Thompson, Director, MRC Biostatistics Unit, Institute of Public Health, Cambridge

Ms Sue Ziebland, Senior Research Fellow, Cancer Research UK, University of Oxford

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, Freelance Consumer Advocate, Bolton

Professor Max Bachmann Professor 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/Laboratory Director, Public Health Laboratory, St Mary's Hospital, Portsmouth

Professor Adrian K Dixon, Professor of Radiology, Addenbrooke's Hospital, Cambridge

Dr David Elliman, Consultant in Community Child Health, London

Professor Glyn Elwyn, Primary Medical Care Research Group, Swansea Clinical School, University of Wales Swansea

Dr John Fielding, Consultant Radiologist, Radiology Department, Royal Shrewsbury Hospital

Dr Karen N Foster, Clinical Lecturer, Dept of General Practice & Primary Care, University of Aberdeen

Professor Antony J Franks, Deputy Medical Director, The Leeds Teaching Hospitals NHS Trust

Mr Tam Fry, Honorary Chairman, Child Growth Foundation, London

Dr Edmund Jessop, Medical Adviser, National Specialist Commissioning Advisory Group (NSCAG), Department of Health, London

Dr Jennifer J Kurinczuk, Consultant Clinical Epidemiologist, National Perinatal Epidemiology Unit, Oxford

Dr Susanne M Ludgate, Medical Director, Medical Devices Agency, London

Dr William Rosenberg, Senior Lecturer and Consultant in Medicine, University of Southampton

Dr Susan Schonfield, CPHM Specialised Services Commissioning, Croydon Primary Care Trust

Dr Margaret Somerville, Director of Public Health, Teignbridge Primary Care Trust

Professor Lindsay Wilson Turnbull, Scientific Director, Centre for MR Investigations & YCR Professor of Radiology, University of Hull

Professor Martin J Whittle, Head of Division of Reproductive & Child Health, 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, Clinical Director, Acute General Medicine SDU, Oxford Radcliffe Hospital

Professor Tony Avery, Professor of Primary Health Care, University of Nottingham

Professor Stirling Bryan, Professor of Health Economics, Health Services Management Centre, University of Birmingham

Mr Peter Cardy, Chief Executive, Macmillan Cancer Relief, London

Dr Christopher Cates, GP and Cochrane Editor, Bushey Health Centre

Professor Imti Choonara, Professor in Child Health, University of Nottingham, Derbyshire Children's Hospital

Mr Charles Dobson, Special Projects Adviser, Department of Health

Dr Robin Ferner, Consultant Physician and Director, West Midlands Centre for Adverse Drug Reactions, City Hospital NHS Trust, Birmingham

Dr Karen A Fitzgerald, Pharmaceutical Adviser, Bro Taf Health Authority, Cardiff

Mrs Sharon Hart, Managing Editor, *Drug & Therapeutics Bulletin*, London

Dr Christine Hine, Consultant in Public Health Medicine, Bristol South & West Primary Care Trust

Professor Stan Kaye, Professor of Medical Oncology, Consultant in Medical Oncology/Drug Development, The Royal Marsden Hospital

Ms Barbara Meredith, Project Manager Clinical Guidelines, Patient Involvement Unit, NICE

Dr Frances Rotblat, CPMP Delegate, Medicines Control Agency, London

Professor Jan Scott, Professor of Psychological Treatments, Institute of Psychiatry, University of London

Mrs Katrina Simister, New Products Manager, National Prescribing Centre, Liverpool

Dr Richard Tiner, Medical Director, Association of the British Pharmaceutical Industry

Dr Helen Williams, Consultant Microbiologist, Norfolk & Norwich University Hospital NHS Trust

Therapeutic Procedures Panel

Members

Chair,
Professor Bruce Campbell,
Consultant Vascular and General Surgeon, Royal Devon & Exeter Hospital

Dr Mahmood Adil, Head of Clinical Support & Health Protection, Directorate of Health and Social Care (North), Department of Health, Manchester

Dr Aileen Clarke, Reader in Health Services Research, Public Health & Policy Research Unit, Barts & the London School of Medicine & Dentistry, Institute of Community Health Sciences, Queen Mary, University of London

Mr Matthew William Cooke, Senior Clinical Lecturer and Honorary Consultant, Emergency Department, University of Warwick, Coventry & Warwickshire NHS Trust, Division of Health in the Community, Centre for Primary Health Care Studies, Coventry

Dr Carl E Counsell, Senior Lecturer in Neurology, University of Aberdeen

Dr Keith Dodd, Consultant Paediatrician, Derbyshire Children's Hospital

Professor Gene Feder, Professor of Primary Care R&D, Barts & the London, Queen Mary's School of Medicine and Dentistry, University of London

Professor Paul Gregg, Professor of Orthopaedic Surgical Science, Department of Orthopaedic Surgery, South Tees Hospital NHS Trust

Ms Bec Hanley, Freelance Consumer Advocate, Hurstpierpoint

Ms Maryann L. Hardy, Lecturer, Division of Radiography, University of Bradford

Professor Alan Horwich, Director of Clinical R&D, The Institute of Cancer Research, London

Dr Phillip Leech, Principal Medical Officer for Primary Care, Department of Health, London

Dr Simon de Lusignan, Senior Lecturer, Primary Care Informatics, Department of Community Health Sciences, St George's Hospital Medical School, London

Dr Mike McGovern, Senior Medical Officer, Heart Team, Department of Health, London

Professor James Neilson, Professor of Obstetrics and Gynaecology, Dept of Obstetrics and Gynaecology, University of Liverpool, Liverpool Women's Hospital

Dr John C Pounsford, Consultant Physician, North Bristol NHS Trust

Dr Vimal Sharma, Consultant Psychiatrist & Hon Snr Lecturer, Mental Health Resource Centre, Victoria Central Hospital, Wirrall

Dr L David Smith, Consultant Cardiologist, Royal Devon & Exeter Hospital

Professor Norman Waugh, Professor 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

Mr John A Cairns,
Professor of Health Economics, Health Economics Research Unit, University of Aberdeen

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 Mary 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 Stephen Cuckle,
Professor of Reproductive Epidemiology, Department of Paediatrics, Obstetrics & Gynaecology, University of Leeds

Professor Nicky Cullum,
Director of Centre for Evidence Based Nursing, University of York

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 Servs, 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 F D Richard Hobbs,
Professor of Primary Care & General Practice, Department of Primary Care & General Practice, University of Birmingham

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 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 Andrew Mortimore,
Consultant in Public Health Medicine, Southampton City Primary Care Trust

Dr Sue Moss,
Associate Director, Cancer Screening Evaluation Unit, Institute of Cancer Research, Sutton

Professor Jon Nicholl,
Director of Medical Care Research Unit, School of Health and Related Research, University of Sheffield

Mrs Julietta Patnick,
National Co-ordinator, NHS Cancer Screening Programmes, Sheffield

Professor Robert Peveler,
Professor of Liaison Psychiatry, University Mental Health Group, Royal South Hants Hospital, Southampton

Professor Chris Price,
Visiting Chair – Oxford, Clinical Research, Bayer Diagnostics Europe, Cirencester

Ms Marianne Rigge,
Director, College of Health, London

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,
Director HSRU/Honorary Consultant in PH Medicine, Department of Public Health, University of Oxford

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