Health Technology Assessment 2004; Vol. 8: No. 6

Effectiveness and efficiency of guideline dissemination and implementation strategies

JM Grimshaw, RE Thomas, G MacLennan, C Fraser, CR Ramsay, L Vale, P Whitty, MP Eccles, L Matowe, L Shirran, M Wensing, R Dijkstra and C Donaldson

February 2004

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.

Effectiveness and efficiency of guideline dissemination and implementation strategies

JM Grimshaw,[1*] RE Thomas,[1] G MacLennan,[1] C Fraser,[1] CR Ramsay,[1] L Vale,[1,2] P Whitty,[3] MP Eccles,[4] L Matowe,[1†] L Shirran,[1] M Wensing,[5] R Dijkstra[5] and C Donaldson[6‡]

[1] Health Services Research Unit, University of Aberdeen, UK
[2] Health Economics Research Unit, University of Aberdeen, UK
[3] Department of Epidemiology and Public Health, University of Newcastle upon Tyne, UK
[4] Centre for Health Services Research, University of Newcastle upon Tyne, UK
[5] Centre for Quality of Care Research, University of Nijmegen, The Netherlands
[6] Department of Community Health Sciences, University of Calgary, Canada

* Corresponding author. Current affiliation: Clinical Epidemiology Programme, Ottawa Health Research Institute and Center for Best Practices, Institute of Population Health, University of Ottawa, Canada

† Current affiliation: Department of Pharmacy Practice, Faculty of Pharmacy, Kuwait University, Kuwait

‡ Current affiliation: Centre for Health Services Research, University of Newcastle upon Tyne, UK

Declared competing interests of authors: none

Published February 2004

This report should be referenced as follows:

Grimshaw JM, Thomas RE, MacLennan G, Fraser C, Ramsay CR, Vale L, *et al*. Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technol Assess* 2004;**8**(6).

Health Technology Assessment is indexed in *Index Medicus*/MEDLINE and *Excerpta Medica*/EMBASE.

NHS R&D HTA Programme

The NHS R&D Health Technology Assessment (HTA) Programme was set up in 1993 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.

Initially, six HTA panels (pharmaceuticals, acute sector, primary and community care, diagnostics and imaging, population screening, methodology) helped to set the research priorities for the HTA Programme. However, during the past few years there have been a number of changes in and around NHS R&D, such as the establishment of the National Institute for Clinical Excellence (NICE) and the creation of three new research programmes: Service Delivery and Organisation (SDO); New and Emerging Applications of Technology (NEAT); and the Methodology Programme.

The research reported in this monograph was identified as a priority by the HTA Programme's Methodology Panel and was funded as project number 94/08/29.

The views expressed in this publication are those of the authors and not necessarily those of the Methodology Programme, HTA Programme or the Department of Health. The editors wish to emphasise that funding and publication of this research by the NHS should not be taken as implicit support for any recommendations made by the authors.

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 editors and publisher have tried to ensure the accuracy of this report but do not accept liability for damages or losses arising from material published in this report. They would like to thank the referees for their constructive comments on the draft document.

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.

M

Abstract

Effectiveness and efficiency of guideline dissemination and implementation strategies

JM Grimshaw,[1*] RE Thomas,[1] G MacLennan,[1] C Fraser,[1] CR Ramsay,[1] L Vale,[1,2] P Whitty,[3] MP Eccles,[4] L Matowe,[1†] L Shirran,[1] M Wensing,[5] R Dijkstra[5] and C Donaldson[6‡]

[1] Health Services Research Unit, University of Aberdeen, UK
[2] Health Economics Research Unit, University of Aberdeen, UK
[3] Department of Epidemiology and Public Health, University of Newcastle upon Tyne, UK
[4] Centre for Health Services Research, University of Newcastle upon Tyne, UK
[5] Centre for Quality of Care Research, University of Nijmegen, The Netherlands
[6] Department of Community Health Sciences, University of Calgary, Canada
[*] Corresponding author. Current affiliation: Clinical Epidemiology Programme, Ottawa Health Research Institute and Center for Best Practices, Institute of Population Health, University of Ottawa, Canada
† Current affiliation: Department of Pharmacy Practice, Faculty of Pharmacy, Kuwait University, Kuwait
‡ Current affiliation: Centre for Health Services Research, University of Newcastle upon Tyne, UK

Objectives: To undertake a systematic review of the effectiveness and costs of different guideline development, dissemination and implementation strategies. To estimate the resource implications of these strategies. To develop a framework for deciding when it is efficient to develop and introduce clinical guidelines.

Data sources: MEDLINE, Healthstar, Cochrane Controlled Trial Register, EMBASE, SIGLE and the specialised register of the Cochrane Effective Practice and Organisation of Care (EPOC) group.

Review methods: Single estimates of dichotomous process variables were derived for each study comparison based upon the primary end-point or the median measure across several reported end-points. Separate analyses were undertaken for comparisons of different types of intervention. The study also explored whether the effects of multifaceted interventions increased with the number of intervention components. Studies reporting economic data were also critically appraised. A survey to estimate the feasibility and likely resource requirements of guideline dissemination and implementation strategies in UK settings was carried out with key informants from primary and secondary care.

Results: In total, 235 studies reporting 309 comparisons met the inclusion criteria; of these 73% of comparisons evaluated multifaceted interventions, although the maximum number of replications of a specific multifaceted intervention was 11 comparisons. Overall, the majority of comparisons reporting dichotomous process data observed improvements in care; however, there was considerable variation in the observed effects both within and across interventions. Commonly evaluated single interventions were reminders, dissemination of educational materials, and audit and feedback. There were 23 comparisons of multifaceted interventions involving educational outreach. The majority of interventions observed modest to moderate improvements in care. No relationship was found between the number of component interventions and the effects of multifaceted interventions. Only 29.4% of comparisons reported any economic data. The majority of studies only reported costs of treatment; only 25 studies reported data on the costs of guideline development or guideline dissemination and implementation. The majority of studies used process measures for their primary end-point, despite the fact that only three guidelines were explicitly evidence based (and may not have been efficient). Respondents to the key informant survey rarely identified existing budgets to support guideline dissemination and implementation strategies. In

general, the respondents thought that only dissemination of educational materials and short (lunchtime) educational meetings were generally feasible within current resources.

Conclusions: There is an imperfect evidence base to support decisions about which guideline dissemination and implementation strategies are likely to be efficient under different circumstances. Decision makers need to use considerable judgement about how best to use the limited resources they have for clinical governance and related activities to maximise population benefits. They need to consider the potential clinical areas for clinical effectiveness activities, the likely benefits and costs required to introduce guidelines and the likely benefits and costs as a result of any changes in provider behaviour. Further research is required to: develop and validate a coherent theoretical framework of health professional and organisational behaviour and behaviour change to inform better the choice of interventions in research and service settings, and to estimate the efficiency of dissemination and implementation strategies in the presence of different barriers and effect modifiers.

Contents

List of abbreviations

A&F	audit and feedback
ANOVA	analysis of variance
ARIMA	autoregressive integrated moving average
CBA	controlled before and after (study)
CCT	controlled clinical trial
C-CCT	cluster allocated controlled clinical trial
CF	C Fraser
C-RCT	cluster randomised controlled trial
D	done
DGH	district general hospital
Edmat	educational materials
Edmeet	educational meetings
EPOC	Cochrane Effective Practice and Organisation of Care (group)
Fin	financial
GM	G MacLennan
IQR	interquartile range
IT	information technology
ITS	interrupted time series (study)
JG	JM Grimshaw
LCP	local consensus processes
LM	L Matowe
LS	L Shirran
ME	MP Eccles
MRC	Medical Research Council
MW	M Wensing
NA	not applicable
NC	not clear
ND	not done
NICE	National Institute for Clinical Effectiveness
NS	not significant
OL	opinion leaders
Org	organisational
Outreach	educational outreach
Patmed	patient mediated
P-CCT	patient allocated controlled clinical trial
PCG	Primary Care Group
P-RCT	patient randomised controlled trial
Profoth	other professional
PW	P Whitty
Reg	regulatory
Rem	reminders
RCT	randomised controlled trial
R&D	research and development
RD	R Dijkstra
RT	RE Thomas
S	significant
SIGN	Scottish Intercollegiate Guidelines Network
SMD	standardised mean difference
Struc	structural
U of AE	unit of analysis error
WHO	World Health Organization

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

Executive summary

Background

Clinical practice guidelines are an increasingly common element of clinical care throughout the world. Such guidelines have the potential to improve the care received by patients by promoting interventions of proven benefit and discouraging ineffective interventions. However, the development and introduction of guidelines are not without costs. In some circumstances, the costs of development and introduction are likely to outweigh their potential benefits. In other circumstances, it may be more efficient to adopt less costly but less effective dissemination and implementation strategies. Local healthcare organisations have relatively few resources for clinical effectiveness activities and policy makers need to consider how best to use these to maximise benefits.

Objectives

The aims of the study were:

- to undertake a systematic review of the effectiveness and costs of different guideline development, dissemination and implementation strategies
- to estimate the resource implications of different development, dissemination and implementation strategies
- to develop a framework for deciding when it is efficient to develop and introduce clinical guidelines based upon the potential costs and benefits of the targeted clinical activity and the effectiveness and costs of guideline development and introduction.

Methods

Systematic review of the effectiveness and efficiency of guideline dissemination and implementation strategies

Data sources

MEDLINE (1966–1998), Healthstar (1975–1998), Cochrane Controlled Trial Register (4th edition 1998), EMBASE (1980–1998), SIGLE (1980–1988) and the specialised register of the Cochrane Effective Practice and Organisation of Care (EPOC) group were searched using a gold standard search strategy developed from handsearches of key journals. The search strategy was 93% sensitive and 18% specific.

Study selection (inclusion criteria)

- *Types of study design*: randomised controlled trials, controlled clinical trials, controlled before and after studies and interrupted time series
- *types of participant*: medically qualified healthcare professionals
- *types of intervention*: guideline dissemination and implementation strategies
- *types of outcome*: objective measures of provider behaviour and/or patient outcome.

Data extraction (and assessment of validity)

Two reviewers independently abstracted data on the methodological quality of the studies (using the Cochrane EPOC group's methodological quality criteria), characteristics of study setting, participants, targeted behaviours and characteristics of interventions. Studies reporting economic evaluations and cost analyses were further assessed against the *British Medical Journal* guidelines for reviewers of economic evaluations.

Data synthesis

Single estimates of dichotomous process variables (e.g. proportion of patients receiving appropriate treatment) were derived for each study comparison based upon the primary end-point (as defined by the authors of the study) or the median measure across several reported end-points. An attempt was made to reanalyse studies with common methodological weaknesses. Separate analyses were undertaken for comparisons of single interventions against 'no-intervention' controls, single interventions against controls receiving interventions, multifaceted interventions against 'no-intervention' controls and multifaceted interventions against controls receiving interventions. The study also explored whether the effects of multifaceted interventions increased with the number of intervention components. For each intervention, the number of comparisons showing a positive direction of effect, the median

effect size across all comparisons, the median effect size across comparisons without unit of analysis errors, and the number of comparisons showing statistically significant effects were reported. A planned meta-regression analysis could not be undertaken owing to the large number of different combinations of multifaceted interventions.

Survey of estimating the feasibility and likely resource requirements of guideline dissemination and implementation strategies in UK settings

Telephone interviews were conducted with key informants from primary and secondary care.

Results (research findings)

Systematic review of the effectiveness and efficiency of guideline dissemination and implementation strategies

In total, 235 studies reporting 309 comparisons met the inclusion criteria. The overall quality of the studies was poor. Seventy-three per cent of comparisons evaluated multifaceted interventions, although the maximum number of replications of a specific multifaceted intervention was 11 comparisons. Overall, the majority of comparisons reporting dichotomous process data (86.6%) observed improvements in care; however, there was considerable variation in the observed effects both within and across interventions. Commonly evaluated single interventions were reminders (38 comparisons), dissemination of educational materials (18 comparisons) and audit and feedback (12 comparisons). There were 23 comparisons of multifaceted interventions involving educational outreach. The majority of interventions observed modest to moderate improvements in care. For example, the median absolute improvement in performance across interventions was 14.1% in 14 cluster randomised comparisons of reminders, 8.1% in four cluster randomised comparisons of dissemination of educational materials, 7.0% in five cluster randomised comparisons of audit and feedback and 6.0% in 13 cluster randomised comparisons of multifaceted interventions involving educational outreach. No relationship was found between the number of component interventions and the effects of multifaceted interventions.

Only 29.4% of comparisons reported any economic data. Eleven reported cost-effectiveness analyses, 38 reported cost consequence analyses (where differences in cost were set against differences in several measures of effectiveness) and 14 reported cost analyses (where some aspect of cost was reported but not related to benefits). The majority of studies only reported costs of treatment; only 25 studies reported data on the costs of guideline development or guideline dissemination and implementation. The majority of studies used process measures for their primary end-point, despite the fact that only three guidelines were explicitly evidence based (and may not have been efficient). Overall, the methods of the economic evaluations and cost analyses were poor. The viewpoint adopted in economic evaluations was only stated in ten studies. The methods to estimate costs were comprehensive in about half of the studies, and few studies reported details of resource use. Owing to the poor quality of reporting of the economic evaluation, data on resource use and cost of guideline development, dissemination and implementation were not available for most of the studies; only four studies provided sufficiently robust data for abstraction.

Survey of estimating the feasibility and likely resource requirements of guideline dissemination and implementation strategies in UK settings

Respondents rarely identified existing budgets to support guideline dissemination and implementation strategies and made frequent comments about using ‘soft money’ or resources for specific initiatives to support such activities. In general, the respondents thought that only dissemination of educational materials and short (lunchtime) educational meetings were generally feasible within current resources.

Conclusions: implications for healthcare and recommendations for research

There is an imperfect evidence base to support decisions about which guideline dissemination and implementation strategies are likely to be efficient under different circumstances. Decision makers need to use considerable judgement about how best to use the limited resources they have for clinical governance and related activities to maximise population benefits. They need to consider the potential clinical areas for clinical effectiveness activities, the likely benefits and costs required to introduce guidelines and the likely benefits and costs as a result of any changes in

provider behaviour. Further research is required to: develop and validate a coherent theoretical framework of health professional and organisational behaviour and behaviour change to inform better the choice of interventions in research and service settings, and to estimate the efficiency of dissemination and implementation strategies in the presence of different barriers and effect modifiers.

Chapter 1
Introduction

Background

Clinical practice guidelines are an increasingly common element of clinical care throughout the world. Health systems are investing substantial resources in the development and introduction of clinical guidelines in the belief that they will inform clinical practice promoting effective and cost-effective healthcare. In the UK, the National Institute for Clinical Effectiveness (NICE) and the Scottish Intercollegiate Guidelines Network (SIGN) have been established to develop national evidence-based guidelines.

Such guidelines have the potential to improve the care received by patients by promoting interventions of proven benefit and discouraging ineffective interventions.[1] However, the development and introduction of guidelines are not themselves without costs. In some circumstances, the costs of development and introduction are likely to outweigh their potential benefits. In other circumstances, it may be more efficient to adopt less costly but less effective dissemination and implementation strategies.

Local healthcare organisations have relatively few resources for clinical effectiveness activities and policy makers need to consider how best to use these to maximise benefits. To make informed judgements, policy makers need to consider a range of factors. First, what are the potential clinical areas for clinical effectiveness activities? These may reflect national or local priorities. Policy makers should consider the prevalence of the condition, whether there are effective and efficient healthcare interventions available to improve patient outcome, and whether there is evidence that current practice is suboptimal. Second, what are the likely benefits and costs required to introduce guidelines? Policy makers need to consider the likely effectiveness of different dissemination and implementation strategies for the targeted condition in their settings and the resources required to deliver the different dissemination and implementation strategies. Third, what are the likely benefits and costs as a result of any changes in provider behaviour?

Despite the current interest in guidelines, there remains uncertainty about the likely effectiveness of different guideline dissemination and implementation strategies and resources required to deliver them. The recent *Effective Health Care* bulletin[2] identified seven systematic reviews of different guideline dissemination and implementation strategies.[3–9] In general, the systematic reviews observed that dissemination and implementation of guidelines could lead to changes in professional behaviour. The *Effective Health Care* bulletin[4] on implementing guidelines concluded that they were more likely to be effective if they take into account local circumstances, are disseminated by an active educational intervention and are implemented by patient specific reminders. Oxman and colleagues[5] concluded that there are "no magic bullets" for provider behaviour change; a range of interventions can lead to provider behaviour change, but no single intervention is always effective for changing behaviour. Wensing and colleagues[8] concluded that multifaceted interventions combining more than one intervention tended to be more effective but may be more expensive. However, many of these reviews failed to account for methodological weaknesses in the primary studies (such as potential unit of analysis errors) and tended to use vote-counting techniques (see Appendix 1 for further discussion of these issues). As a result, they provide little information about the likely effect sizes of different interventions and how this might vary, for example, across different settings, different targeted groups and different clinical activities. None of them considered the resources required to deliver the intervention or the likely efficiency of the interventions. As a result, purchasers and providers have little to guide them about when it would be worthwhile to devote resources to guideline development and introduction for specific clinical problems.

The study aimed to develop a framework for deciding when it is worthwhile to develop and introduce clinical guidelines, based upon the potential costs and benefits of the targeted clinical activity, the current levels of activity (and thus potential for health gain), and the likelihood that the development and introduction of clinical guidelines will incur such costs and achieve such benefits.

Aims of study

The aims of the study were:

- to undertake a systematic review of the effectiveness and costs of different guideline development, dissemination and implementation strategies
- to estimate the resource implications of different development, dissemination and implementation strategies
- to develop a framework for deciding when it is efficient to develop and introduce clinical guidelines based upon the potential costs and benefits of the targeted clinical activity and the effectiveness and costs of guideline development and introduction.

Structure of this report

Chapter 2 discusses the methods of the systematic review of guideline dissemination and implementation strategies. Appendices 1–3 provide additional details of the statistical methods used, development of search strategies and data checklist. Chapter 3 describes the characteristics of the included studies. Chapter 4 describes the results of the systematic review. Appendices 4–6 include bibliographic details of included studies, tables of the characteristics of included studies and results. Chapter 5 describes the methods and results of a review of economic evaluations and cost analyses of guideline implementation strategies undertaken by the included studies. Chapter 6 describes the results of telephone interviews with key informants about the feasibility of using different dissemination and implementation strategies and factors that might influence the resources required to deliver them. Chapter 7 discusses the results of the study and its strengths and limitations, and presents the main conclusions and recommendations of the study.

Chapter 2

Systematic review of guideline dissemination and implementation strategies: aims and methods

Aims

The aims of the review were:

- to identify rigorous evaluations of the introduction of clinical guidelines into medical practice
- to estimate the effectiveness of guideline dissemination and implementation strategies to promote improved professional practice.

Methods

The study followed the methods proposed by the Cochrane Effective Practice and Organisation of Care (EPOC) group.[10]

Selection criteria for studies

Studies were selected for inclusion if they met the following criteria.

Study design

- Randomised controlled trials (RCT), involving either individual randomisation at the level of the patient (P-RCT) or cluster randomisation at the level of professional, practice or healthcare organisation (C-RCT)
- controlled clinical trials (CCT), involving either individual allocation at the level of the patient (P-CCT) or cluster allocation at the level of professional, practice or healthcare organisation (C-CCT)
- controlled before and after (CBA) studies
- interrupted time series (ITS) designs.

Further details describing these study designs and the rationale for including them in the review can be found in the statistics appendix (Appendix 1).

Participants

- Medically qualified healthcare professionals.

If the guidelines were aimed at multiprofessional groups or other healthcare professionals, studies were included only if the results for medical healthcare professionals were reported separately or medical healthcare professionals represented more than 50% of the targeted population. Studies evaluating the introduction of guidelines targeting undergraduate medical students, patients or the general public were excluded.

Interventions

For the purpose of this review, the broad definition of clinical guidelines as "systematically developed statements to assist practitioner decisions about appropriate health care for specific clinical circumstances"[11] was adopted. Studies that evaluated an intervention that was considered by the reviewers to meet the above criteria, regardless of whether the term 'guideline' was explicitly mentioned, were included in the review.

Outcomes

- Objective measures of provider behaviour and/or patient outcome.

Studies that use aggregate self-report data on provider behaviour, e.g. 'do you usually measure patients' blood pressure during visits?' were excluded. Studies that evaluated only medical professionals' satisfaction and/or knowledge were excluded.

Search strategy

Electronic searches

Electronic searches were undertaken of the following databases:

- MEDLINE (1966–1998)
- Healthstar (1975–1998)
- Cochrane Controlled Trial Register (4th edition 1998)
- EMBASE (1980–1998)
- SIGLE (1980–1988)
- the EPOC specialised register.

Details of the search strategies used and their development are given in Appendix 2.

Bibliographies of previous systematic reviews

In addition, the reviewers checked the reference lists of 51 systematic reviews of professional behaviour change strategies identified in the *Effective Health Care* bulletin on 'Getting evidence into practice'.[2]

There was no restriction on language of publication.

Screening of search results

RT and CF screened the search results to identify potentially relevant studies. Hard copies of potentially relevant studies were obtained and assessed against the inclusion criteria of the review by RT and LS. Disagreements were resolved by consensus in discussion with JG.

Data abstraction and quality assessment (see Appendix 3 for the data abstraction checklist)

Two reviewers (RT and one of RD, ME, JG, LM, GM, LS, MW, PW) independently abstracted data on:

- study design
- methodological quality of the studies (using the EPOC methodological quality criteria; see *Box 1*)
- participants (characteristics of professionals, including clinical speciality, level of training, age, years since graduation, proportion of eligible providers participating in study; characteristics of patients, including clinical problem, age, gender, ethnicity; numbers of patients, providers, practices/hospitals/communities participating in the study)
- study settings (reimbursement system, setting of care, academic status, country)
- targeted behaviours
- characteristics of interventions (characteristics of clinical guidelines, whether a gap analysis had been undertaken[13] and type of intervention (see below for further details)
- study results (see below for further details).

Classification of interventions

Dissemination and implementation strategies were classified according to a taxonomy of professional, organisational, financial and regulatory interventions that was developed by EPOC (see *Box 2* for categories of professional interventions). The majority of interventions targeted professionals. There were also some interventions targeting patients (e.g. patient-mediated interventions, patient financial incentives); given the small number of these interventions, they were collapsed into one category called patient directed interventions. Similarly, all organisational and professional financial interventions were collapsed into single categories (organisational and financial, respectively)

Derivation of study effect sizes (see Appendix 1 for further details)

A study could report one or all of the following end-points: dichotomous process of care variable (e.g. proportion of patients receiving appropriate treatment); continuous process of care variable (e.g. number of prescriptions issued by providers) dichotomous outcome of care (e.g. proportion of patients who have stopped smoking) and continuous outcome of care (e.g. mean symptom score). Data were abstracted on each type of end-point. Where studies reported more than one measure of each end-point, the primary measure (as defined by the authors of the study) or the median measure was abstracted. For example, if the comparison reported five dichotomous process of care variables and none of them was denoted the primary variable, then the effect sizes for the five variables were ranked and the median value was taken.

The hypothesised direction of effect differed between studies, with some studies expecting an increase in outcome and others expecting a decrease. In all cases the effect size was standardised so that a positive difference between postintervention percentages or means was a good outcome.

The type of data abstracted varied by study design. For C-RCTs, P-RCTs, CCTs and CBAs data were abstracted on baseline performance, postintervention performance, absolute difference in postintervention performance, and significance level for absolute difference in postintervention period (and relative difference in postintervention performance and standardised mean difference for continuous data only). For ITS, data were abstracted on the change in slope and level. The reviewers attempted to reanalyse those ITS studies that either were analysed inappropriately or did not report comparable results (see below and Appendix 1 for further details).

Where comparable data could not be abstracted, usually because of the complexity of the design or analysis, the results were summarised in prose in the results table (Appendix 6) and separately in the text (Chapter 4).

BOX 1 *Cochrane EPOC methodological quality criteria*

1. Quality criteria for RCTs and CCTs
   (a) Concealment of allocation
   - DONE — Unit of allocation was institution, team or professional and any random process explicitly described, e.g. use of random number tables, OR unit of allocation was patient or episode of care and some form of centralised randomisation scheme, an on-site computer system or sealed opaque envelopes used
   - NOT CLEAR — Allocation procedure not described explicitly OR unit of allocation was patient or episode of care and reported use of 'list' or 'table', 'envelopes' or 'sealed envelopes' for allocation
   - NOT DONE — Use of alternation, such as reference to case record numbers, dates of birth, day of the week or any other such approach OR unit of allocation was patient or episode of care and reported use of any allocation process that is entirely transparent before assignment, such as an open list of random numbers or assignments OR allocation was altered by investigators, professionals or patients.

   (b) Follow-up of professionals (protection against exclusion bias)
   - DONE — Outcome measures for ≥ 80% of professionals randomised (Do not assume 100% follow-up unless stated explicitly)
   - NOT CLEAR — Not specified
   - NOT DONE — Outcome measures for < 80% of professionals randomised

   (c) Follow-up of patients or episodes of care
   - DONE — Outcome measures for ≥ 80% of patients randomised or patients who entered the trial (Do not assume 100% follow-up unless stated explicitly)
   - NOT CLEAR — Not specified
   - NOT DONE — Outcome measures for < 80% of patients randomised or patients who entered the trial.

   (d) Blinded assessment of primary outcome(s)[a] (protection against detection bias)
   - DONE — Stated explicitly that primary outcome variables were assessed blindly OR outcome variables are objective, e.g. length of hospital stay, drug levels assessed by a standardised test
   - NOT CLEAR — Not specified.
   - NOT DONE — Outcomes not assessed blindly.

   (e) Baseline measurement
   - DONE — Performance or patient outcomes measured prior to the intervention, and no substantial differences present across study groups
   - NOT CLEAR — Baseline measures not reported, or unclear whether baseline measures are different across study groups
   - NOT DONE — Differences at baseline in main outcome measures likely to undermine the postintervention differences, e.g. differences between groups before the intervention similar to those found postintervention.

   (f) Reliable primary outcome measure(s)[b]
   - DONE — Two or more raters with agreement ≥ 90% or kappa ≥ 0.8 OR outcome assessment is objective, e.g. length of hospital stay, drug levels assessed by a standardised test
   - NOT CLEAR — Reliability not reported for outcome measures obtained by chart extraction or collected by an individual
   - NOT DONE — Two or more raters with agreement < 90% or kappa < 0.8.

   (g) Protection against contamination
   - DONE — Allocation by community, institution or practice and unlikely that control group received the intervention
   - NOT CLEAR — Professionals allocated within a clinic or practice and possible that communication between experimental and control group professionals could have occurred
   - NOT DONE — Likely that control group received the intervention, e.g. cross-over trials or if patients rather than professionals were randomised.

2. Quality criteria for controlled before and after (CBA) designs
   (a) Baseline measurement
   - DONE — Performance or patient outcomes measured before the intervention, and no substantial differences present across study groups

continued

BOX 1 *Cochrane EPOC methodological quality criteria (cont'd)*

NOT CLEAR	Baseline measures not reported, or unclear whether baseline measures are different across study groups
NOT DONE	Differences at baseline in main outcome measures likely to undermine the postintervention differences, e.g. differences between groups before the intervention similar to those found postintervention.

(b) Characteristics of study and control

DONE	Characteristics of study and control providers are reported and similar
NOT CLEAR	It is not clear, e.g. characteristics are mentioned in the text but no data are presented
NOT DONE	There is no report of characteristics either in the text or a table OR if baseline characteristics are reported and there are differences between study and control providers.

(c) Blinded assessment of primary outcome(s)[a] (protection against detection bias)

DONE	Stated explicitly that primary outcome variables were assessed blindly OR outcome variables are objective, e.g. length of hospital stay, drug levels assessed by a standardised test
NOT CLEAR	Not specified
NOT DONE	Outcomes were not assessed blindly.

(d) Protection against contamination

DONE	Allocation by community, institution or practice and unlikely that control group received the intervention
NOT CLEAR	Professionals allocated within a clinic or practice and possible that communication between experimental and control group professionals could have occurred
NOT DONE	Likely that control group received the intervention, e.g. cross-over trials or if patients rather than professionals were randomised.

(e) Reliable primary outcome measure(s)[b]

DONE	Two or more raters with agreement ≥ 90% or kappa ≥ 0.8 OR outcome assessment is objective, e.g. length of hospital stay, drug levels assessed by a standardised test
NOT CLEAR	Reliability not reported for outcome measures obtained by chart extraction or collected by an individual
NOT DONE	Two or more raters with agreement < 90% or kappa < 0.8

(f) Follow-up of professionals (protection against exclusion bias)

DONE	Outcome measures for ≥ 80% of professionals randomised (Do not assume 100% follow-up unless stated explicitly)
NOT CLEAR	Not specified
NOT DONE	Outcome measures for < 80% of professionals randomised.

(g) Follow-up of patients

DONE	Outcome measures for ≥ 80% of patients randomised or patients who entered the trial (Do not assume 100% follow-up unless stated explicitly)
NOT CLEAR	Not specified
NOT DONE	Outcome measures for < 80% of patients randomised or patients who entered the trial.

3. Quality criteria for ITSs

Protection against secular changes

(a) The intervention is independent of other changes

DONE	The intervention occurred independently of other changes over time
NOT CLEAR	Not specified (will be treated as NOT DONE if information cannot be obtained from the authors)
NOT DONE	Reported that the intervention was not independent of other changes in time.

(b) There are sufficient data points to enable reliable statistical inference

DONE	At least 20 points are recorded before the intervention AND the authors have done a traditional time series analysis (ARIMA model) (or a post hoc analysis can be done) **OR** at least 3 points are recorded pre- and postintervention AND the authors have done a repeated measures analysis (or a post hoc analysis can be done)

continued

BOX 1 *Cochrane EPOC methodological quality criteria (cont'd)*

	OR at least 3 points are recorded pre- and postintervention AND the authors have used ANOVA or multiple *t*-tests (or a post hoc analysis can be done) AND there are at least 30 observations per data point
NOT CLEAR	Not specified, e.g. number of discrete data points not mentioned in text or tables (treated as NOT DONE if information cannot be obtained from the authors)
NOT DONE	Any of the conditions above are unmet.
(c) Formal test for trend (complete this section if authors have used ANOVA modelling)	
DONE	Formal test for change in trend using appropriate method is reported (e.g. see Cook and Campbell)[12] (or can be redone)
NOT CLEAR	Not specified (will be treated as NOT DONE if information cannot be obtained from the authors)
NOT DONE	Formal test for change in trend has not been done.
Protection against detection bias	
(d) Intervention unlikely to affect data collection	
DONE	Reported that intervention itself was unlikely to affect data collection, e.g. sources and methods of data collection were the same before and after the intervention
NOT CLEAR	Not specified (treated as NOT DONE if information cannot be obtained from the authors)
NOT DONE	Intervention itself was likely to affect data collection, e.g. any change in source or method of data collection reported.
(e) Blinded assessment of primary outcome(s)[a]	
DONE	Stated explicitly that primary outcome variables were assessed blindly OR outcome variables are objective, e.g. length of hospital stay, drug levels assessed by a standardised test
NOT CLEAR	Not specified (treated as NOT DONE if information cannot be obtained from the authors)
NOT DONE	Outcomes were not assessed blindly.
(f) Completeness of data set	
DONE	Data set covers 80–100% of total number of participants or episodes of care in the study
NOT CLEAR	Not specified (will be treated as NOT DONE if information cannot be obtained from the authors)
NOT DONE	Data set covers less than 80% of the total number of participants or episodes of care in the study.
(g) Reliable primary outcome measure(s)[b]	
DONE	Two or more raters with agreement ≥ 90% or kappa ≥ 0.8 OR outcome assessment is objective, e.g. length of hospital stay, drug levels assessed by a standardised test
NOT CLEAR	Reliability not reported for outcome measures obtained by chart extraction or collected by an individual (will be treated as NOT DONE if information cannot be obtained from the authors)
NOT DONE	Two or more raters with agreement < 90% or kappa < 0.8.

[a] Primary outcome(s) are those variables that correspond to the primary hypothesis or question as defined by the authors. In the event that some of the primary outcome variables were assessed in a blind fashion and others were not, score each separately on the back of the form and label each outcome variable clearly.

[b] In the event that some outcome variables were assessed in a reliable fashion and others were not, score each separately on the back of the form and label each outcome variable clearly.

ANOVA, analysis of variance; ARIMA, autoregressive integrated moving average.

BOX 2 *Classification of professional interventions from EPOC taxonomy*

- **Distribution of educational materials**: distribution of published or printed recommendations for clinical care, including clinical practice guidelines, audiovisual materials and electronic publications. The materials may have been delivered personally or through mass mailings.
- **Educational meetings**: healthcare providers who have participated in conferences, lectures, workshops or traineeships.
- **Local consensus processes**: inclusion of participating providers in discussion to ensure that they agreed that the chosen clinical problem was important and the approach to managing the problem was appropriate.
- **Educational outreach visits**: use of a trained person who met with providers in their practice settings to give information with the intent of changing the provider's practice. The information given may have included feedback on the performance of the provider(s).
- **Local opinion leaders**: use of providers nominated by their colleagues as 'educationally influential'. The investigators must have explicitly stated that their colleagues identified the opinion leaders.
- **Patient-mediated interventions**: new clinical information (not previously available) collected directly from patients and given to the provider, e.g. depression scores from an instrument.
- **Audit and feedback**: any summary of clinical performance of healthcare over a specified period. The summary may also have included recommendations for clinical action. The information may have been obtained from medical records, computerised databases or observations from patients.
  The following interventions are excluded:
  – provision of new clinical information not directly reflecting provider performance which was collected from patients, e.g. scores on a depression instrument, abnormal test results. These interventions should be described as patient mediated
  – feedback of individual patients' health record information in an alternative format (e.g. computerised). These interventions should be described as organisational.
- **Reminders**: patient- or encounter-specific information, provided verbally, on paper or on a computer screen, which is designed or intended to prompt a health professional to recall information. This would usually be encountered through their general education, in the medical records or through interactions with peers, and so remind them to perform or avoid some action to aid individual patient care. Computer-aided decision support and drugs dosage are included.
- **Marketing**: use of personal interviewing, group discussion ('focus groups'), or a survey of targeted providers to identify barriers to change and subsequent design of an intervention that addresses identified barriers.
- **Mass media**: (1) varied use of communication that reached great numbers of people including television, radio, newspapers, posters, leaflets and booklets, alone or in conjunction with other interventions; (2) targeted at the population level.
- **Other**: other categories to be agreed in consultation with the EPOC editorial team.

Reanalyses of studies (see Appendix 1 for further details)

Within-group pre–post comparisons

For C-RCTs, P-RCTs, CCTs and CBAs, an attempt was made to reanalyse the across group postintervention comparison if studies only reported or analysed within-group pre–post comparisons. If reanalysis was not possible, the point estimates for the across-group postintervention comparison was reported without any estimate of the statistical significance.

Potential unit of analysis errors

In many C-RCT studies the practitioner was randomised but during the statistical analyses individual patient data were analysed as if there was no clustering within practitioner. C-RCT studies that do not account for clustering during analysis have 'unit of analysis errors'.[14] Although the point estimate of effect is unlikely to be biased, the p-values are likely to be artificially extreme and confidence intervals overly narrow, increasing the chances of spuriously significant findings and misleading conclusions.[15] If C-RCTs had potential unit of analysis errors, the reviewers attempted to reanalyse the trial using cluster level data. If reanalysis was not possible the point estimates were reported without any estimate of the statistical significance.

ITS studies

For ITS studies, each study was reanalysed (where possible) using time series regression methods (see Appendix 1). Data were derived from tables of results presented in the original studies. Where information on individual values was reported graphically only, data were derived by measuring each data point manually. This was done using a flatbed scanner to obtain a digital image of the graph. The images were then imported into a

graphics package, gridlines applied and the corresponding values read off. In some studies, derivation of data points from graphs was not possible owing to the scales used. Time series regressions were undertaken to estimate two effect sizes for each ITS comparison. First, a change in the level of outcome immediately after the introduction of the intervention was estimated. This was done by extrapolating the preintervention regression line to the first point postintervention. The difference between this extrapolated point and the postintervention regression estimate for the same point gave the change in level estimate. Further mathematical details are available from the authors. Second, a change in the slopes of the regression lines was estimated (calculated as postintervention slope minus preintervention slope). Both of these effect sizes were necessary for interpreting the results of each study. For example, there could have been no change in the level immediately after the intervention, but there could have been a significant change in slope. Reporting only the change in level estimate in this example would have misled the reader. The direction of effect was standardised so that a positive level or slope estimate was considered a good outcome and a negative estimate was a poor outcome.

Data handling and management

All included studies were given a unique identifying number. Data were entered onto an Access database. Any disagreements were resolved by consensus.

Data synthesis and presentation

Given the expected extreme heterogeneity within the review and the number of studies with potential unit of analysis errors, there was no plan to undertake formal meta-analysis. Further discussion of this issue can be found in the analytical framework section of Appendix 1. Previous qualitative systematic reviews of implementation strategies have largely used vote-counting methods that add up the number of positive and negative studies and conclude whether the interventions were effective on this basis.[2,16] However, this method does not incorporate the precision of the estimate from primary studies, does not provide an estimate of the effect size of the intervention and has low statistical power. A more explicit analytical framework was used within this review.

Separate analyses were undertaken for comparisons of single interventions against 'no-intervention' controls (i.e. 'usual care' or control groups that did not receive any interventions), single interventions against 'intervention' controls (i.e. control groups that did receive an intervention), multifaceted interventions against 'no-intervention' controls and multifaceted interventions against 'intervention' controls. For the multifaceted interventions, separate analyses were undertaken for combinations including educational outreach and combinations with more than four comparisons. The reviewers also explored whether the effects of multifaceted interventions increased with the number of components. Although the original plan was to undertake a metaregression analysis to estimate the effects of different interventions, the number of different combinations of multifaceted interventions evaluated proved problematic (see Appendix 1 for further details).

Bibliographic details of the included studies are reported in Appendix 4, details of the characteristics of included studies (including quality assessment) in Appendix 5 and the results of all comparisons in Appendix 6. Dichotomous process of care measures were used as the primary effect size for each comparison, for two pragmatic reasons: first, they were reported considerably more frequently in the studies, and second, continuous process of care measures were less stable. For example, a relative percentage change in a continuous measure depends on the scale being used: a comparison that shifts from a mean of 1 to 2 will show the same relative improvement as one that shifts from 25 to 50. To counter this standardised mean differences were calculated where possible, but there were rarely enough data presented in the paper to do this. For completeness, dichotomous and continuous outcome results are also reported.

For C-RCT, P-RCT, CCT and CBA comparisons,

- the number of comparisons showing a positive direction of effect
- the median effect size across all comparisons
- the median effect size across comparisons without unit of analysis errors
- and the number of comparisons showing statistically significant effects

were reported (separately for each study design). This allows the reader to assess the consistency of effects across different study designs and across comparisons where the statistical significance is known.

The paper also presents stacked bar charts of estimates of dichotomous process measures (across all designs) against observed effect size, distinguishing between studies reporting significant effects, non-significant effects and studies with unit of analysis errors. This allows the reader to assess the median and range of effect sizes across all studies.

To describe the size of effect (absolute difference across postintervention measures) for process dichotomous measures, the following terms were used:

- 'small' to describe effect sizes ≤ 5%
- 'modest' to describe effect sizes > 5% and ≤ 10%
- 'moderate' to describe effect sizes >10 and ≤ 20%
- 'large' to describe effect sizes >20%.

For ITS comparisons, the significance of changes in level and slope is reported.

Chapter 3

Systematic review of guideline dissemination and implementation strategies: details of included studies

Results of literature searches

The search strategy produced over 150,000 hits. The titles and abstracts of these were screened and around 5000 were initially identified as potentially relevant. The full text of 863 potentially relevant reports of studies was retrieved and assessed for inclusion in this review. In total, 285 reports of 235 studies yielding 309 separate comparisons were included in the review (see *Figure 1*). Bibliographic details of the included studies (including all relevant papers identified) are provided in Appendix 4. Appendix 5 tables details of all the included studies. Details of the excluded studies are available from the authors.

Methodological quality of included studies

The studies included 110 (46.8%) C-RCTs, 29 (12.3%) P-RCTs, 7 (3.0%) C-CCTs, 10 (4.3%) P-CCT, 40 CBAs (17.0%) and 39 (16.6%) ITS designs.

Methodological quality

Methodological quality was assessed using the Cochrane EPOC group criteria (see Appendix 3 for further details). Overall, the quality of studies was difficult to determine owing to poor reporting. The quality of the studies by design and allocation level is summarised in *Tables 1–3*.

RCTs and CCTs (*Table 1*)

Concealment of allocation was 'not clear' or 'not done' for the majority of the RCTs and CCTs. Protection against contamination was 'done' for just over half (54%) of the C-RCTs (i.e. it was unlikely that the control group received the intervention), but not clear for 41% of them. In contrast, protection against contamination was not done for the majority of P-RCTs and P-CCTs.

It was not clear whether outcomes had been blindly assessed for over 70% of C-RCTs and CCTs and 59% of P-RCTS. The reliability of the outcome measurement and follow-up of ≥ 80% of the professionals included was not clear from the reports of over 70% of the randomised and

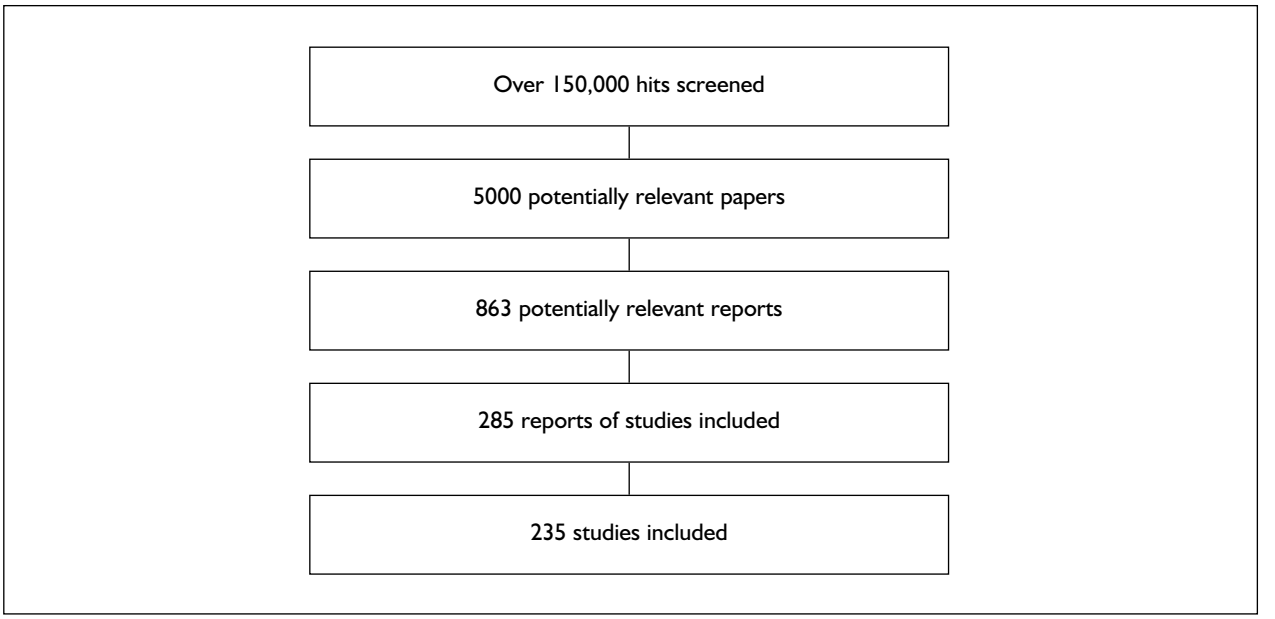

FIGURE 1 *Flowchart of included studies*

TABLE 1 *Quality criteria for RCTs and CCTs*

Criterion		Done *n* (%)	Not clear *n* (%)	Not done *n* (%)
Concealment of allocation	C-RCT	29 (26.4)	80 (72.7)	1 (0.9)
	P-RCT	4 (13.8)	24 (82.8)	1 (3.4)
	C-CCT	0	2 (28.6)	5 (71.4)
	P-CCT	0	0	10 (100)
Protection against contamination	C-RCT	60 (54.5)	45 (41)	5 (4.5)
	P-RCT	0	5 (17.2)	24 (82.8)
	C-CCT	0	5 (71.4)	2 (28.6)
	P-CCT	0	0	10 (100)
Blinded assessment of outcome	C-RCT	21 (19.1)	85 (77.3)	4 (3.6)
	P-RCT	11 (37.9)	17 (58.6)	1 (3.4)
	C-CCT	2 (28.6)	5 (71.4)	0
	P-CCT	2 (20.0)	8 (80.0)	0
Reliable outcome measure	C-RCT	27 (24.5)	80 (72.7)	3 (2.7)
	P-RCT	3 (10.3)	26 (89.7)	0
	C-CCT	1 (14.3)	5 (71.5)	1 (14.3)
	P-CCT	0	10 (100)	0
Outcomes measured at baseline and no differences	C-RCT	46 (41.8)	58 (52.7)	6 (5.5)
	P-RCT	9 (31.0)	20 (69.0)	0
	C-CCT	1 (14.3)	6 (85.7)	0
	P-CCT	0	10 (100)	0
Follow-up of professionals	C-RCT	28 (25.5)	78 (70.9)	4 (3.6)
	P-RCT	0	29 (100)	0
	C-CCT	0	7 (100)	0
	P-CCT	1 (10)	9 (90)	0
Follow-up of patients	C-RCT	29 (26.4)	78 (70.9)	3 (2.7)
	P-RCT	16 (55.2)	10 (34.5)	3 (10.3)
	C-CCT	4 (57.1)	3 (42.9)	0
	P-CCT	3 (30)	7 (70)	0

controlled clinical trials. Follow-up of patients was not clear for 71% of C-RCTs. Measurement of outcomes at baseline and no differences at baseline was done for 42% of C-RCTs and 14% of C-CCTs.

CBAs (*Table 2*)

The criterion 'characteristics of study and control sites reported and similar' was not clear or not done for 77.5% of CBAs and 'outcomes measured at baseline and no differences' was not clear or not done for 57.5% of CBAs. Protection against contamination was done for 62.5% of studies.

Blinded assessment of outcome, reliability of outcome measurement and follow-up of ≥ 80% of patients was not clear for over 75% of studies. Follow-up of ≥ 80% of professionals was not clear for 67.5% of studies.

ITSs (*Table 3*)

The intervention was deemed independent of other changes in only 36% of the ITS studies and deemed not to affect data collection in 92% of studies, but the completeness of the data set was not clear in 90% of studies. Sixteen (41%) of the 39 ITS studies were analysed appropriately. Blinded assessment and reliable outcomes were not clear for 14/39 (36%) of the ITS studies, but done for the remainder.

Other methodological issues

Sample size calculations

Most of the RCTs, CCTs and CBA studies did not report a sample size calculation (135/196, 69%). In 44 (22%) studies it was clear that a sample size calculation had been done and was fully reported,

TABLE 2 *Quality criteria for CBA studies*

Criterion	Done *n* (%)	Not clear *n* (%)	Not done *n* (%)
Characteristics of study and control reported and similar	9 (22.5)	17 (42.5)	14 (35.0)
Protection against contamination	25 (62.5)	14 (35)	1 (2.5)
Outcomes measured at baseline and no differences	17 (42.5)	16 (40.0)	7 (17.5)
Blinded assessment of outcome	6 (15)	31 (77.5)	3 (7.5)
Reliable outcome measure	6 (15)	34 (85)	0
Follow-up of professionals	12 (30)	27 (67.5)	1 (2.5)
Follow-up of patients	4 (10)	35 (87.5)	1 (2.5)

TABLE 3 *Quality criteria for ITS designs*

Criterion	Done *n* (%)	Not clear *n* (%)	Not done *n* (%)
The intervention is independent of other changes	14 (36)	24 (61)	1 (3)
The intervention is unlikely to affect data collection	36 (92)	3 (8)	0
Blinded assessment of outcome	25 (64)	14 (36)	0
Reliable outcome measure	25 (64)	14 (36)	0
Completeness of data set	3 (7)	35 (90)	1 (3)
Analysed appropriately	16 (41)	11 (28)	12 (31)

but 15 (34%) of these appeared to be retrospective calculations. In the remaining 17 studies it was not clear whether a sample size calculation had been done.

Potential unit of analysis errors

Potential unit of analysis errors were present in the main analyses of 53% C-RCTs, 86% of C-CCTs and 83% of CBAs.

Data points and intervals: ITS designs

The median numbers of data points before and after the intervention for ITS studies were ten [interquartile range (IQR) 5–17] and 12 (IQR 7–24), respectively. The interval between data points was 1 month in 64% of studies. The minimum time interval was 5 days and the maximum was 1 year.

Data reported

Process of care was measured in 95% of studies; 67% of studies reported dichotomous process measures and 39% reported continuous process measures. Outcome of care was reported in 22% of studies; 14% reported dichotomous outcome measures and 14% reported continuous outcome measures.

Twenty-nine per cent of studies reported economic data (see Chapter 5 for further details).

Details of comparisons

Seventy-nine per cent of studies involved only one comparison of an intervention group versus control group, 13% involved two comparisons (e.g. three-arm C-RCT) and 8% involved three or more comparisons.

In 69% of RCTs, CCTs and CBAs the control group did not receive any intervention, in 22% the control received a single intervention and in 10% the control received more than one intervention.

Details of participants

Number of allocation units: RCTs, CCTs and CBAs

The number of allocation units in each study group was reported in 77% of the studies. The

remainder tended to report an overall number taking part in the study or the number of other units per study group, for example, practices were allocated to study groups but the number of individual GPs per group was reported. The median number of allocation units per study arm was 18 (IQR 2–65). The studies that allocated to study arm by cluster (C-RCTs, C-CCTs, CBAs) had a median number of seven units per arm (IQR 1–24). The median for non-cluster (P-RCTs, P-CCTs) studies was 113 (IQR 61–423).

Proportion of eligible providers or allocation units participating

The percentage of eligible providers or allocation units in the target sampling frame that participated was not clear in 186 (79%) studies. Where it was reported, the median value was 80% (IQR 45–95%).

Setting and professionals

The studies were conducted in 14 different countries: 71% were conducted in the USA, 11% in the UK, 6% in Canada, and 3% in Australia and The Netherlands. The remaining studies were conducted in Denmark, France, Germany, Israel, Mexico, New Zealand Norway, Sweden and Thailand.

The most common setting was primary care (39%), followed by inpatient settings (19%) and generalist outpatient or ambulatory care settings (although all of these studies were undertaken in the USA and may be equivalent to primary care) (19%). Thirty-six (15%) studies were based within mixed settings, either inpatient and outpatient, at the interface between settings, or a mix of community- and hospital-based care. The remaining studies were set in nursing homes or long-term care facilities (3%), emergency departments (2%), specialist outpatient care (1%) and a military medical clinic (0.4%). The setting was not clear from the reports of three (1%) studies.

Physicians or doctors alone were the target of the intervention in 174 (74%) of the studies; the remaining studies targeted physicians and other health professionals and workers within the organisations such as nurses, nurse practitioners, pharmacists, dieticians, physician's assistants and office staff.

Most studies (57%) involved only one medical speciality, most commonly general practice or family medicine (24%).

The level of training of providers was not clear from the reports of 51% studies. Twenty-six per cent involved both 'fully trained' providers and providers 'in training', 14% targeted only providers in training and 8% involved only fully trained providers.

Details of interventions

Characteristics of the clinical guidelines

The source of the guidelines was a national professional expert body or national government body in 35% of studies, local clinicians in 30%; and 10% were from some other source. The source of the guidelines was not clear in 25% of studies. The composition of the guideline development group was not clear in 81% of studies. The evidence base of the guideline recommendations was not clear in 94% of studies. They appeared to be based on good evidence in only 3% of studies.

The purpose of the recommendations was appropriate management in 81% of studies, cost containment in 4% and both appropriate management and cost containment in 14%. The purpose was not clear or other in three (1%) studies.

An increase in established management was the nature of the desired change in 40% of studies. In 25%, a modification in management was desired (i.e. an increased management frequency in one activity and a decrease in another) and a reduction in established management was required in 15% of studies. The nature of desired change was not clear in 11% of studies and combinations of desired change were required in 9%.

Targeted behaviours

Forty-seven per cent of studies targeted just one type of provider behaviour. The single behaviour most frequently targeted was general management of the problem in 19% of studies, followed by prescribing in 14% of studies and test ordering in 8%. The most frequently targeted behaviours in combination were prescribing (34% of studies), prevention (32% of studies), patient education and advice (31% of studies), test ordering (23% of studies) and general management (23% of studies). Other targeted behaviours included diagnosis, discharge planning, financial, procedures, professional to patient communication, record keeping and referrals.

Intervention strategies

Eighty-four of the 309 comparisons (27%) involved a study group receiving a single guideline

implementation intervention strategy versus a 'no-intervention' or 'usual care' control group and 1% of comparisons involved single intervention groups compared with a control group receiving a single intervention. One hundred and thirty-six (44%) compared a group receiving a multifaceted intervention (>1 intervention) with a no-intervention or usual care control group. Multifaceted intervention groups were compared with a control group receiving an intervention (≥ 1 intervention) for 85 (27%) of the comparisons.

The most frequent single intervention evaluated against a no-intervention control was reminders in 13% of all comparisons, followed by dissemination of educational materials in 6% of comparisons, audit and feedback in 4% of comparisons, and patient-directed interventions in 3% of comparisons (*Table 4*).

The intervention strategy used most frequently as part of multifaceted interventions was educational materials (evaluated in 48% of all comparisons) (*Table 5*), followed by educational meetings (41%), reminders (31%), and audit and feedback (24% comparisons).

One hundred and seventeen studies (including 136 comparisons) evaluated a total of 68 different combinations of interventions against a 'no-intervention' control group, and 61 studies (including 85 comparisons) evaluated 58 different combinations of interventions against a control that also received one or more interventions.

TABLE 4 *Single interventions versus 'no-intervention' controls*

Intervention	No. of comparisons (% of total, 309)
Educational materials	18 (6)
Educational meetings	3 (1)
Consensus processes	0 (0)
Educational outreach	1 (0.3)
Opinion leaders	0 (0)
Patient-directed interventions	8 (3)
Audit and feedback	12 (4)
Reminders	38 (13)
Other professional (including mass media and marketing)	2 (1)
Financial interventions	0 (0)
Organisational interventions	2 (1)
Structural interventions	0 (0)
Regulatory interventions	0 (0)

TABLE 5 *Interventions used in multifaceted interventions*

Intervention	No. of comparisons (% of total, 309)
Educational materials	147 (48)
Educational meetings	126 (41)
Consensus processes	16 (5)
Educational outreach	35 (11)
Opinion leaders	6 (2)
Patient-directed interventions	56 (18)
Audit and feedback	73 (24)
Reminders	95 (31)
Other professional (including mass media and marketing)	19 (6)
Financial interventions	12 (4)
Organisational interventions	45 (15)
Structural interventions	37 (12)
Regulatory interventions	3 (1)

Chapter 4

Systematic review of guideline dissemination and implementation strategies: results

This chapter presents the results of the systematic review of guideline dissemination and implementation strategies. The results were synthesised and reported in the following structured format (see also in Appendix 1, 'Analytical framework used in this review'):

- Evaluations of single interventions against 'no-intervention' control groups.
- Evaluations of single interventions against 'intervention' control groups.
- Evaluations of multifaceted interventions against 'no-intervention' controls:
  - all comparisons
  - multifaceted interventions incorporating educational outreach
  - combinations of educational materials and educational meetings
  - combinations of educational materials, and audit and feedback
  - combinations of reminders and patient-mediated interventions
  - combinations of educational materials, educational meetings, and audit and feedback
  - combinations of educational materials, educational meetings and organisational interventions.
- Evaluations of multifaceted interventions against 'intervention' controls:
  - all comparisons
  - multifaceted interventions incorporating educational outreach
  - educational materials and reminders compared with educational materials
  - educational materials, educational meetings and reminders compared with educational materials and educational meetings.

For each intervention above, study demographics, dichotomous process measures, continuous process measures, dichotomous outcome measures, continuous outcome measures, ITS results and summary are reported. References to included studies denoted 'A number' are given in Appendix 4. The results for all comparisons are presented in Appendix 6.

Evaluations of single interventions against no-intervention control groups

Eighty-one studies (involving 84 comparisons) evaluated single interventions compared with a no-intervention control. There were no comparisons evaluating educational outreach, local consensus processes, local opinion leaders, financial, structural or regulatory interventions as a single intervention.

Educational materials

Eighteen studies (involving 18 comparisons) evaluated the effects of disseminating educational materials, including seven C-RCTs (A15, A44, A60, A86, A140, A151, A168), two P-RCTs (A76, A91), two CBAs (A167, A198) and seven ITS (A19, A121, A167, A184, A188, A206, A234). One C-RCT (A140) also presented data on continuous process end-points based upon a CBA comparison. Eight studies took place in the USA, four in the UK, three in Australia, two in Canada and one in The Netherlands. The majority of studies were based in primary care settings (n = 10). The targeted behaviours were general management of a clinical problem in seven studies, prescribing in three studies, prevention services in three studies, test ordering in three studies and procedures in two studies.

Results from RCT, CCT and CBA comparisons

Dichotomous process measures (Figure 2). Five comparisons reported dichotomous process data, including four C-RCT comparisons (A15, A60, A151, A168) and one P-RCT comparison (A76). All C-RCT comparisons observed improvements in care; the median effect was +8.1% (range +3.6 to +17%) absolute improvement in performance. Two comparisons had potential unit of analysis errors (A15, A60) and the significance of one comparison could not be determined (A151). The remaining comparison without a potential unit of analysis error observed an effect of +6% [not significant (NS)] (A168). The P-RCT comparison observed an absolute deterioration in care of –8.3% (NS).

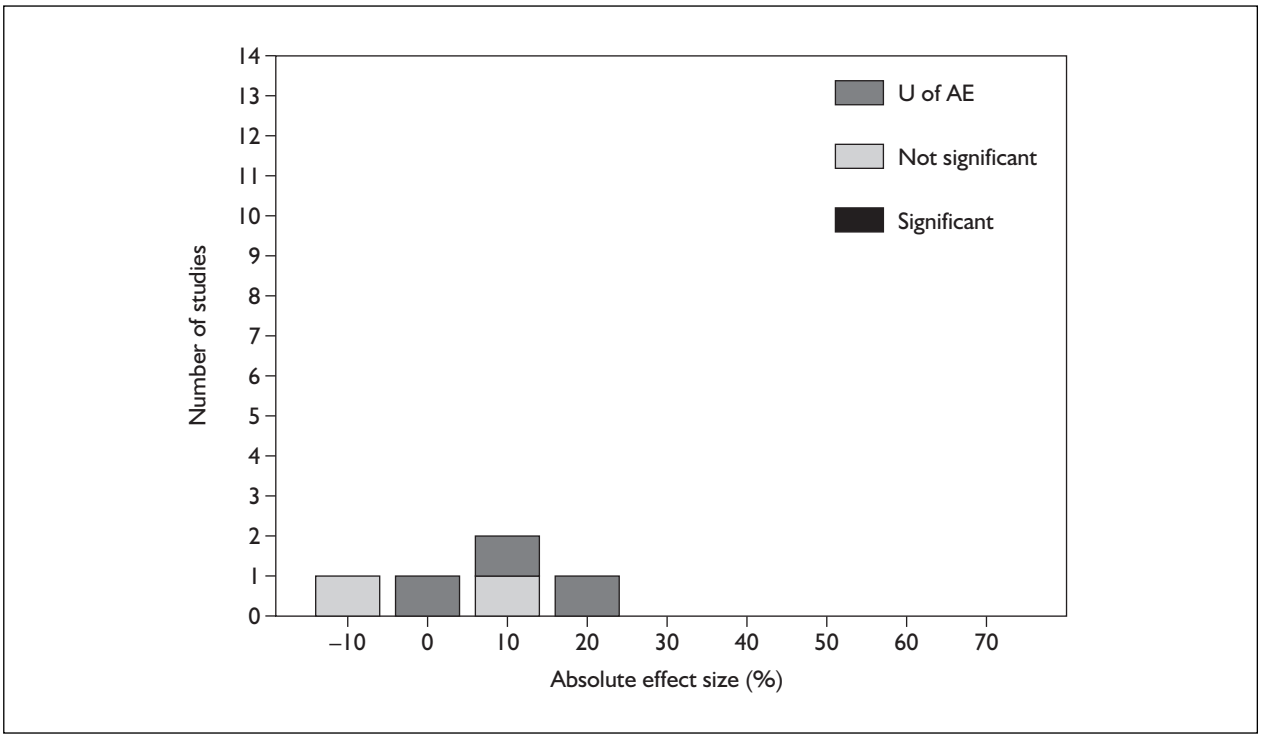

FIGURE 2 *Effect sizes observed in dichotomous process data in RCT, CCT and CBA comparisons of educational materials against a no-intervention control. U of AE: unit of analysis error*

Continuous process measures. Five comparisons reported continuous process data, including three C-RCT comparisons (A44, A86, A151) and two CBA comparisons (A140, A198). One C-RCT comparison (A151) reported a +34.7% relative improvement in performance and a standardised mean difference (SMD) of +0.25 (not significant). It was not possible to abstract comparable data for two C-RCT comparisons (A44, A86). Both reported non-significant effects but one comparison had a potential unit of analysis error (A86). The two CBA comparisons observed +77.3% (A140) and +100.3% (A198) relative improvements in performance. It was only possible to calculate an SMD for one comparison, +0.53 (A140). Both comparisons had potential unit of analysis errors.

Dichotomous outcome measures. Two comparisons reported dichotomous outcome data, including one P-RCT (A91) and one CBA comparison (A167). The P-RCT comparison reported a median effect of –4.6% absolute reduction in outcome (NS). It was not possible to abstract comparable data for the CBA comparison (A167), which reported a non-significant effect.

Continuous outcome measures. One C-RCT comparison (A140) reported a median effect of +17.1% relative improvement in outcome, SMD +0.86 ($p < 0.05$).

Results from ITS comparisons

Six comparisons were reanalysed using time series regression. The remaining comparison did not provide sufficient data for reanalysis but had undertaken time series regression of the change in slope (A121). Two comparisons reported significant improvements in level (A19, A170); however, both also reported deterioration in slope (statistically significant for A170), suggesting a decay effect. One comparison reported a significant deterioration in both level and slope (A234). The comparison that was not reanalysed (A121) observed a significant improvement in slope (data on change in level not reported).

Summary

The majority of studies evaluating dissemination of educational materials observed improvements in process of care. The effects were modest (absolute improvement across four C-RCT comparisons +8.1%, range +3.6 to +17%, relative improvement in one C-RCT of +0.25 SMD). One P-RCT comparison observed a deterioration in care. The statistical significance of two dichotomous process comparisons and one continuous process comparison could be determined; none was

significant. Three of six ITS comparisons observed significant improvements in performance (although two also observed possible decay effects), and one of the ITS comparisons observed a significant deterioration in performance. Only four studies were conducted in UK settings. These results suggest that educational materials may have a modest effect on guideline implementation that may be short lived. However, the evidence base is sparse and of poor quality.

Educational meetings

Three C-RCTs (A86, A191, A202,) (involving three comparisons) evaluated the effects of educational meetings against a no-intervention control. One study took place in the USA, one in the UK and the other in The Netherlands. Two studies were based in primary care settings and one in an inpatient setting. General management was the targeted behaviour of two studies and referrals of the other.

Results from RCT, CCT and CBA comparisons

Dichotomous process measures. One C-RCT comparison reported dichotomous process of care results (A191); the median effect size was +1% absolute improvement in performance; however, there was a potential unit of analysis error.

Continuous process measures. One C-RCT comparison reported continuous process measures (A202); the median effect was +27% relative improvement in performance. Insufficient data were presented to calculate an SMD. The comparison also had a potential unit of analysis error. It was not possible to abstract comparable data for one C-RCT (A86) that reported non-significant effects but had a potential unit of analysis error.

Continuous outcome measures. One C-RCT comparison reported continuous outcome measures (A191) and observed a median effect size of –3.6% relative change in performance. Insufficient data were presented to calculate an SMD and the comparison also had a potential unit of analysis error.

Summary

There are relatively few evaluations of educational meetings against a no-intervention control. The results suggest that the effects, if any, are likely to be small.

Audit and feedback

Ten studies (involving 12 comparisons) evaluated the effects of audit and feedback, including seven C-RCTs (A86, A125, A129, A130, A194, A209, A228,), one P-RCT (A174 – three comparisons), one CBA (A147) and one ITS (A75). Eight studies took place in the USA and two in the UK. Four studies were based in a generalist outpatient or ambulatory care setting, three in inpatient settings, two in primary care settings, and the setting was unclear in one study. The targeted behaviour was general management in three studies, prevention services in three studies, test ordering in three studies and discharge planning in one study.

Results from RCT, CCT and CBA comparisons

Dichotomous process measures (Figure 3). Six comparisons reported dichotomous process data, including five C-RCT comparisons (A129, A130, A194, A209, A228) and one CBA comparison (A147). All five C-RCT comparisons observed improvements in care. Across all comparisons, the median effect was +7.0% (range +1.3 to +16.0%) absolute improvement in performance. Three comparisons had potential unit of analysis errors (A129, A194, A209). The two remaining comparisons observed effects of +5.2% (NS) (A228) and +13% ($p < 0.05$) (A130). The CBA comparison (A147) observed an absolute improvement in performance of +32.6%, this study had a potential unit of analysis error.

Continuous process measures. Six comparisons reported continuous process measures, including three C-RCT comparisons (A86, A125, A129) and one P-RCT comparison (A174 – three comparisons). One C-RCT comparison (A129) observed +8.5% relative improvement in performance, SMD +0.2; however, the study had a potential unit of analysis error. In two C-RCT comparisons (A86, A125) it was not possible to abstract comparable data. The authors of both studies stated that the effects of audit and feedback were not significant; however, there was a potential unit of analysis error in both comparisons.

Two out of three P-RCT comparisons (A174) observed improvements in care; across all comparisons, the median effect was a +15.4% (range 0 to +20.3%) relative improvement in performance. Insufficient data were provided to calculate an SMD and the significance of these comparisons could not be determined.

Results from ITS comparisons

One comparison (A75) was reanalysed using time series regression and observed a significant change in level but not in slope.

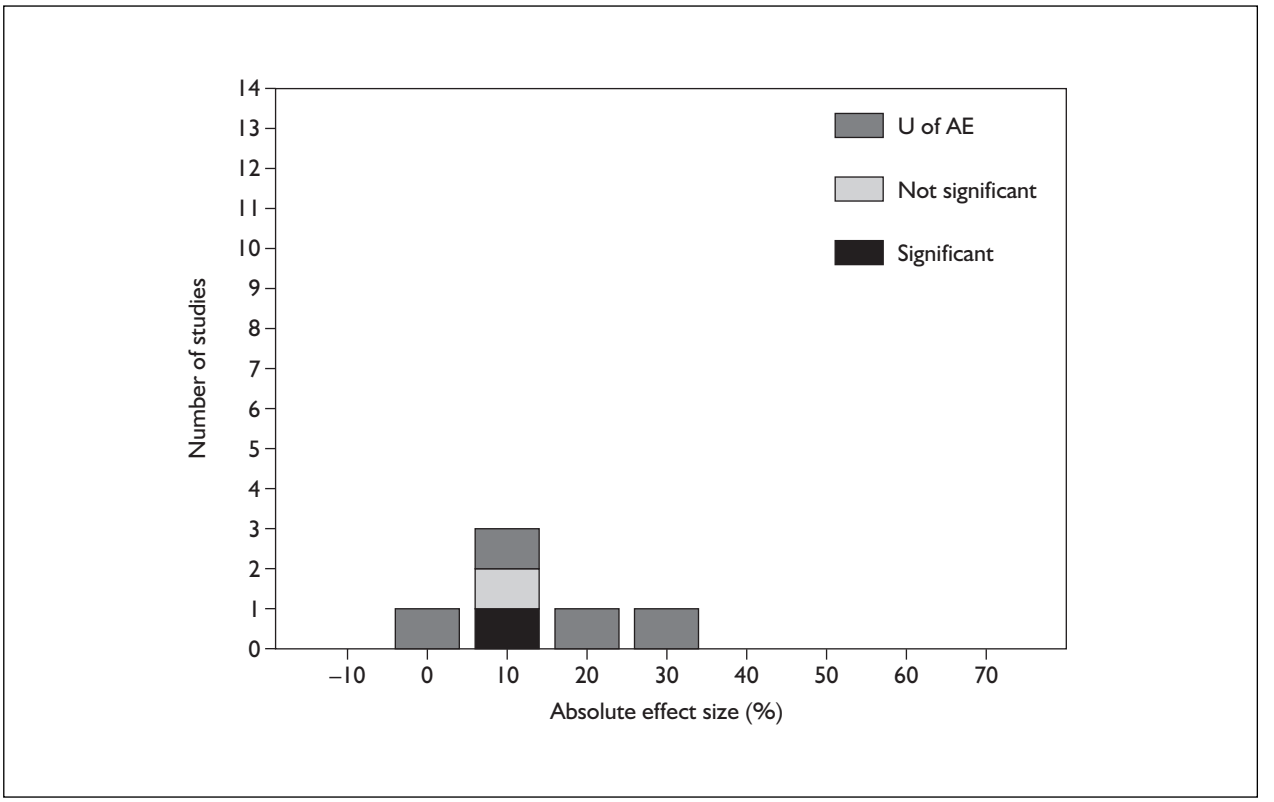

FIGURE 3 *Effect sizes observed in dichotomous process data in RCT, CCT and CBA comparisons of audit and feedback against a no-intervention control*

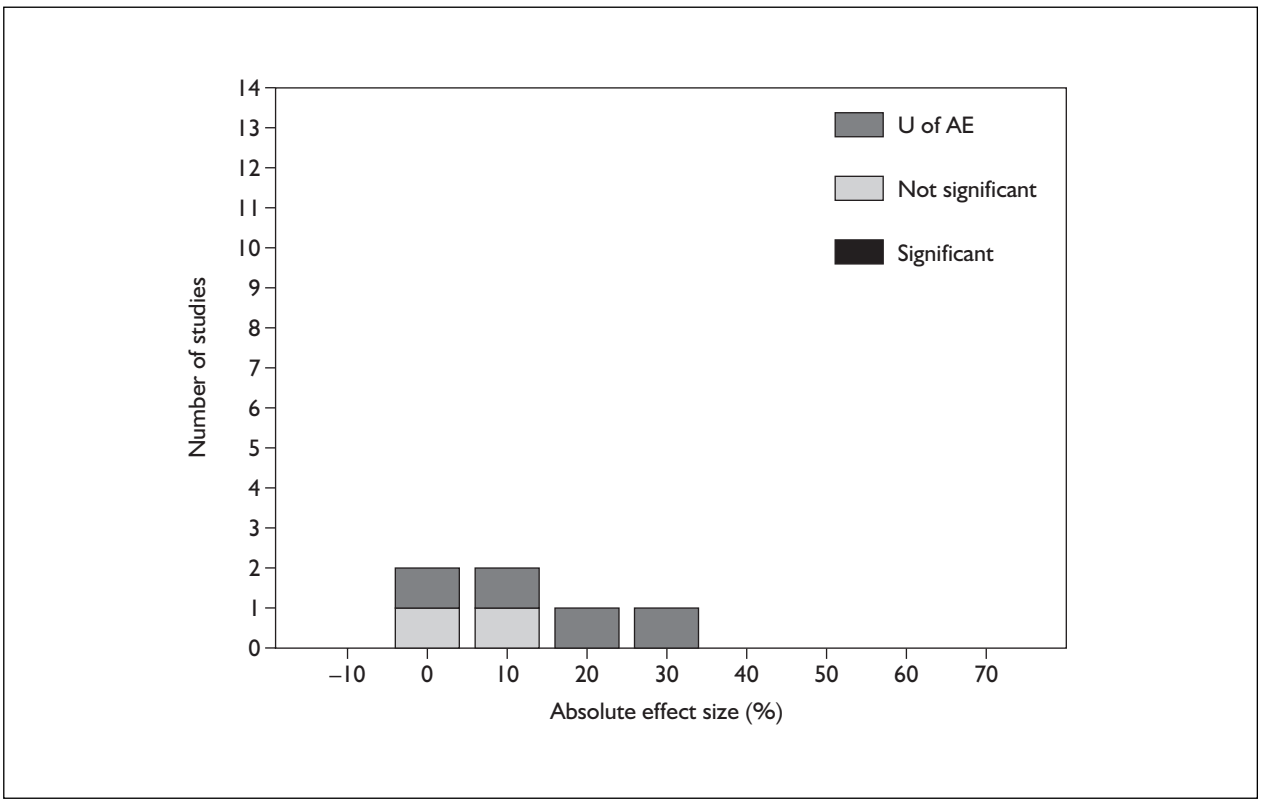

FIGURE 4 *Effect sizes observed in dichotomous process data in RCT, CCT and CBA comparisons of patient-directed interventions against a no-intervention control*

Summary
All studies evaluating audit and feedback observed improvements in care. The effects were modest, with an absolute improvement across five C-RCT comparisons of +7.0% (range +1.3 to +16.0%) and a relative improvement of +0.2 standardised mean difference across one study. The statistical significance of two process dichotomous comparisons could be determined; one was statistically significant. One CBA comparison observed large absolute improvements in performance (+32.6%), but had a unit of analysis error. One ITS comparison observed a significant change in level. Only two studies were conducted in UK settings, both of which targeted test ordering. The results suggest that audit and feedback may have a modest effect on guideline implementation.

Patient-directed interventions

Seven studies (involving eight comparisons) evaluated the effects of patient-directed interventions, including four C-RCTs (A24, A111, A179 – two comparisons, A219) and three P-RCTs (A18, A34, A76). Six studies took place in the USA and one in Canada. Four studies were based in primary care settings and three in generalist outpatient or ambulatory care settings. The targeted behaviour was prevention services in five studies and general management of the problem in the remaining two.

Results from RCT, CCT and CBA comparisons
Dichotomous process measures (Figure 4). Six comparisons reported dichotomous process data, including three C-RCT comparisons (A111, A179 – two comparisons) and three P-RCT comparisons (A18, A34, A76). All three C-RCT comparisons observed improvements in care; the median effect size was +20.8% (range +10.0 to +25.4%) absolute improvement in performance (however, all three comparisons had potential unit of analysis errors). All three P-RCT comparisons reporting dichotomous process of care variables observed improvements in care. The median effect size was +1.0% (range 0.8 to +9.0%) absolute improvement in performance; the statistical significance of one comparison could not be determined (A18) and the results of the other two comparisons were non-significant (A34, A76).

Continuous process measures. One C-RCT comparison (A24) observed a median effect of –9.1% relative deterioration in performance and an SMD of -0.67; however, the study had a potential unit of analysis error.

Continuous outcome measures. Two C-RCT comparisons (A24, A219) observed median effects of +9.10% and +5.1% relative improvement in performance and SMDs of +6.00 and +0.09; however, both comparisons had potential unit of analysis errors.

Summary
All studies observed improvements in care. The effects were moderate to large. The median absolute effects were +20.8% (range +10.0 to +25.4%) across three C-RCT comparisons and +1.0% (+0.8 to +9.0%) across three P-RCTs. The effect sizes from the P-RCT comparisons may underestimate the effects of the intervention if there was a contamination effect (see Appendix 1). Consequently, greater weight should be given to the C-RCTs. Unfortunately, all C-RCTs had unit of analysis errors. None of the studies was conducted in UK settings and the majority of the studies targeted preventive services. These results suggest that patient-mediated interventions may result in moderate to large improvements in performance, especially when targeting preventive services.

Reminders

Thirty-eight studies (involving 38 comparisons) evaluated the effects of reminders against a no-intervention/usual care control, including 16 C-RCTs (A10, A38, A40, A47, A81, A111, A119, A128, A134, A135, A146, A156, A179, A180, A182, A209), nine P-RCTs (A13, A16, A34, A37, A42, A98, A133, A204, A207), two C-CCTs (A46, A189), eight P-CCTs (A90, A112, A142, A222, A223, A224, A225, A226), two CBAs (A101, A187) and one ITS (A12).

Thirty-four studies were conducted in the USA, two in Israel, one in Canada and one in Thailand. Fifteen studies (40%) were based in generalist outpatient or ambulatory care settings, 11 (29%) in primary care settings, eight in impatient settings, two in mixed settings, one in a specialist outpatient setting and one in a military medical centre. The targeted behaviour was prevention services in 19 studies, general management in 13 studies, prescribing in three studies, and discharge planning, financial and procedures in the remaining three studies. Five studies were conducted in the Regenstrief Institute (A133, A134, A135, A156, A209) and five in the Ceders Sinai Medical Center (A222, A223, A224, A225, A226).

Results from RCT, CCT and CBA comparisons
Dichotomous process measures (Figure 5). Thirty-three comparisons reported dichotomous process data,

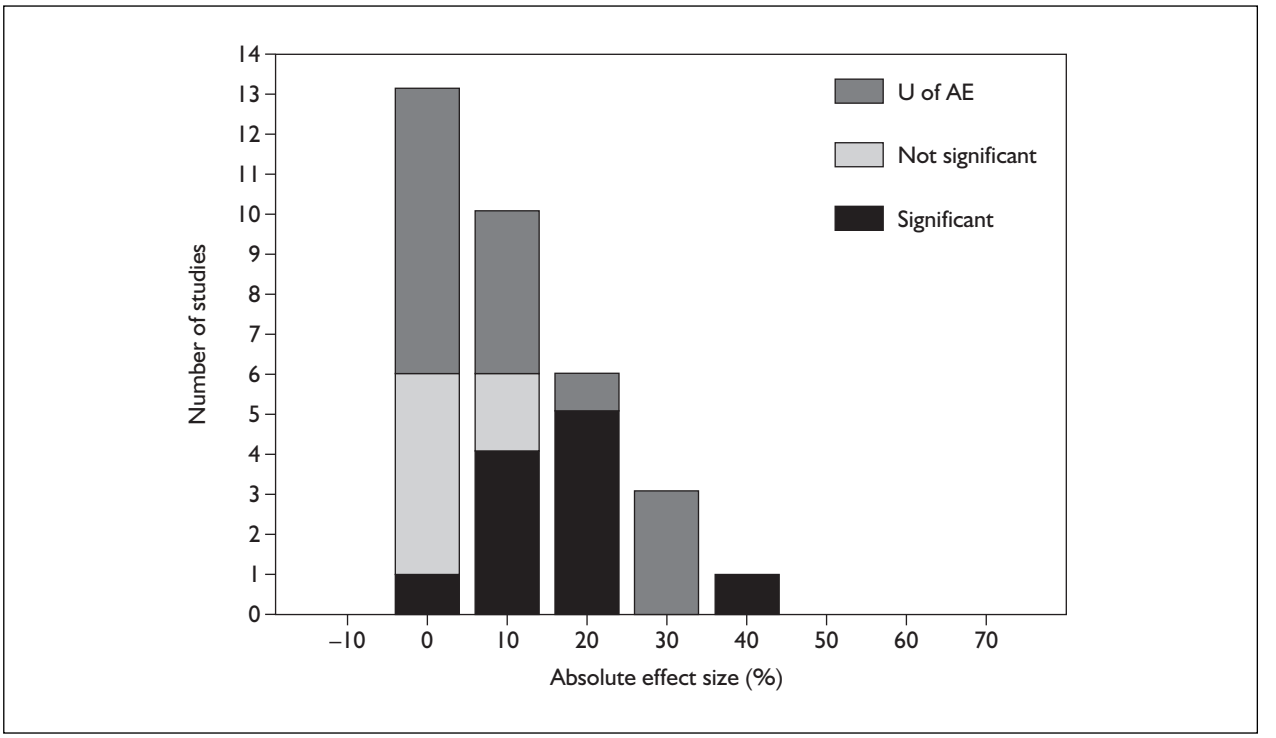

FIGURE 5 *Effect sizes observed in dichotomous process data in RCT, CCT and CBA comparisons of reminders against a no-intervention control*

including 15 C-RCT comparisons (A10, A38, A40, A47, A81, A111, A119, A128, A134, A135, A156, A179, A180, A182, A209), eight P-RCT comparisons (A13, A16, A34, A37, A98, A133, A204, A207), eight CCT comparisons (A46, A90, A112, A142, A222, A224, A225, A226) and two CBA comparisons (A101, A187).

Twelve of 14 C-RCT comparisons reported improvements in care; the median effect was +14.1% (range –1.0 to +34.0%) absolute improvement in performance. Eleven comparisons had potential unit of analysis errors; the remaining three comparisons (A119, A134, A135) observed a median effect of +20.0% (range +13 to +20%); all were significant. Comparable data could not be abstracted for one C-RCT comparison (A10) that reported no significant changes in overall compliance and had a potential unit of analysis error.

Seven of eight P-RCT comparisons reported improvements in care; the median effect was +5.4% (range –1 to +25.7%). Three comparisons were statistically significant (A13, A37, A207) and one study had a potential unit of analysis error (A133).

One C-CCT (A46) reported an absolute improvement in care of +4.3% but had a potential unit of analysis error. Six of the seven P-CCT comparisons reported improvements in care; the median effect was +10.0% (range 0 to +40.0%) absolute improvement in care. Four comparisons were statistically significant (A90, A142, A222, A224).

Two CBA comparisons (A101, A187) observed effects of +3.6% and +10% absolute improvements in performance; both had potential unit of analysis errors.

Continuous process measures. Ten comparisons reported continuous process data, including two C-RCT comparisons (A146, A182), one P-RCT comparison (A207), one C-CCT (A189) and six P-CCT comparisons (A90, A112, A223, A224, A225, A226). One C-RCT comparison (A182) observed a relative improvement in performance of +16.7%; however, the SMD could not be calculated and there was a potential unit of analysis error. Comparable data could not be abstracted for the other C-randomised comparison (A146). The P-RCT comparison observed a relative deterioration in performance of –3.3% and –0.28 SMD; however, the significance of the comparison could not be ascertained.

The C-CCT (A189) observed a relative improvement in performance of +32.0% and an

SMD of +0.15; however, this was not significant. Five out of six P-CCT comparisons observed improvements in performance; the median effect was +5.7% (range –41.8 to +36.0%) relative improvement in performance. SMDs were calculable for six comparisons. Improvements were observed in five of these comparisons; the median effect was +0.11 (range –0.81 to +0.22). Three comparisons were statistically significant, one observed a relative deterioration in performance of –41.8% (SMD –0.81) (A225) the other two observed relative improvements in performance of +25.7% (SMD +0.22) (A224) and +36% (SMD not calculable) (A223).

Dichotomous outcome measures. Four comparisons reported dichotomous outcome measures. One C-CRT (A182) observed an absolute improvement of 3% but had a potential unit of analysis error. Three CCT comparisons reporting dichotomous outcome measures (A90, A225, A224) observed improvements; the median effect was +1.9% (range +1.0 to +6.8%) absolute improvement in performance. None of the effects was statistically significant.

Continuous outcome measures. Four comparisons reported continuous outcome data, including one P-RCT comparison (A42) and three CCT comparisons (A90, A225, A224). The P-RCT comparison was designed to test equivalence and suggested similar results for computer- and physician-managed patients; however, there was a potential unit of analysis error. It was not possible to abstract comparable data about the effect size. Two of the three CCT comparisons observed improvements; the median effect was +1.9% (range –2.0 to +7.2%) relative improvement in performance. The median SMD was +0.11 (range –0.07 to + 0.11). None of the studies was statistically significant.

Results from ITS comparisons

One comparison (A12) was reanalysed using time series regression and observed positive statistically significant changes in both the level and slope.

Summary

Reminders were the most frequently evaluated single intervention. The results were moderate. Improvements in dichotomous process measures were observed in 28 out of 33 comparisons across all designs. Across all C-RCTs, the median effect size was +14.1% (range –1.0 to +34.0%). The statistical significance of three C-RCT process dichotomous comparisons could be determined; all were significant. Across all P-RCTs, the median effect was +5.4% (range –1 to +25.7%); three of eight comparisons were statistically significant. Again, the difference in effect sizes between C-RCTs and P-RCTs suggests that there may be a contamination effect in P-RCTs (and P-CCTs). One C-CCT observed an improvement of +4.3% but had a potential unit of analysis error. Across all P-CCTs, the median effect was +10.0% (range +0 to +40%); four of seven comparisons were statistically significant. Two CBA comparisons observed effects of +3.6% and +10%; both had unit of analysis errors. Reminders have been tested across a wide range of targeted behaviours and in a wide range of settings (although none of the studies was based in UK settings). These results suggest that reminders may have a moderate effect on guideline implementation.

Other professional interventions

Three other studies (involving three comparisons) evaluated a professional intervention versus a control receiving no intervention or usual care, including two C-RCTs (A227, A86) and one P-RCT (A80).

One study was conducted in the USA, one in Canada and one in the UK. Two were based in primary care settings, the other in an inpatient setting. General management was the targeted behaviour in two studies and referral in the third study.

One C-RCT comparison (A227) evaluated the provision of nicotine replacement gum to physicians and observed a +39.1% absolute improvement in performance (percentage of patients reporting that physicians mentioned smoking) and +0.5% absolute improvement in outcome (mean percentage per practice ceased smoking, 1 year prevalence). There was a potential unit of analysis error in the dichotomous process analysis and the significance of the dichotomous outcome analysis could not be determined.

The other C-RCT comparison (A86) evaluated the effects of interviewing GPs about outpatient referrals. Comparable data could not be abstracted. The author states that the intervention had no effect, there was a potential unit of analysis error.

The P-RCT comparison (A80) evaluated the effects of a rapid rule-out protocol in the management of chest pain and observed a +71.8% relative improvement in performance and an SMD of +0.31 (statistically significant).

Organisational interventions

Two studies evaluated the effects of organisational interventions, including one C-RCT (A181) and one CBA (A93). One of the studies took place in the USA, the other in The Netherlands. One was based in a primary care setting, the other in an inpatient setting. General management was the targeted behaviour in one of the studies and prescribing in the other study.

The C-RCT comparison (A181) evaluated the effects of continuity of care in diabetes in primary care. It observed a +2.1% relative improvement in the level of glycosylated haemoglobin; the SMD could not be calculated and the study had a potential unit of analysis error.

The CBA comparison (A93) evaluated the effects of clinical pharmacy services (revision of professional role) on prescribing in an orthopaedic unit. They observed a –16.2% relative deterioration in performance and an SMD of –0.21; however, the study had a potential unit of analysis error.

Evaluations of single interventions against intervention control groups

Three studies involving three comparisons evaluated single interventions against control groups that received interventions, including one C-RCT comparison (A118) and two CBA comparisons (A78, A210). Two studies were based in the USA and one in Canada. Two were set in ambulatory care settings and one in an inpatient setting. Two targeted general management of a problem and the third prescribing.

The C-RCT comparison (A118) evaluated whether requiring physicians to respond to computer reminders increased compliance compared with reminders alone. It observed 7% absolute improvement in the study group; however, there was a potential unit of analysis error.

One CBA comparison (A78) compared educational materials and reminders (control group). There was a –67% absolute deterioration in care in the study (educational materials) group; the study had a potential unit of analysis error.

The other CBA comparison (A210) compared reminders and patient-directed intervention (control). There was a +5.6% absolute improvement in care in the study (reminders) group; the study had a potential unit of analysis error.

Evaluations of multifaceted interventions

A total of 178 studies (including 222 comparisons) evaluated multifaceted interventions: 137 comparisons (44.3%) evaluated 68 different combinations compared with a control group that did not receive an intervention, and 85 (27.5%) comparisons evaluated 55 different combinations compared with a control group that also received an intervention. The maximum number of comparisons of the same combination of interventions against a no-intervention control group was 11. The maximum number of comparisons of interventions against control group that received an intervention was six. This presented considerable difficulties for the synthesis and interpretation of these studies. It was planned originally to undertake a meta-regression analysis to estimate the effects of different interventions, but the number of multifaceted studies proved problematic. Meta-regression allowing for interaction effects between interventions was not possible owing to the large number of additional variables required and because combinations of some interventions were highly correlated (see Appendix 1 for further details). This section describes the results of comparisons of multifaceted interventions against a control group that did not receive interventions (no-intervention controls) and against control groups that did receive interventions and summarises the results for different combinations of interventions with more than five comparisons. Finally, the study tested whether the effectiveness of multifaceted interventions increases with the number of interventions.

Multifaceted interventions against no-intervention controls

A total of 117 studies (including 136 comparisons) evaluated 68 different combinations of interventions (including 26 combinations of two interventions, 19 combinations of three interventions, 16 combinations of four interventions and seven interventions of five or more interventions) (*Table 6*). The studies included 46 C-RCTs (A1, A9, A17, A22, A23, A24, A28, A31, A32, A39, A48, A51, A55, A56, A59, A64, A65, A85, A95, A96, A100, A102, A106, A111, A116, A117, A129, A137, A138, A141, A144, A149, A154, A158, A168, A183, A186, A192, A194, A196, A202, A208, A209, A219, A227, A235), 13 P-RCTs (A3, A16, A18, A21, A34, A53, A76, A103, A109, A123,

TABLE 6 *Summary of comparisons for multifaceted interventions against a no-intervention control group*

No. of interventions in multifaceted interventions	No. of different combinations of multifaceted interventions	No. of comparisons of multifaceted interventions	Maximum no. of comparisons relating to specific combination of interventions
2	26	62	11
3	19	44	8
4	16	21	3
5	4	6	3
6	2	2	1
7	1	1	
Total	68	136	

A145, A163, A177), one C-CCT (A97), 27 CBAs (A5, A27, A49, A50, A54, A66, A69, A71, A74, A77, A83, A101, A110, A115, A127, A147, A160, A161, A167, A171, A172, A173, A175, A176, A178, A213, A216) and 30 ITS (A8, A11, A25, A26, A29, A41, A52, A58, A63, A68, A70, A73, A82, A87, A88, A107, A113, A124, A143, A150, A159, A164, A190, A195, A200, A201, A203, A217, A218, A231).

Eighty-two studies were undertaken in the USA, 16 in the UK, six in Canada, five in Australia and the remaining eight in France, Mexico, New Zealand, The Netherlands, Norway, Sweden and Thailand. Forty-four studies were based within primary care settings, 25 in mixed settings, 20 in inpatient settings, 14 in generalist outpatient or ambulatory care settings and 12 in nursing home or long-term care facilities, accident and emergency settings, specialist outpatient clinics and a military medical centre. The setting was not clear in two studies. The targeted behaviour was general management of a problem in 45 studies, prevention in 30 studies, prescribing in 21 studies, test ordering in 12 studies and procedures, financial management and referral in the nine remaining studies.

Results from RCT, CCT and CBA comparisons

Dichotomous process measures. Seventy-eight comparisons (derived from 62 studies) evaluating 49 different combinations of interventions reported dichotomous process data, including 41 C-RCT comparisons (A1 – two comparisons, A9, A17 – two comparisons, A23, A28 – two comparisons, A31, A32 – two comparisons, A102, A55 – three comparisons, A48, A51, A56, A64, A65, A95, A111, A116, A117, A129 – two comparisons, A137, A141, A144, A154 – two comparisons, A168, A183, A186, A194, A196, A209, A227, A235, A208, A39), ten P-RCT comparisons (A3, A16, A18 – two comparisons, A34, A76, A103, A123, A145, A177), one CCT comparison (A97), 26 CBA comparisons (A5, A49, A50, A54, A66, A69 – four comparisons, A71, A101, A110, A115, A147, A160, A172, A171, A173, A175, A176 – four comparisons, A213, two comparisons, A216). The majority of comparisons evaluated combinations of two (n = 36), three (n = 25) or four (n = 10) interventions. There were few replications of evaluations for different combinations of interventions.

Table 7 summarises the median effect size of the absolute improvement in performance across studies for each combination of interventions.

Comparable data could not be abstracted for one C-RCT (A208) and one CBA (A110). The C-RCT comparison evaluated a combination of educational materials, educational meetings and organisational interventions and observed significant improvements in care. The CBA comparison evaluated a combination of educational materials, educational meetings, audit and feedback, and educational outreach and observed a significant improvement in the use of antibiotic prophylaxis in surgery.

Continuous process measures. Thirty comparisons (derived from 27 studies) reported continuous process data, including 18 C-RCT comparisons (A9, A22, A24, A56, A59, A85, A96, A100, A116, A129 – two comparisons, A138, A149, A158, A186, A192, A202, A208), three P-RCT comparisons (A109, A145, A177), one CCT (cluster) comparison (A97) and eight CBA comparisons (A27, A74, A77, A83 – three comparisons, A161, A178). Fourteen comparisons evaluated combinations of two interventions, five comparisons evaluated combinations of three interventions, four comparisons evaluated combinations of four interventions and one comparison evaluated a combination of five interventions. *Table 8* summarises the median effect size of the relative

TABLE 7 *Summary of comparisons of multifaceted interventions compared to no-intervention control groups reporting dichotomous process data*

Combination of interventions	No. of comparisons	Median % absolute difference across studies	Range % absolute difference across studies	Study IDs
Two intervention combinations				
A&F, Fin	1	6.7	NA	A95
A&F, LCP	1	–9	NA	A194
A&F, Outreach	2	13.7	+10 to +17.4	A55, A64
A&F, Profoth	1	68	NA	A123
Edmat, A&F	2	7.4	+7.0 to +7.8	A69, A176
Edmat, Edmeet	5	3.0	–3.0 to +10	A28, A55, A97, A129, A154
Edmat, Org	1	12.1	NA	A103
Edmat, Outreach	6	1.3	–5.6 to +13.1	A48, A49, A56, A154, A172, A171
Edmat, Patmed	1	7.8	NA	A76
Edmat, Profoth	1	12	NA	A168
Edmat, Rem	1	1.1	NA	A69
Edmat, Org	1	5.0	NA	A51
Edmeet, Outreach	1	7.0	NA	A144
Edmeet, Rem	1	15.0	NA	A175
Org, Struc	1	45.0	NA	A18
Outreach, Org	1	2.2	NA	A186
Rem, A&F	2	9.2	+2.7 to +15.7	A32, A209
Rem, Patmed	5	17.0	+1.3 to +25.1	A16, A31, A34, A111, A147
Rem, Struc	2	12.5	+8.0 to + 17.0	A177, A235
Three intervention combinations				
Edmat, A&F, LCP	1	18.0	NA	A176
Edmat, Edmeet, A&F	3	43.0	+2.6 to +9.0	A23, A39, A129
Edmat, Edmeet, Org	3	1.0	+0.4 to +6.3	A115, A213, A213
Edmat, Edmeet, Outreach	3	11.0	+8.4 to +16.4	A65, A173, A196
Edmat, Edmeet, Profoth	1	60.0	NA	A227
Edmat, Org, Struc	1	2.2	NA	A69
Edmat, Rem, A&F	2	26.0	+25.0 to + 27.0	A176, A176
Edmat, Rem, Patmed	2	9.5	+5.6 to +13.4	A117, A137
Edmat, Rem, Struc	1	11.1	NA	A69
Edmeet, A&F, LCP	2	24.0	+16.8 to +33.2	A102, A160
Edmeet, Rem, A&F	1	–2.0	NA	A18
Edmeet, Rem, Patmed	2	23.5	+20.0 to +27.0	A17, A17
Rem, A&F, Patmed	1	16.9	NA	A32
Rem, A&F, Struc	1	1.5	NA	A183
Four intervention combinations				
Edmat, Edmeet, A&F, Org	1	2.0	NA	A116
Edmat, Edmeet, A&F, Outreach	1	6.0	NA	A55
Edmat, Edmeet, A&F, Struc	1	0.3	NA	A216
Edmat, Edmeet, Profoth, Org	1	10.0	NA	A50
Edmat, Edmeet, LCP, Profoth	1	6.0	NA	A71
Edmat, Edmeet, Outreach, Patmed	1	–4.0	NA	A28
Edmat, Edmeet, Outreach, Profoth	1	2.2	NA	A9
Edmat, Edmeet, Rem, Org	1	4.0	NA	A1
Edmat, Rem, Outreach, Patmed	1	–2.0	NA	A141
Edmat, Rem, Patmed, Org, Struc	1	24.0	NA	A3
Rem, Fin, Org, Struc	1	54.0	NA	A145
Rem, Patmed, Org, Struc	1	8.9	NA	A54

continued

TABLE 7 *Summary of comparisons of multifaceted interventions compared to no-intervention control groups reporting dichotomous process data (cont'd)*

Combination of interventions	No. of comparisons	Median % absolute difference across studies	Range % absolute difference across studies	Study IDs
Five intervention combinations				
Edmat, Edmeet, Patmed, Org, Struc	1	24.3	NA	A101
Edmat, Edmeet, Rem, A&F, Org	1	–2.0	NA	A1
Edmat, Edmeet, Rem, Org, Struc	1	19.6	NA	A5
Six intervention combinations				
Edmat, Edmeet, A&F, Patmed, Profoth, Fin	1	15.0	NA	A66

A&F: audit and feedback; Fin: financial; LCP: local consensus processes; Outreach: educational outreach; Profoth: other professional; Edmat: educational materials; Edmeet: educational meetings; Org: organisational; Patmed: patient mediated; Rem: reminders; Struc: structural; NA: not applicable.

TABLE 8 *Summary of comparisons of multifaceted interventions compared to no-intervention control groups reporting continuous process data*

Combination of interventions	No. of comparisons	Relative % difference (range)	SMD (range)	Study IDs
Two intervention combinations				
A&F, Profoth	1	–17.2	NC	A149
Edmat, A&F	2	+9.1 (0 to +18.1)	0.23	A178, A192
Edmat, Edmeet	4	+18.75 (+1.2 to +40.2)	0.09 (+0.03 to +0.15)	A22, A97, A109, A129
Edmat, LCP	1	–38.6	NC	A100
Edmat, Outreach	2	+15.7 (+11.3 to +20.0)	NC	A56, A161
Edmeet, Org	1	+245.0	NC	A202
Org, Outreach	1	+5.5	NC	A186
Rem, Struc	2	+130.9 (+81 to +180.7)	NC	A27, A177
Three intervention combinations				
Edmat, Edmeet, A&F	3	+19.6 (–13.1 to +400.0)	0.12 (–2.1 to +0.45)	A74, A129, A158
Edmat, Edmeet, Outreach	1	+24.0	NC	A138
Four intervention combinations				
Edmat, Edmeet, A&F, Org	1	+32.2	NC	A116
Edmat, Edmeet, Outreach, Profoth	1	+15.0	NC	A9
Edmat, Edmeet, Rem, Patmed	1	+23.1	1.5	A24
Edmeet, Outreach, OL, Profoth	1	+1.7	NC	A59
Rem, Fin, Org, Struc	1	+7.6	4	A145
Five intervention combinations				
Edmat, Edmeet, Rem, Patmed, Org	1	NC	2.5	A77

OL, opinion leaders; NC, not clear.

TABLE 9 *Summary of comparisons of multifaceted interventions compared to no-intervention control groups reporting dichotomous outcome data*

Combination of interventions	No. of comparisons	Median % absolute difference across studies	Range % absolute difference across studies	Study IDs
Two intervention combinations				
A&F, Outreach	1	3	NA	A64
Edmat, Edmeet	2	11.5	10.0 to 13.0	A97, A109
Edmeet, Outreach	1	–1	NA	A144
Rem, Org	1	21	NA	A21
Rem, Struc	1	8	NA	A177
Three intervention combinations				
Edmat, Edmeet, A&F	1	0.8	NA	A127
Edmat, Edmeet, Profoth	1	3.8	NA	A227
Edmeet, Org, Profoth	1	–8.3	NA	A163
Rem, A&F, Struc	1	–2.2	NA	A183
Four intervention combinations				
Edmat, Edmeet, Outreach, Profoth	1	–7	NA	A9
Edmat, Rem, Outreach, Patmed	1	–7	NA	A141
Edmeet, Outreach, OL, Profoth	1	0.1	NA	A59
Six intervention combinations				
Edmat, Edmeet, A&F, Patmed, Profoth, Fin	1	3	NA	A66

improvement in performance and SMD across comparisons for each combination of interventions.

Comparable data could not be abstracted for six comparisons (A83 – three comparisons, A85, A96, A208). Two C-RCT comparisons (A85, A96) observed no significant effects. The third C-RCT (A208) observed a significant reduction in waiting times following referral. One CBA reported three comparisons (A83). Two comparisons evaluating combinations of educational materials and meetings observed significant increases in peak flow monitoring. The third comparison, evaluating a combination of educational materials, educational meetings and reminders, observed a significant increase in oral B_2 antagonist use.

Dichotomous outcome measures. Fifteen comparisons reported dichotomous outcome data, including eight C-RCT comparisons (A9, A59, A64, A117, A141, A144, A183, A227), four P-RCT comparisons (A21, A109, A163, A177), one CCT comparison (A97) and two CBA comparisons (A66, A127). Six studies evaluated five combinations of two interventions, four studies evaluated four combinations of three interventions, three studies evaluated three combinations of four interventions and one study evaluated one combination of six interventions. *Table 9* summarises the median effect size of the absolute improvement in performance across comparisons for each combination of interventions. Comparable data were not available for one C-RCT comparison (A117) that observed improvements in outcome.

Continuous outcome measures. Sixteen comparisons (derived from 15 studies) reported continuous outcome data, including 11 C-RCT comparisons (A24, A59, A64, A85, A106, A116, A141, A219 – two comparisons, A235, A208), four P-RCT comparisons (A53, A145, A163, A177) and one CBA comparison (A173). Five comparisons evaluated four combinations of two interventions, one comparison evaluated one combination of three interventions, five comparisons evaluated four combinations of four interventions and two comparisons evaluated two combinations with five or more interventions. *Table 10* summarises the median effect size of the relative improvement in outcome and SMD across comparisons for each combination of interventions.

Comparable data could not be abstracted for three C-RCT comparisons (A85, A106, A208). All three comparisons observed no significant effects of the intervention.

Results from ITS comparisons

Thirty comparisons evaluated the effects of multifaceted interventions (A8, A11, A25, A26, A29, A41, A52, A58, A63, A70, A68, A73, A82,

TABLE 10 *Summary of comparisons of multifaceted interventions compared to no-intervention control groups reporting continuous outcome data*

Combination of interventions	No. of comparisons	Relative % difference (range)	SMD (range)	Study IDs
Two intervention combinations				
A&F, Outreach	1	0	0	A64
Rem, Struc	2	–0.4 (–3.0 to +2.2)	0.115 (–0.13 to +0.36)	A235, A177
Patmed, Profoth	1	32		A53
Edmeet, Org	1	19.4		A163
Three intervention combinations				
Edmat, Edmeet, Outreach	1	13.9	2.375	A173
Four intervention combinations				
Edmeet, Outreach, OL, Profoth	1	2.7		A59
Edmat, Edmeet, Rem, Patmed	1	9.3	6.0	A24
Rem, Fin, Org, Struc	1	2		A145
Edmat, Rem, Outreach, Patmed	1	–1.4		A141
Edmat, Edmeet, A&F, Org	1	0.1		A116
Five or more intervention combinations				
Edmat, Edmeet, Rem, A&F, Org	1	0.9	0.003	A219
Edmat, Edmeet, Rem, A&F, Patmed, Profoth, Org	1	3	0.1	A219

A87, A88, A107, A113, A124, A143, A150, A159, A164, A190, A195, A200, A201, A203, A217, A218, A231). Nineteen comparisons were reanalysed using time series regressions (A8, A11, A25, A41, A52, A70, A73, A82, A88, A113, A124, A143, A150, A164, A190, A201, A203, A218, A231) and comparable data were available from three comparisons (A29, A195, A217). Nineteen of 21 comparisons observed improvements in changes on level, nine comparisons were statistically significant (A11, A25, A29, A70, A82, A88, A150, A217, A231) (*Table 11*). Fourteen of 21 comparisons observed improvements in changes in slope; four comparisons were significant (A25, A29, A124, A190). One of seven comparisons observing deterioration in change in slope was statistically significant (A195).

Comparable data could not be abstracted and reanalysis undertaken for eight studies (A26, A58, A63, A68, A87, A107, A159, A200). One study (A200) presented data in control charts and reported significant improvements in care, but no quantification was given.

Additional analyses

A number of comparisons evaluated discrete combinations of interventions (e.g. 11 comparisons evaluated combinations of educational materials and educational meetings). In this section, the results of combinations of interventions that had more than four comparisons are summarised. Twenty-three comparisons evaluating multifaceted interventions including educational outreach were identified. Six additional combinations of multifaceted interventions had more than four comparisons, including: educational materials plus educational meetings; educational materials plus audit and feedback, reminders and patient-directed interventions; educational materials, educational meetings and audit and feedback; educational materials, educational meetings and audit and feedback; and educational materials, educational meetings and organisational interventions.

Multifaceted interventions incorporating educational outreach

Twenty-two studies (involving 23 comparisons) evaluated the multifaceted interventions including educational outreach against a no-intervention control, including 14 C-RCTs (A9, A28, A48, A55 – two comparisons, A56, A59, A64, A65, A138, A141, A144, A154, A186, A195), six CBAs (A49, A110, A161, A171, A172, A173) and two ITS (A195, A200).

Eleven different multifaceted interventions incorporating educational outreach were evaluated, including four combinations of two interventions, one combination of three interventions, five combinations of four interventions and one

TABLE 11 *Summary of results of ITS comparisons of multifaceted interventions*

Study ID	Combination of interventions	Study reanalysed	Change in level	Significance	Change in slope	Significance
Two intervention combinations						
A164	A&F, Org	Yes	+	NS	+	NS
A8	Edmat, Edmeet	Yes	+	NS	+	NS
A41	Edmat, Edmeet	Yes	+	NS	–	NS
A203	Edmat, Org	Yes	+	NS	+	NS
A150	Edmat, Org	Yes	+	S	–	NS
A195	Edmat, Outreach	No			–	S
A201	Edmat, Reg	Yes	+	NS	+	NS
A190	Legislation, Fin	Yes	–	NS	+	S
A218	Rem, Patmed	Yes	+	NS	–	NS
A88	Rem, Org	Yes	+	S	+	NS
Three intervention combinations						
A29	Edmat, A&F, Org	No	+	S	+	S
A11	Edmat, Edmeet, A&F	Yes	+	S	+	NS
A70	Edmat, Edmeet, A&F	Yes	+	S	–	NS
A217	Edmat, Fin, Org	No	+	S		
A25	Edmat, Fin, Org	Yes	+	S	+	S
A124	Edmat, Mass media, Org	Yes	+	NS	+	S
A122	Edmat, Rem, Patmed	Yes	+	NS	+	NS
A82	Edmeet, A&F, LCP	Yes	+	S	+	NS
A73	Edmeet, A&F, Org	Yes	+	NS	+	NS
Four intervention combinations						
A231	Edmat, Edmeet, Rem, Org	Yes	+	S	–	NS
A143	Edmeet, Rem, A&F, Org	Yes	–	NS	+	NS
Five intervention combinations						
A52	Edmat, Edmeet, Rem, A&F, Org	Yes	+	NS	–	NS

Comparable data were only available for 22 comparisons.
Reg, regulatory; NS, not significant; S, significant.

combination of six interventions. Common combinations included educational materials and educational outreach (eight comparisons), educational materials, educational meetings and educational outreach (four comparisons), audit and feedback and educational outreach (two comparisons) and educational materials, educational meetings, audit and feedback and educational outreach (two comparisons).

Twelve studies were conducted in the USA, four in the UK, four in Australia and two in Sweden. Twelve studies were based in primary care settings and five in nursing homes, the remaining studies were based in mixed settings (two) and inpatient settings (two), and in one the setting was not clear. Ten studies targeted prescribing, nine targeted general management of a clinical problem, two targeted prevention and one targeted procedures.

Results from RCT, CCT and CBA comparisons

Dichotomous process measures. Eighteen comparisons reported dichotomous process data, including 13 C-RCT comparisons (A9, A28, A55 – two comparisons, A48, A56, A64, A65, A141, A144, A154, A186, A196) and five CBA comparisons (A49, A110, A171, A172, A173).

Eleven of the C-RCT comparisons observed improvements in performance; the median effect was +6.0% (range –4 to +17.4%) absolute improvement in performance. Statistical significance could be determined for five comparisons (A28, A55, A56, A64); the median effect size across these studies was +10.0% (range –4 to +17.4%) absolute improvement in performance (only one study observing a +17.4% absolute improvement in performance was statistically significant, A64). Seven studies had potential unit of analysis errors (A48, A65, A141, A144, A154, A186, A196) and the significance of the postintervention comparison could not be determined in one study (A9).

Two of four CBA comparisons (A49, A171, A172, A173) reporting dichotomous process of care

results observed positive improvements in performance. Across all studies the median effect was +7.3 (range –5.6 to + 16.4%) absolute improvement in performance. All these studies had potential unit of analysis errors. Comparable data could not be abstracted for one study (A110) that reported significant improvements in use of antibiotic prophylaxis in surgery.

Continuous process measures. Six comparisons reported continuous process data, including five C-RCT comparisons (A9, A56, A59, A138, A186) and one CBA comparison (A161). All C-RCT comparisons observed improvements in performance; the median effect was +15.0% (range +1.7 to +24.0%) relative improvement in performance. No studies reported sufficient data to calculate SMDs. One study had a potential unit of analysis error (A186) and the significance of the postintervention comparison could not be determined in two studies (A9, A56). The remaining two studies were not statistically significant.

One CBA (A161) observed 11.3% relative improvement in performance. The SMD could not be calculated and the study had a potential unit of analysis error.

Dichotomous outcome measures. Five C-RCT comparisons reported dichotomous outcome data (A9, A59, A64, A141, A144); the median effect was –1.0% (range –7.0 to +3.0%) absolute improvement in outcome. Only two comparisons observed improvements in outcome. Three studies had potential unit of analysis errors (A9, A64, A141) and the significance of one comparison could not be determined (A144). The remaining comparison was not significant (A59).

Continuous outcome measures. Four comparisons reported continuous outcome data, including three C-RCT comparisons (A59, A64, A141) and one CBA comparison (A173). The median effect of the C-RCT comparisons was 0% (range –1.4 to 2.7%). The SMD could be calculated for one comparison and was 0 (A64). Two comparisons had potential unit of analysis errors (A64, A141) and the third comparison was not significant (A59). One CBA comparison (A173) observed a +13.9% relative improvement in outcome and an SMD of +2.8; however, the study had a potential unit of analysis error.

Results from ITS comparisons

Two studies reported continuous process data (A195, A200). One study reported statistically significant improvements in performance (A195) but could not be reanalysed. The second study (A200) displayed results as control charts; significant changes were reported but no quantification was given.

Additional analyses

Exploratory subgroup analyses were undertaken of comparisons evaluating combinations of educational materials and educational outreach (eight comparisons) and educational materials, educational meetings and educational outreach (four comparisons).

Educational materials and educational outreach

Eight studies evaluated the effects of educational materials and educational outreach visits against a no-intervention control, including three C-RCTs (A48, A56, A154), four CBAs (A49, A161, A171, A172) and one ITS (A195). Six comparisons reported dichotomous process data (including three C-RCT comparisons – A48, A56, A154 and three CBA comparisons – A49, A171, A172). Improvements in care were observed in four comparisons; the median effect size was +1.2% (range –5.6 to +13.1%) absolute improvement in performance. Five comparisons had potential unit of analysis errors and the remaining study was non-significant (A56). Both studies (A56, A161) reporting continuous outcome data observed improvement in care but neither reported sufficient data to calculate a standardised mean effect size and the significance of the comparisons was uncertain. One ITS comparison reported significant improvements in performance but could not be reanalysed (A195).

Educational materials, educational meetings and educational outreach

Four studies evaluated the effects of educational materials, educational meetings and educational outreach visits against a no-intervention control, including three C-RCTs (A65, A138, A196) and one CBA (A173). Three comparisons reported dichotomous process data (including two C-RCT comparisons – A65, A196 and one CBA comparisons – A173). Improvements in care were observed in all comparisons, the median effect size was +11.0% (range +8.4 to +16.4%) absolute improvement in performance. All comparisons had potential unit of analysis errors. One C-RCT comparison reported continuous process data and observed a non-significant +24.0% relative improvement in performance; insufficient data were provided to calculate an SMD (A138). One CBA comparison reported continuous outcome data observing a 13.9% relative improvement and

an SMD of +2.38; the comparison had a potential unit of analysis error (A173).

Relationship between number of interventions and effect size

Exploratory analyses were undertaken to determine whether the number of interventions influenced the effectiveness of multifaceted interventions including educational outreach compared to a no-intervention control group using dichotomous process data. The median effect of combinations of two interventions was +4.6%, three interventions +11.0% and four interventions +0.1% (*Table 12*).

Summary

The majority of studies evaluating multiple educational outreach against a no-intervention control observed absolute improvements in performance of care. However, the effects were modest, with absolute improvements of +6.0% (range –4 to +17.4%) across 11 C-RCTs and +7.3% (range –5.6 to +17.4%) across four CBAs. The statistical significance of five studies could be calculated, and only one was significant. No studies reporting continuous process measures provided sufficient information to calculate SMDs. Exploratory subgroup analyses suggested that:

- combinations of educational materials and educational outreach may be relatively ineffective
- combinations of educational materials, educational meetings and educational outreach may have modest to moderate effects
- there was no clear relationship between number of interventions and effect size.

Four studies were conducted in the UK. Ten studies targeted prescribing behaviours. These results suggest that multiple intervention strategies including educational outreach may have a modest effect on guideline implementation, especially when targeting prescribing behaviours.

Combinations of educational materials and educational meetings

Ten studies (involving 11 comparisons) evaluated combinations of educational materials and educational meetings, including five C-RCTs (A22, A28, A55, A129, A154), one P-RCT (A109), one C-CCT (A97), one CBA (A83 – two comparisons) and two ITS (A8, A41).

Seven studies (eight comparisons) were conducted in the USA, two in the UK and one in France. Five studies were based in primary care settings. The targeted behaviour was general management of a problem in seven studies (eight comparisons).

Results from RCT, CCT and CBA comparisons

Dichotomous process measures. Seven comparisons reported dichotomous process data, including four C-RCT comparisons (A28, A55, A129, A154), one C-CCT comparison (A97) and two CBA comparisons (A83). Three C-RCT comparisons observed improvements in care; the median effect was +1.9% (range –3.0 to +5.0%) absolute improvement in care. Two comparisons had potential unit of analysis errors (A129, A154) and the remaining two comparisons were non-significant. The C-CCT comparison observed an absolute improvement of +10%; however, the study had a potential unit of analysis error. Comparable data could not be abstracted for the

TABLE 12 *Effectiveness of multifaceted interventions including educational outreach by number of interventions*

No. of interventions	Dichotomous process data (% absolute improvement in performance)
Two	
No. of studies	10
Median effect	+4.6
Range	–5.6 to +17.4
Three	
No. of studies	3
Median effect	+11.0
Range	+8.4 to +16.4
Four	
No. of studies	4
Median effect	+0.1
Range	–4.0 to +6.0

CBA comparisons; both reported significant increases in peak flow monitoring (A83).

Continuous process data. Four comparisons reported continuous process data, including two C-RCT comparisons (A22, A129), one P-RCT (A109) and one C-CCT comparison (A97). The two C-RCT comparisons observed relative improvements in performance of +1.2% (A129) and 33.8% (A22). An SMD could only be calculated for one comparison, +0.03 (A129). One comparison (A129) had a potential unit of analysis error and the postintervention significance of the other comparison could not be determined. The P-RCT comparison observed a non-significant improvement of 3.7%, but the SMD could not be calculated. The C-CCT comparison (A97) observed a relative improvement of +40.2%, with an SMD of +0.15; however, the study had a potential unit of analysis error.

Dichotomous outcome data. One C-CCT comparison (A97) observed an increase of +13.0% but had a potential unit of analysis error.

Results from ITS comparisons
Two ITSs (A8, A41) observed no significant improvements in care.

Summary
All comparisons observed small to modest improvements in process of care. The effects were small (absolute improvement across four C-RCT comparisons +1.9%, range –3.0 to +5.0%) to modest (one C-CCT comparison +10%). Two dichotomous process C-RCT comparisons were non-significant. Two dichotomous process CBA comparisons reported statistically significant results. One dichotomous process P-RCT comparison was non-significant. Two ITS comparisons were non-significant. Only two studies were conducted in UK settings. These results suggest that educational materials and educational meetings in combination may have, at best, a small effect on guideline implementation. However, the evidence base is sparse and of poor quality.

Combinations of educational materials and audit and feedback

Four studies evaluated the effects of combinations of educational materials and audit and feedback, including one C-RCT comparison (A192) and three CBA comparisons (A69, A176, A178). Two studies were based in the UK, one in Norway and one in the USA. Two studies were based in inpatient settings, one in family or general practice and one in mixed settings. General management was the targeted behaviour for two studies, whereas test ordering and prescribing were the behaviours in the other two studies.

Results from RCT, CCT and CBA comparisons
Dichotomous process data. Two CBA comparisons reported dichotomous process data and observed improvements of +7% (A176) and +7.8% (A69); both studies had potential unit of analysis errors.

Continuous process data. One C-RCT comparison (A192) observed a relative improvement of +18.1% and an SMD of 0.23. One CBA comparison observed no change (0%) in performance; the SMD could not be calculated. Both studies had potential unit of analysis errors.

Summary
There were relatively few comparisons of educational materials and audit and feedback. Three of four comparisons observed improvements in care. The effects were modest (median absolute improvement across two C-RCT comparisons of 7.4%). These results suggest that effects are likely to be modest.

Combinations of reminders and patient-directed interventions

Six studies evaluated combinations of reminders and patient-directed interventions, including two C-RCT comparisons (A31, A111), two P-RCT comparisons (A16, A34), one CBA comparison (A147) and one ITS comparison (A218). All studies were conducted in the USA. Three studies were based in primary care settings and three in ambulatory care/outpatient settings. General management was the targeted behaviour for all studies.

Results from RCT, CCT and CBA comparisons
Dichotomous process measures. Five comparisons reported dichotomous process measures, including two C-RCT comparisons (A31, A111), two P-RCT comparisons (A16, A34) and one CBA comparison (A147). The two C-RCT comparisons observed absolute improvements of +17.0% (A31) and +20.0% (A111), but both studies had potential unit of analysis errors. The two P-RCT comparisons observed absolute improvements of +1.3% (A34) and +6.0% (A16); neither was statistically significant. The CBA comparison observed an absolute improvement of +25.1% but had a potential unit of analysis error.

Results from ITS comparisons
One ITS comparison observed no significant improvements in performance (A218).

Summary
All comparisons observed improvements in performance. Effect sizes derived from two C-CRT comparisons and one CBA comparison suggest moderate to large effects (all comparisons had potential unit of analysis errors). Effect sizes derived from two P-RCTs were considerably smaller and non-significant, probably due to a contamination effect. The results suggest that combinations of reminders and patient-directed interventions may lead to moderate effects.

Combinations of educational materials, educational meetings, and audit and feedback

Eight studies evaluated combinations of educational materials, educational meetings, and audit and feedback, including four C-RCTs (A23, A158, A129, A39), two CBAs (A74, A127) and two ITS studies (A11, A70). Six studies were conducted in the USA and two in the UK. Three studies were set in primary care settings, two in ambulatory care/outpatient settings, two in inpatient settings and one in a mixed hospital setting. General management was the targeted behaviour in four studies and test ordering in four studies.

Results from RCT, CCT and CBA comparisons

Dichotomous process measures. Three C-RCT comparisons (A23, A129, A39) reported dichotomous process data; the median effect was +3.0% (range +2.6 to +9.0%) absolute improvement in care. All three studies had potential unit of analysis errors.

Continuous process measures. Three comparisons reported continuous process data. The two C-RCT comparisons observed +19.6% (A129) and +400% (A158) relative improvement in performance. It was possible to calculate an SMD for one study of +0.45 (A129). One study had a potential unit of analysis error (A129) and it was not possible to determine the statistical significance in the other study (A158). One CBA comparison (A74) observed –13.1% relative deterioration in performance and had an SMD of –0.21. The study had a potential unit of analysis error.

Dichotomous outcome measures. One CBA comparison (A127) observed an improvement of +1.0%, but the study had a potential unit of analysis error.

Results from ITS comparisons

Two comparisons (A11, A70) observed significant improvements in level following the intervention.

Summary
The observed effects were small (median absolute improvement across three C-RCT comparisons of +3.0%, relative improvement across one C-RCT +0.45 SMD). These results suggest that educational materials, educational meetings, and audit and feedback in combination may have, at best, a small effect on guideline implementation.

Combinations of educational materials, educational meetings and organisational interventions

Six studies (seven comparisons) evaluated combinations of educational materials, educational meetings and organisational interventions, including one C-RCT (A208), two CBAs (A115, A213 – two comparisons) and three ITS (A63, A68, A87). Four studies were based in the USA and two in the UK. Four studies were set in primary care settings. General management was the targeted behaviour in three studies, prescribing in two studies and test ordering in the other study.

Results from RCT, CCT and CBA comparisons

Dichotomous process measures. Three CBA comparisons (A115, A213 – two comparisons) reported dichotomous process data; the median effect was +1.0% (range +0.4 to +6.3%) absolute improvement in performance. All three studies had potential unit of analysis errors.

Continuous process measures. One C-RCT comparison (A208) reported a significant reduction in waiting times.

Continuous outcome measures. One C-RCT comparison (A208) reported no significant differences in patient outcome.

Results of ITS comparisons

Three ITS comparisons were identified (A63, A68, A87). However, these had not been appropriately analysed and it was not possible to reanalyse them.

Summary
The observed effects were small (median absolute improvement across three CBA comparisons of +1.0%). These results suggest that educational materials, educational meetings and organisational interventions in combination may have, at best, a small effect on guideline implementation.

Multifaceted interventions against intervention controls

Sixty-one studies (including 85 comparisons) evaluated 58 different combinations of interventions

TABLE 13 *Summary of comparisons for multifaceted interventions against control groups that received interventions*

No. of interventions in multifaceted interventions	No. of interventions in control group	No. of different combinations of multifaceted interventions	No. of comparisons	Maximum no. of comparisons of specific combinations of study and control interventions
2	1	13	30	6
	2	1	1	1
3	1	9	9	1
	2	8	13	4
	3	2	2	1
4	1	5	5	1
	2	5	7	2
	3	4	6	2
5	1	3	4	1
	2	3	3	1
	3	1	1	1
6	1	1	1	1
7	1	3	3	1
Total		58	85	

(including 14 combinations of two interventions, 19 combinations of three interventions, 14 combinations of four interventions and 11 combinations of five or more interventions) (*Table 13*). The maximum number of studies evaluating a specific combination of study interventions against a specific combination of control interventions was six (*Table 13*).

The studies included 56 C-RCT comparisons (A2, A6, A7, A20 – three comparisons, A35, A36, A43, A45 – three comparisons, A57, A61 – three comparisons, A62, A67, A72, A79 – two comparisons, A84 – two comparisons, A92 – two comparisons, A94 – two comparisons, A114, A120, A122 –two comparisons, A126, A132, A136 – two comparisons, A139, A148, A152, A153 – two comparisons, A155 – three comparisons, A157, A166, A169, A197, A212, A215, A219, A220, A221 – two comparisons, A230 – two comparisons, A232), eight P-RCT comparisons (A30, A33, A104, A105, A193 – two comparisons, A204, A214), five C-CCT comparisons (A199 – two comparisons, A205, A211, A229), two P-CCT comparisons (A108, A165) and 14 CBA comparisons (A4 – three comparisons, A14 – two comparisons, A89, A99, A131 – three comparisons, A162, A185, A210, A233).

Forty-six studies took place in the USA, six in Canada, four in the UK, three in The Netherlands, and one each in France and Germany. Twenty-eight studies were based in primary care settings, 16 in generalist outpatient or ambulatory care settings, eight in inpatient settings, seven in mixed settings and two in accident and emergency settings. The targeted behaviour was general management in 27 studies, prevention in 27 studies, prescribing in four studies, test ordering in two studies and discharge planning in one study.

Results from RCT, CCT and CBA comparisons

Dichotomous process measures. Seventy-nine comparisons (derived from 55 studies) evaluating 49 different combinations of interventions reported dichotomous process data, including 52 C-RCT comparisons (A2, A6, A7, A20 – three comparisons, A36, A43, A45 – three comparisons, A57, A61 – three comparisons, A62, A67, A72, A79 – two comparisons, A84 – two comparisons, A92 – two comparisons, A94 – two comparisons, A114, A120, A122 – two comparisons, A126, A132, A136 – two comparisons, A148, A153 – two comparisons, A155 – three comparisons, A157, A169, A197, A212, A215, A219, A220, A221 – two comparisons, A230 – two comparisons, A232), seven P-RCT comparisons (A30, A33, A104, A105, A193 – two comparisons, A204), five C-CCT comparisons (A199 – two comparisons, A205, A211, A229), one CCT comparison (A165) and 14 CBA comparisons (A4 – three comparisons, A14 – two comparisons, A89, A99, A131 – three comparisons, A162, A185, A210, A233).

Table 14 summarises the median effect size of the absolute improvement in performance across studies for each combination of study and control interventions. Comparable data could not be abstracted for two CBA comparisons (A14 – two comparisons) evaluating the effects of a combination of educational materials, educational

TABLE 14 *Summary of comparisons of multifaceted interventions compared with controls that received interventions reporting dichotomous process data*

Study interventions	Control interventions	No. of studies	Median effect size % absolute difference across studies	Range % absolute difference across studies	Study IDs
Two study interventions compared with one control intervention					
A&F, LCP	Edmat	1	–2.7	NA	A122
A&F, Outreach	A&F	1	6	NA	A99
A&F, Patmed	Patmed	1	9	NA	A136
Edmat, A&F	Edmat	2	3.8	–17 to 24.6	A57, A185
Edmat, A&F	A&F	1	0.9	NA	A221
Edmat, Edmeet	Edmeet	1	–1	NA	A61
Edmat, Edmeet	Edmat	2	2	–5, 9	A230, A230
Edmat, Rem	Edmat	2	20.2	19 to 21.4	A157, A220
Edmeet, A&F	A&F	1	5.8	NA	A89
Edmeet, Org	Edmat	1	8.7	NA	A232
Edmeet, Patmed	Edmeet	1	10.3	NA	A193
Edmeet, Rem	Edmeet	5	8.8	3.9 to 32	A43, A92, A92, A148, A165
Rem, A&F	Rem	1	29.2	NA	A120
Rem, Patmed	Patmed	2	12.3	3.6 to 21	A136, A210
Rem, Patmed	Edmeet	1	11	NA	A61
Rem, Patmed	Rem	1	9.9	NA	A211
Rem, Struc	Rem	1	2.8	NA	A205
Rem, Struc	Patmed	1	2	NA	A212
Two study interventions compared with two control interventions					
Edmat, Outreach	Edmat Outreach	1	–5.2	NA	A169
Three study interventions compared with one control intervention					
Edmat, A&F, Outreach	A&F	1	1.5	NA	A221
Edmat, Edmeet, OL	Edmat	1	10.8	NA	A122
Edmat, Edmeet, Outreach	Edmat	1	24.3	NA	A153
Edmat, Edmeet, Rem	Profoth	1	23	NA	A7
Edmat, Rem, Profoth	Profoth	1	2.6	NA	A30
Edmeet, A&F, Patmed	Edmat	1	–0.3	NA	A132
Edmeet, Patmed, Org	Edmeet	1	30.7	NA	A104
Edmeet, Rem, Patmed	Edmeet	1	13.7	NA	A193
Rem, A&F, LCP	LCP	1	0.5	NA	A229
Three study interventions compared with two control interventions					
Edmat, A&F, LCP	Edmat, A&F	1	5.3	NA	A2
Edmat, Edmeet, A&F	Edmat, Edmeet	1	3.1	NA	A20
Edmat, Edmeet, Patmed	Edmat, Edmeet	1	2.8	NA	A94
Edmat, Edmeet, Rem	Edmat, Edmeet	4	8.7	–2.7 to 34	A20, A45, A84, A131
Edmat, Patmed, Struc	Edmat, Patmed	1	25	NA	A162
Rem, Patmed, Struc	Rem, Patmed	1	10	NA	A72
Rem, Patmed, Struc	Patdir, Struc	1	2	NA	A204
Three study interventions compared with three control interventions					
Edmat, Rem, Org	Edmat, Rem, Org	1	13	NA	A4
Edmeet, Rem, A&F	Edmeet, Rem, A&F	1	4.4	NA	A155
Four study interventions compared with one control intervention					
Edmat, Edmeet, A&F, Org	Edmat	1	0.9	NA	A79
Edmat, Edmeet, A&F, Outreach	Edmat	1	0	NA	A215

continued

TABLE 14 *Summary of comparisons of multifaceted interventions compared with controls that received interventions reporting dichotomous process data (cont'd)*

Study interventions	Control interventions	No. of studies	Median effect size % absolute difference across studies	Range % absolute difference across studies	Study IDs
Edmat, Edmeet, OL, Profoth	Edmat	1	2	NA	A197
Edmat, Edmeet, Patmed, Struc	Struc	1	9	NA	A67
Edmat, Edmeet, Rem, Patmed	Edmeet	1	33	NA	A61
Four study interventions compared with two control interventions					
Edmat, Edmeet, Patmed, Org	Edmat, Edmeet	2	13.7	9.1 to 18.2	A94, A105
Edmat, Edmeet, Rem, A&F	Edmat, Edmeet	1	–9.6	NA	A20
Edmat, Edmeet, Rem, Patmed	Edmat, Edmeet	2	43	34 to 56	A45, A45
Edmat, Edmeet, Rem, Struc	Edmat, Edmeet	1	6	NA	A131
Four study interventions compared with three control interventions					
Edmat, Rem, OL, Org	Edmat, Rem, Org	1	2	NA	A4
Edmeet, Rem, A&F, Patmed	Edmeet, Rem, A&F	2	4.5	2.8 to 6.2	A155, A155
Edmeet, Rem, Patmed, Org	Rem, Patmed, Org	1	–5.7	NA	A199
Edmeet, Rem, Patmed, Org	Edmeet, Patmed, Org	1	1.5	NA	A199
Edmeet, Rem, Patmed, Struc	Edmeet, Patmed, Struc	1	12	NA	A33
Edmat, Edmeet, A&F,	Edmat, OL, Org	1	2.5	NA	A6
Five study interventions compared with one control intervention					
Edmat, Edmeet, Rem, Outreach, Patmed	Profoth	1	–1.2	NA	A233
Edmat, Edmeet, Rem, Outreach, Patmed	Edmat	1	0	NA	A153
Edmat, Edmeet, Rem, Patmed, Org	Edmeet	1	–1	NA	A36
Five study interventions compared with two or more control interventions					
Edmat, Edmeet, Outreach, Fin, Org	Edmat, Fin	1	14	NA	A62
Edmat, Edmeet, Rem, A&F, Fin	Edmat, Edmeet	1	6.2	NA	A84
Edmat, Edmeet, Rem, Org, Struc	Edmat, Edmeet	1	8	NA	A131
Edmat, Rem, LCP, Org, Struc	Edmat, Rem, Org	1	7	NA	A4
Six or more study interventions compared with one control intervention					
Edmat, Edmeet, A&F, LCP, Org, Struc	Edmat	1	2.2	NA	A79
Edmat, Edmeet, Rem, A&F, Outreach, OL, LCP	Edmat	1	12	NA	A114
Edmat, Edmeet, Rem, A&F, Outreach, Patmed, Struc	Edmat	1	8.1	NA	A126
Edmat, Edmeet, Rem, A&F, Patmed, Consultation facility, Phone hotline	Patmed	1	5	NA	A219

TABLE 15 *Summary of comparisons of multifaceted interventions compared with controls that received interventions reporting continuous process data*

Study interventions	Control interventions	No. of studies	Relative % difference (range)	SMD (range)	Study IDs
Two study interventions compared with one control intervention					
Edmat, Rem	Edmat	2	9.6	NC	A108, A157
Edmeet, Rem	Edmeet	1	−1.2	NC	A152
Edmeet, Org	Edmat	1	9.5	0.27	A232
Three study interventions compared with one control intervention					
Edmeet, Patmed, Org	Edmeet	1	21.6	0.33	A104
Edmat, Edmeet, Outreach	Edmat	1	34.9	NC	A153
Three study interventions compared with two control interventions					
Edmat, Edmeet, Rem	Edmat, Edmeet	1	157	0.95	A45
Four study interventions compared with two control interventions					
Edmat, Edmeet, Rem, Patmed	Edmat, Edmeet	2	257.1	1.57	A45, A45
Edmat, Edmeet, Patmed, Org	Edmat, Edmeet	1	12.2	0.25	A105
Five study interventions compared with one control intervention					
Edmat, Edmeet, Rem, Outreach, Patmed	Edmat	1	−1	NC	A153
Five study interventions compared with two control interventions					
Edmat, Edmeet, Outreach, Fin, Org	Edmat, Fin	1	149	4.9	A62

meetings, and audit and feedback compared with a combination of educational materials and educational meetings. A median increase of +3.9% was reported, but the statistical significance of this was unclear.

Continuous process measures. Thirteen comparisons (derived from ten studies) reported continuous process data, including ten C-RCT comparisons (A45 – three comparisons, A62, A152, A153 – two comparisons, A157, A166, A232), two P-RCT comparisons (A104, A105) and one CCT comparison (A108). *Table 15* summarises the median effect size of the relative improvement in performance across studies for each combination of study and control interventions.

Comparable data were not abstractable for one C-RCT comparison (A166). The study compared a combination of educational materials, audit and feedback, and educational outreach with a combination of audit and feedback, and educational outreach. The results were mainly non-significant, but the effect of participation in patient care appraisal was significant.

Dichotomous outcome measures. Seventeen comparisons (derived from 11 studies) reported dichotomous outcome measures, including 13 C-RCT comparisons (A7, A45 – three comparisons, A61 – three comparisons, A79 – two comparisons, A132, A139, A230 – two comparisons), three P-RCT comparisons (A105, A104, A214) and one CBA comparison (A233). *Table 16* summarises the median effect size of the absolute improvement in outcome across studies for each combination of study and control interventions.

Continuous outcome measures. Nine comparisons (from eight studies) reported continuous outcome measures, including seven C-RCT comparisons (A6, A35, A36, A79 – two comparisons, A132, A232), one P-RCT comparison (A214) and one CCT comparison (A108). *Table 17* summarises the median effect size of the relative improvement in outcome across studies for each combination of study and control interventions.

Additional analyses

This section summarises the results of multifaceted interventions including educational outreach compared with other interventions, and three combinations of study interventions compared with combinations of control interventions that had more than four comparisons: educational materials and reminders compared with educational materials; educational meetings and reminders compared with educational meetings; and educational materials, educational meetings and reminders compared with educational materials and educational meetings.

TABLE 16 *Summary of comparisons of multifaceted interventions compared with controls that received interventions reporting dichotomous outcome data*

Study interventions	Control interventions	No. of comparisons	Median % absolute difference across studies	Range % absolute difference across studies	Study IDs
Two study interventions compared with one control intervention					
Edmat, Edmeet	Edmeet	2	–6.5	–1 to +14	A61, A139
Edmat, Edmeet	Edmat	2	–2	–3 to –1	A230, A230
Edmat, Rem	Edmat	1	7	NA	A214
Rem, Patmed	Edmeet	1	20	NA	A61
Three study interventions compared with one control intervention					
Edmat, Edmeet, Rem	Profoth	1	–2	NA	A7
Edmeet, Patmed, Org	Edmeet	1	20.4	NA	A104
Edmeet, A&F, Patmed	Edmat	1	1.4	NA	A132
Three study interventions compared with two control interventions					
Edmat, Edmeet, Rem	Edmat, Edmeet	1	6.4	NA	A45
Four study interventions compared with one control intervention					
Edmat, Edmeet, A&F, Org	Edmat	1	–0.9	NA	A79
Edmat, Edmeet, Rem, Patmed	Edmeet	1	24	NA	A61
Four study interventions compared with two control interventions					
Edmat, Edmeet, Rem, Patmed	Edmat, Edmeet	2	3.5	+3.2 to +3.7	A45, A45
Edmat, Edmeet, Patmed, Org	Edmat, Edmeet	1	21.7	NA	A105
Five study interventions compared with one control intervention					
Edmat, Edmeet, Rem, Outreach, Patmed	Profoth	1	–3.6	NA	A233
Six study interventions compared with one control intervention					
Edmat, Edmeet, A&F, LCP, Org, Struc	Edmat	1	2.2	NA	A79

TABLE 17 *Summary of comparisons of multifaceted interventions compared with control interventions reporting continuous outcome data*

Study interventions	Control intervention	No. of comparisons	Median % relative difference across studies (range)	SMD	Study IDs
Two study interventions compared with one control intervention					
Edmat, Rem	Edmat	2	–8.2 (–16.3 to 0)	–0.32	A108, A214
Edmeet, Org	Edmat	1	10	0.28	A232
Three study interventions compared with one control intervention					
Edmeet, A&F, Patmed	Edmat	1	6.4	NC	A132
Three study interventions compared with two control interventions					
Edmeet, Patmed, Org	Edmat, Edmeet	1	31.3	NC	A35
Four study interventions compared with one control intervention					
Edmat, Edmeet, A&F, Org	Edmat	1		NA	A79
Five study interventions compared with one control intervention					
Edmat, Edmeet, A&F, OL, Org	Edmat	1	6.1	NC	A6
Edmat, Edmeet, Rem, Patmed, Org	Edmeet	1	–20	NC	A36
Six study interventions compared with one control intervention					
Edmat, Edmeet, A&F, LCP, Org, Struc	Edmat	1	5.7	NA	A79

Multifaceted interventions including educational outreach compared with other interventions

Ten studies (involving 11 comparisons) evaluated multifaceted interventions including educational outreach against an intervention control, including nine C-RCT comparisons (A62, A114, A126, A153 – two comparisons, A166, A169, A215, A221) and two CBA comparisons (A99, A233). Six studies were conducted in the USA, two in The Netherlands, one in Canada and one in the UK. The majority of studies ($n = 7$) were based in primary care settings. The targeted behaviour was prevention in five studies, general management of a problem in three studies and prescribing in two studies. All but three comparisons had a single intervention control. The most common control intervention was dissemination of educational materials ($n = 5$).

Results from RCT, CCT and CBA comparisons

Dichotomous process measures. Eight C-RCT comparisons (A62, A114, A126, A153 – two comparisons, A169, A215, A221) and two CBA comparisons (A233, A99) reported dichotomous process measures. The median effect across the eight C-RCTs was +4.5% (range –5.2 to +24.3%) absolute improvement in performance; five comparisons observed improvements in care. Three comparisons had potential unit of analysis errors (A114, A153 – two comparisons) and the statistical significance of one comparison could not be determined (A114). Only one of the remaining comparisons was statistically significant (A62). The two CBA comparisons observed effects of –1.2% absolute deterioration in care (A233) and +6.0% absolute improvement in care (A99); both studies had potential unit of analysis errors.

Five C-RCT comparisons (A114, A126, A153 – two comparisons, A215) evaluated multifaceted interventions including educational outreach compared with an educational materials control. Three out of five studies observed improvements in care; across all studies, the median effect size was +8.1% (range 0 to +24.3%). Three comparisons had potential unit of analysis errors (A126, A153 – two comparisons), one did not have sufficient data to determine significance (A114) and the remaining comparison was non-significant (A215).

One C-RCT comparison (A221) and one CBA comparison (A99) evaluated multifaceted interventions including educational outreach compared with an audit and feedback control. Both observed improvements of care of +0.9% (NS) (A221) and +6.0% (potential unit of analysis error) (A99).

Continuous process measures. Four C-RCT comparisons reported continuous process measures (A62, A153 – two comparisons, A166). Across three comparisons (A62, A153 – two comparisons), the median process continuous effect was +34.9% (range –1 to +149.0%) relative improvement in performance. For two comparisons insufficient data were reported to calculate SMDs and the statistical significance of these comparisons was unclear (A153). The other comparison (A62) observed an improvement of +149.0% and an SMD of 4.9 ($p < 0.001$). Comparable data could not be abstracted for the remaining C-RCT comparison (A166), which reported mainly non-significant results, although the effect on participation in patient care appraisal was significant.

Two C-RCT comparisons (A153 – two comparisons) evaluated multifaceted interventions including educational outreach compared with an educational materials control, and observed relative changes of –1.0% and +34.9%. Insufficient data were reported to calculate SMDs and the statistical significance of these comparisons was unclear.

Dichotomous outcome measures. One CBA comparison (A233) observed an absolute deterioration of –3.6% in a dichotomous outcome measure; however, the study had a potential unit of analysis error.

Summary

The majority of studies evaluating multiple educational outreach against an intervention control observed absolute improvements in performance of care. The effects were modest and less than those observed in comparisons of multiple interventions including educational outreach compared with a no-intervention control. Educational outreach appeared to be more effective than educational materials in three out of five comparisons and audit and feedback in two out of two comparisons.

Educational materials and reminders compared with educational materials

Four studies evaluated educational meetings and reminders compared with educational meetings, including two C-RCT comparisons (A157, A220), one P-RCT comparison (A214) and one CCT comparison (A108). Two studies were based in the USA, and one each in The Netherlands and the UK. Two studies were set in inpatient settings, one in mixed hospital settings, and one in an accident and emergency department. General management

was the targeted behaviour in two studies and test ordering and discharge planning in the other two studies.

Dichotomous process measures. Two C-RCT comparisons reported dichotomous process data and observed absolute improvements in process of care of +19% (potential unit of analysis error) (A220) and 21.4% ($p < 0.05$) (A157).

Continuous process measures. One C-RCT (A157) and one CCT (A108) comparison reported continuous process measures and observed relative improvements in process of care of +6.2% (potential unit of analysis error) (A157) and +14.0% (NS) (A108). Neither study provided sufficient data to calculate the SMD.

Dichotomous outcome measures. One P-RCT comparison (A214) observed an absolute improvement in patient outcome of +7.0%, but the significance of the comparison was uncertain.

Continuous process measures. One P-RCT comparison (A214) observed 0% relative change and 0.0 SMD (non-significant). One CCT comparison observed a relative deterioration in patient outcome of –16.3% and an SMD of –0.32 (NS).

Summary
The combination of educational materials and reminders appears more effective than educational materials alone.

Educational meetings and reminders compared with educational meetings

Five studies (involving six comparisons) evaluated educational meetings and reminders compared with educational meetings, including five C-RCT comparisons (A43, A92 – two comparisons, A148, A152) and one CCT comparison (A165). All studies were based in the USA, were set in ambulatory care settings and targeted general management of a problem.

Dichotomous process measures. Five comparisons, including four C-RCT comparisons (A43, A92 – two comparisons, A148) and one CCT comparison (A165) reported dichotomous process data. All C-RCT comparisons observed improvements in performance; the median effect size was +7.9% (range +3.9 to +32.0%) (three comparisons had potential unit of analysis errors, one had insufficient data to determine significance). The CCT observed an improvement in performance of +24% (NS).

Continuous process measures. One CRCT (A152) comparison observed a relative deterioration of –1.2% in performance. Insufficient data were provided to calculate the SMD and the study had a potential unit of analysis error.

Summary
The combination of educational meetings and reminders appears more effective than educational meetings alone.

Educational materials, educational meetings and reminders compared with educational materials and educational meetings

Four studies evaluated educational materials, educational meetings and reminders compared with educational materials and educational meetings, including three C-RCT comparisons (A20, A45, A84) and one CBA comparison (A131). All studies were based in the USA and targeted general management. Two studies were set in ambulatory care settings, one in an inpatient setting and one in a primary care setting.

Dichotomous process measures. All four comparisons reported dichotomous process measures. Two of three C-RCT comparisons observed improvements in performance; the median effect was +13.3% (range –2.7 to +34%) absolute improvement in performance. Two comparisons had potential unit of analysis errors; the third comparison was statistically significant (A84). The CBA comparison observed an improvement of +4%; however, there was a potential unit of analysis error.

Continuous process measures. One C-RCT comparison (A45) observed a relative improvement of +157.0% and an SMD of +0.95 ($p < 0.05$).

Dichotomous outcome measures. One C-RCT comparison (A45) observed an absolute improvement of +6.4%, but the statistical significance of this comparison could not be determined.

Summary
The combination of educational materials, educational meetings and reminders appears more effective than educational materials and educational meetings alone.

Does the effectiveness of multifaceted interventions increase with the number of interventions?

Table 18 reports the median effect sizes of multifaceted interventions by number of

TABLE 18 *Summary of median effect sizes for evaluations of multifaceted interventions by number of interventions in the study and control arms*

		Median absolute effect size across studies (no. of studies)			
No. of components in study arm	**No. of components in control arm**	**0**	**1**	**2**	**3**
1		10.2 (n = 52)	5.3 (n = 4)	–	–
2		7.4 (n = 36)	8.8 (n = 26)	–5.2 (n = 1)	–
3		11.0 (n = 25)	10.8 (n = 9)	4.7 (n = 10)	8.7 (n = 2)
4		4.0 (n = 11)	2.0 (n = 5)	13.7 (n = 6)	2.4 (n = 6)
5		21.8 (n = 4)	–0.5 (n = 4)	8.0 (n = 3)	8.0 (n = 1)
6		15.0 (n = 1)	2.2 (n = 1)	–	–
7		–	10.1 (n = 1)	–	–

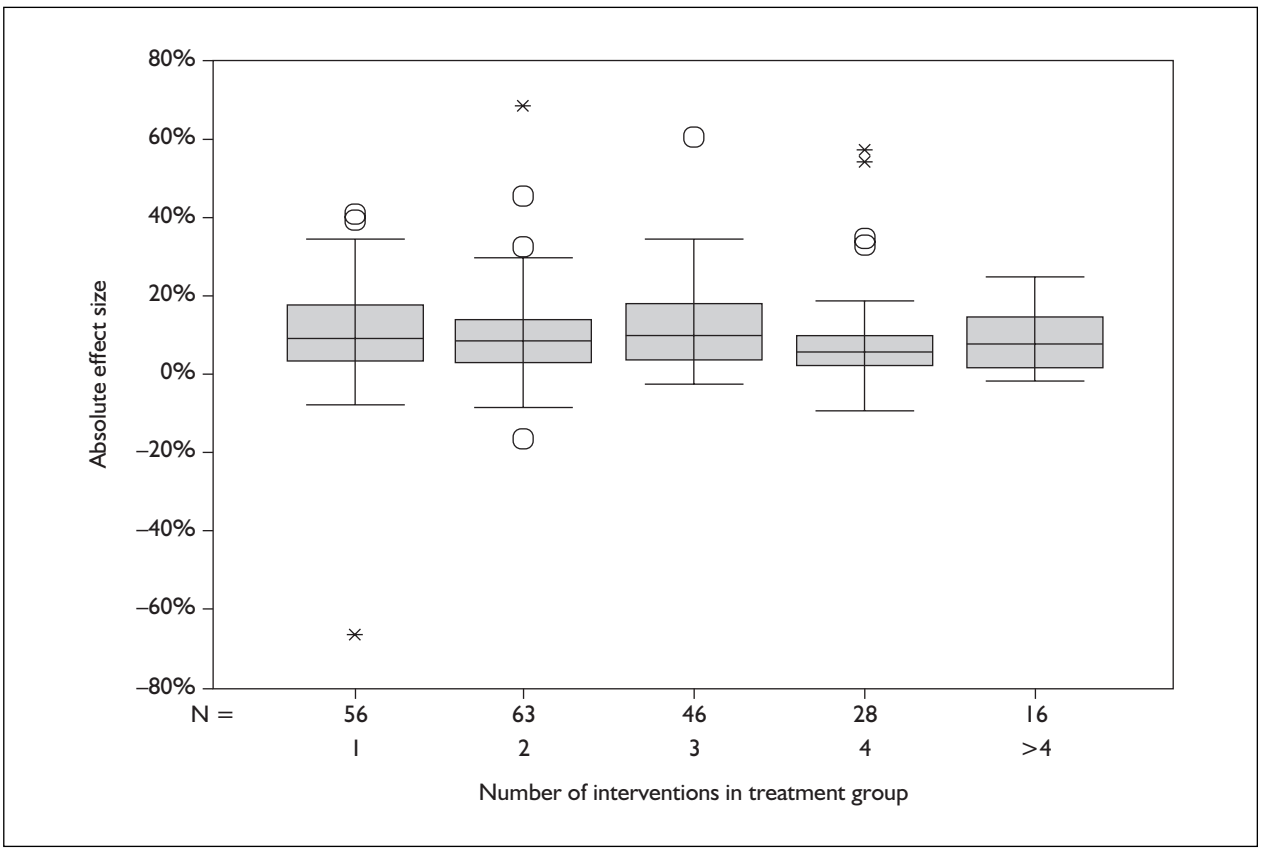

FIGURE 6 *Effect sizes of multifaceted interventions by number of interventions*

interventions in study and control groups. *Figure 6* illustrates the spread of effect sizes for increasing number of interventions in the study group, using boxplots. Visually, there appeared to be no relationship between effect size and number of interventions. For studies with no-intervention control groups, there was no statistical evidence of a relationship between the number of interventions used in the study group and the effect size (Kruskal–Wallis test, $p = 0.18$). In addition, there was no evidence of a difference between studies that used multiple intervention control groups and studies with multiple intervention study groups (Kruskal–Wallis test, $p = 0.69$).

Chapter 5

Systematic review of guideline dissemination and implementation strategies: economic evaluations and cost analyses

This chapter reviews the use of economic appraisal in evaluations of guideline dissemination and implementation strategies. The aim is to determine the frequency and methods of economic appraisal used and to summarise existing evidence on the relative efficiency of guideline dissemination and implementation strategies and the costs of guideline development, dissemination and implementation. From this, conclusions will be drawn about the strengths and weaknesses of the existing economic analyses of guideline implementation and recommendations will be made as to how such analyses can be improved.

The next section of this chapter briefly outlines how an economic evaluation of guideline implementation strategies may differ from an economic evaluation of clinical interventions, and highlights why an analyst may legitimately choose not to include all stages of guideline dissemination and implementation in an economic evaluation. Following this, the methods and results of the review of economic appraisals are presented and, in the final section, the strengths and weaknesses of the existing economic and cost analyses of guideline implementation are discussed.

Structure for the economic evaluation of guideline implementation

Economic evaluation of guideline implementation strategies should be based on the same basic principles as an evaluation of a standard health technology (a new drug or a new type of surgery).[17] A standard health technology assessment limits itself to the consideration of the costs and benefits of providing a treatment (e.g. administering drug A or drug B) and the consequences of that treatment. The evaluation of guideline implementation strategies is different in that the breadth of costs and benefits that could be considered is wider. Determining whether the implementation of a guideline is worthwhile involves determining whether the guideline itself represents an efficient use of resources and the most efficient way of supporting practitioners to adopt the guideline.

There are three distinct stages in guideline development and introduction could be considered in an economic evaluation:[18]

1. development of the guideline
2. dissemination and implementation of the guideline
3. treatment effects and costs as a consequence of behaviour change.

Although the structure of an economic evaluation of guideline implementation strategies could include the costs and benefits from each of the three stages, it may sometimes be legitimate to design an economic evaluation of more limited scope. Whether such limitations are appropriate depends on what justification is given for the limitation. Such justification should be explicit and supported by appropriate evidence.

One legitimate reason to limit the evaluation to the costs and benefits of changing practice may be if the guideline has already been shown to represent efficient practice.[19] Therefore, the costs and consequences of adopting the recommendations of the guideline need not necessarily be assessed in a primary study if the practice advocated by the guideline has been shown to be efficient in the healthcare setting in which the implementation will take place and at the desired scale of implementation.

A further reason for legitimately limiting the scope of the economic evaluation relates to the perspective the evaluation takes. If a societal perspective was adopted then all costs (including those that fall on the patient) and all benefits, such as those gained by health professionals in terms of improved knowledge, job satisfaction that may arise during development, dissemination and implementation of the guideline, should be considered.

A final reason for not measuring some costs and benefits is when the assumption can be made that their inclusion will not change the policy decision. For example, if it is believed, a priori, that a guideline will reduce the costs of treatment while maintaining or improving the outcomes of patients, it may be felt legitimate to exclude the costs of development, dissemination and implementation as they could not possibly cancel out any savings in treatment costs. However, this would limit the transferability of results to the application of the same implementation strategy for a guideline addressing the same issues in a similar setting.

The extent to which economic evaluations and cost analyses included in this review have considered all costs and benefits from each stage is assessed as part of the review. Where relevant costs and benefits have been excluded, consideration will be given to whether their exclusions are justified on the basis of stated perspective, unimportance to final conclusions or previous work showing that adoption of the guideline recommendation would be efficient.

Methods

Search and selection criteria

The economic evaluations and cost analyses included in this review were identified as part of the systematic review reported in Chapters 2–4.

Selection criteria for the review of cost and economic analyses

Two reviewers identified any study that reported information on cost. A second reviewer (L Vale) then determined whether the study reported a detailed cost analysis or an economic evaluation. Studies were defined as cost analyses if they failed to relate costs to effectiveness/benefits. A loose definition of economic evaluation was used in that studies had to report evidence on costs and at least surrogate end-points for effectiveness/benefits.

Review of methodological quality

Included studies in this section were assessed against the *British Medical Journal* guidelines for reviewers of economic evaluations (*Box 3*).[20] These guidelines are designed to improve the quality of economic evaluations, and their recommendations cover three broad areas including study design, data collection, and analysis and interpretation of results. If studies only reported cost analyses, they were not assessed against the criteria relating to benefits. The criteria were used not as a scoring system, but rather as a means of summarising those aspects of an economic evaluation that are generally considered to be important.

BOX 3 *Criteria used to assess the quality of economic evaluations and cost analyses*

Design of economic evaluation

1. Research question stated
2. Importance of question stated
3. Viewpoint of analysis stated and defined
4. Rationale for choosing alternative programmes or interventions compared stated
5. Alternatives being compared clearly defined
6. Form of economic evaluation used stated
7. Choice of form of economic evaluation justified in relation to question addressed

Data collection

8. Source(s) of effectiveness estimates stated
9. Details of design and results of effectiveness study given
10. Details of methods of synthesis of estimates of health effects given
11. Primary outcome measure(s) for economic evaluation clearly stated
12. Methods to value health states and other benefits stated
13. Details of subjects from whom valuations were obtained given
14. Productivity changes (if included) reported separately
15. Relevance of productivity changes to study question discussed
16. Quantities of resources reported separately from their unit costs
17. Methods for estimation of quantities and unit costs described
18. Currency and price data recorded
19. Details of currency and price adjustments for inflation or currency conversion given
20. Details of any model used
21. Choice of model used and key parameters on which it is based justified

Analysis

22. Time horizon of costs and benefits stated
23. Discount rate(s) stated
24. Choice of rate(s) justified
25. Explanation given if costs and benefits are not discounted
26. Details of statistical tests and confidence interval given for stochastic data
27. Approach to sensitivity analysis given
28. Choice of variable for sensitivity analysis justified
29. Ranges of which variables are varied stated
30. Relevant alternatives compared
31. Incremental analysis reported
32. Major outcomes presented in an aggregated as well as a disaggregated form
33. Answer to study question given
34. Conclusions follow from data reported
35. Conclusions accompanied by appropriate caveats

Review of results

Data were abstracted on resource use and cost of guideline development, dissemination and implementation, and summarised according to the type of dissemination and implementation strategies adopted.

Results

Studies reporting cost analyses or economic evaluations

Sixty-three of 235 (29.4%) studies (involving 78 comparisons) reported either cost analyses or economic evaluations that attempted to assess the costs or cost-effectiveness of different guideline implementation strategies (A3, A8, A10, A11, A12, A18 – three comparisons, A23, A29, A30, A32 – two comparisons, A41, A51, A52, A53, A55 – three comparisons, A57, A61 – three comparisons, A63, A64, A72, A73, A82, A87, A93, A96, A100, A101 – two comparisons, A104, A106, A107, A109, A110, A112, A114, A120, A123, A129 – three comparisons, A133, A142, A143, A144, A145, A147 – two comparisons, A154 – two comparisons, A158, A159, A164, A169, A179 – three comparisons, A189, A196, A200, A203, A207, A208, A217, A221 – two comparisons, A223, A224, A226, A228, A229, A234).

General characteristics

Characteristics of the included studies are described in *Table 19*. The majority of the studies were conducted in the USA and aimed to improve management. Thirty-six studies tried to change one behaviour and the remainder targeted several behaviours (to a maximum of six behaviours). The study population always involved clinicians, but in 14 studies other healthcare professionals were involved (most commonly nurses).

TABLE 19 *General characteristics of included studies*

Study characteristic	Type	No. of studies
Study design	C-RCT	24
	P-RCT	10
	C-CCT	2
	P-CCT	5
	CBA	4
	ITS	18
Country of origin	USA	45
	UK	11
	Canada	3
	Other	4
Rationale for study	Improve management	36
	Cost containment	19
	Both management and cost containment	8
Targeted behaviour	General management	24
	Patient education and advice	14
	Prescribing	28
	Preventive services	13
	Referrals	8
	Test ordering	24
	Other	18
Study population	Physicians	49
	Physicians, nurses	3
	Physicians, nurses, pharmacists	1
	Physicians, nurses, pharmacists, other	1
	Physicians, nurses, other	2
	Physicians, unclear	1
	Physicians, other	6

Interventions evaluated

Table 20 describes the interventions evaluated. The majority of comparisons evaluated multifaceted interventions [58 of 78 comparisons (74%)] and the maximum number of behavioural change interventions employed was seven. Ten studies evaluated reminders as a single intervention.

TABLE 20 *Interventions evaluated in economic evaluations and cost analyses*

Intervention type	No. of times used
Educational materials	29
Educational meetings	25
Consensus process	5
Outreach visits	10
Patient-mediated	3
Audit and feedback	24
Reminders	20
Other	17
Financial	4
Organisational	10
Structural	11

Methodological quality of economic evaluations

Design criteria

The viewpoint adopted in the economic evaluation was stated in ten studies (A3, A8, A23, A32, A57, A123, A145, A159, A208, A217) and justified in five (A23, A57, A159, A208, A217). Four studies took the perspective of the health service (A23, A57, A159, A217), five studies took the perspective of the hospital or providers (A3, A8, A32, A217, A123) and one study took a societal perspective (A208).

As a result of the inclusion criteria of the systematic review, the alternatives being compared were clearly defined.

Only 12 of the 63 studies (19%) studies reported the form of economic evaluation used; nine studies reported that they had undertaken cost-effectiveness analyses (A11, A23, A32, A104, A109, A114, A144, A145, A154) and three comparisons reported that they had undertaken cost–benefit analyses (A12, A221, A208). On further inspection, two of the studies that stated they had undertaken a cost–benefit analysis appeared to have undertaken cost–consequence analyses (presenting differences in cost set against differences in several measures of effectiveness) (A221, A208) and one study appeared to have undertaken a cost analysis (A12). One study reported that it had undertaken a cost analysis; however, on closer inspection, this appeared to be a cost–consequence analysis (A57).

The remaining studies did not define the form of economic evaluation undertaken. Thirty-five appeared to be cost–consequences analyses, two cost–effectiveness analyses and 13 were cost analyses, as they reported some aspect of cost (e.g. staff or material costs) but made no effort to relate costs to benefits.

Thirty-eight studies reported costs of treatment changes as a result of guideline dissemination and implementation and 11 reported costs of guideline dissemination and implementation (A8, A10, A23, A32, A55, A120, A142, A158, A179, A228, A229). Two studies reported costs of guideline development, dissemination, implementation and treatment (A87, A208). Twelve studies reported costs of both guideline dissemination and implementation and treatment (A11, A29, A61, A64, A72, A114, A133, A196, A207, A221, A224, A226).

Thirteen studies that reported more than just treatment costs could be defined as cost–consequence analyses, five as cost-effectiveness analyses and seven as cost analyses.

Data collection criteria

Sources of effectiveness estimates

As a result of the inclusion criteria for the systematic review, the sources of effectiveness estimates and details of the effectiveness study design were always available. However, the methodological weaknesses of the primary studies often undermined the effectiveness results. For example, the statistical significance of benefits was uncertain in 20 C-RCTs, C-CCTs and CBAs that had potential unit of analysis errors (A10, A23, A34, A51, A57, A61, A72, A77, A93, A96, A101, A106, A147, A154, A158, A179, A196, A207, A208, A228) and 11 ITS studies were inappropriately analysed in the published reports (A11, A12, A41, A52, A73, A107, A143, A159, A164, A203, A234).

Primary end-point for economic evaluation

In many studies the primary end-point for the economic evaluation was not clearly stated. Only five studies attempted to measure patient outcomes or other economic measures of benefit (A104, A145, A208, A224, A226). The remainder relied on process measures for their primary end-point. As discussed above, it is justifiable to focus on such intermediate end-points if the guideline being evaluated has been shown to be efficient at the target level of implementation. However, the effectiveness evidence base for the vast majority of guidelines evaluated in the systematic review was uncertain (see Chapter 3).

Methods for estimation of costs

The methods used to estimate costs were reasonably comprehensive in 29 studies. The price year was given in only 13 studies (A10, A30, A32, A41, A52, A57, A73, A87, A101, A110, A158, A221, A234) and adjustments for inflation were only made in five (A29, A32, A57, A101, A226) (although this was not needed in four studies owing to the short duration of the analyses – A57, A73, A110, A221). Only three provided details of methods, price year and adjustments for inflation (A32, A57, A101). One of these three looked at the cost of implementation only (A32) and the other two looked at some of the costs of treatment only.

Details of resource use for included costs were provided only in nine studies (A11, A30, A73, A87, A145, A147, A179, A208, A229). Five of these focused on treatment costs only (A11, A30, A73, A145, A147), two looked at the costs of implementation only (A179, A229) and two looked at the costs from all three stages (A87, A208). These analyses were not comprehensive in terms of the costs considered, but one did attempt to bring uncosted resource use and unquantified benefits into the decision-making process by using a balance sheet approach (A208).

Overall, no study gave reasonably complete information on the estimation of cost and covered all three relevant stages.

Details of model used

All studies were based on primary data and no modelling was performed.

Analysis criteria

Discounting was not undertaken or mentioned for any of the economic evaluations in which data collection on costs and benefits was longer than 12 months. Only 16 studies reported some form of sensitivity analysis on cost and cost-effectiveness (A3, A23, A29, A41, A82, A87, A101, A107, A114, A169, A184, A189, A208, A221, A224, A234). In the majority of cases the effect on results was assessed by changes in a single variable (e.g. a cost of a procedure) in one-way sensitivity analysis (A23, A29, A41, A82, A87, A101, A107, A114, A169, A189, A221). One study investigated the effect of changes in personnel and salary level on cost-effectiveness (A179), one study investigated the effect of changing each aspect of cost independently (multiple one-way sensitivity analysis) (A208) and three studies provided a confidence interval around an estimate of cost saving (A3, A224, A234).

Incremental analysis of costs and effects or cost-effectiveness was not performed in nine of the 46 analyses [the remaining three papers did not require incremental analysis as one was unknown (A73) and two detected no differences in either costs or effects (A100, A112)] where it was appropriate (excluding the 14 cost analyses from this analysis) (A23, A63, A72, A104, A169, A179, A207).

In one study initially planned as a cost-effectiveness analysis, the implementation strategy was cost saving (A145). In all but 11 of the 38 cost–consequences analyses it was concluded that the implementation strategy was efficient (A30, A53, A63, A100, A112, A129, A158, A208, A221, A226, A229). In five of these studies the implementation strategy was more costly but no more effective (A100, A112, A221, A226, A229). These conclusions must be treated with suspicion, as must the other conclusions about cost and efficiency, given the limitations in methodology and reporting reported above.

Summary of resource and cost estimates for guideline development, dissemination and implementation

Owing to the generally poor quality of reporting of the economic evaluations, data on resource use and cost of guideline development, dissemination and implementation are not available for most of them. Four analyses provided reasonable information on resource use and on the methods of cost estimation (A87, A179, A208, A229). All but two of these analyses reported resource use and cost in the implementation stage. These two analyses covered all four stages (A87, A208). Details of these analyses are summarised in *Tables 21a* (details of studies), *21b* (resource use in guideline development stage) and *21c* (resource use in guideline dissemination and implementation stage). Two studies used the preparation and dissemination of educational materials alongside educational meetings (A87, A208). The last study reported resource use for audit and feedback.

The resource information reported for audit and feedback may not be transferable to another setting, as the resource use refers to computer and staff time and the study was conducted in the late 1970s (A229). Since this time computer equipment has changed substantially.

Discussion

The purpose of economic evaluation is to provide information with which to aid judgements about the use of scarce resources. By identifying and critiquing the available evidence systematically on guideline implementation strategies, this chapter has sought to provide information to decision-makers and researchers to aid their decisions on how best to get guidelines into practice. Although the number of studies included in the review is large, the multifaceted nature of many of the implementation strategies adopted and the multitude of policy issues addressed precluded the presentation of results on cost-effectiveness of alternative implementation strategies in any meaningful form. This problem was anticipated at the outset of this study and for this reason data on resource use and costs of guideline development, dissemination and implementation were sought. It was felt that these data would be most useful for those who are planning the implementation of a guideline in their own setting.

The quantity of usable data on resource use and cost was comparatively small given the number of studies included in the review. However, these data still provide a valuable resource for researchers and have been shown to come from analyses of a reasonable quality.

In general, the methodological quality of the included studies was poor. This finding is in common with other reviews of economic evaluations[21] and in part should be expected given the loose interpretation of what defines an economic evaluation. In another recently completed review of economic evaluations the

TABLE 21a *Details of studies reporting resource consequences of guideline development, dissemination and implementation strategies*

Study	Intervention	Behaviours targeted	Setting
Gurwitz [A87]	**Distribution of educational materials, educational meetings**, other professional behavioural change (list of patients receiving target therapy), financial interventions (changes to reimbursable products)	Prescription of H_2 receptor antagonists (treatment of gastrointestinal disorders) in nursing-home patients	Boston, USA Long-term care facility Time of study: 1988–1990 Costs in 1989 US dollars Cost of providing financial change not stated
Thomas [A208]	**Distribution of educational materials, educational meetings**, organisational changes (open-access clinic)	Guidelines for two urological problems: prostatism due to benign prostatic hyperplasia, and microscopic haematuria in the general population	Grampian, Scotland Family practice Time of study: 1995–1996 Costs in UK pounds Year not stated, possibly 1995/6 Costs of providing organisational change intervention not stated
Rosser [A179]	**Reminders** about requirement for a procedure/test	Influenza vaccination Cervical screening Blood pressure screening Tetanus boosters	Ontario, Canada Family practice Time of study: 1984/5 Unclear whether costs in US or Canadian dollars Year that costs relate to not stated
Winickoff [A229]	Reminders, **audit and feedback**, local consensus process	Management of hypertension	Boston, USA Primary care Time of study unclear, but the study took 18 months Costs in 1979 US dollars Only costs of providing audit and feedback were reported

Data in **bold** identify the intervention on which resource use data are provided.

TABLE 21b *Resources used in the guideline development stage*

Study	Area of resource use	Quantity of resource use	Monetary cost for development costs
Gurwitz [A87]	Literature review by MD/pharmacist	20 hours (also includes preparation of documentation which should be considered as part of dissemination/implementation costs)	$650 (at $32.50/hour)
Thomas [A208]	GP/nurse/clinician researcher time spent at meetings (including travel time)	217 hours (163 leisure hours; 54 work hours)	£1944 (work hours only); £9029 including valuation of leisure time
	Research staff preparing for development meetings	182 hours	£2676
	Travel costs	Mileage or unit costs not reported, but 1998 UK Automobile Association cost per mile	£462
	Consumables	Not detailed	£3329

TABLE 21c *Resources used in the dissemination and implementation stage*

Study	Cost-generating event	Quantity of resources used	Cost
Gurwitz [A87]	*Educational materials*		
	Documentation preparation by MD/pharmacist	See *Table 21b*	See *Table 21b*
	Review of medical records for documents by MD/pharmacist	25 hours	$812.50
	Printing costs	Resource use not recorded, but materials required for 16 providers	$200
	Educational meetings		
	Preparation time for presentation by MD/pharmacist	5 hours	$32.50
	Group discussions attended by 16 members of staff and MD/pharmacist	Time not stated	Not costed
Thomas [A208]	*Educational materials*		
	Consumables (printing folders)	Resource use not detailed, but guideline was disseminated to approximately 300 GPs and 74 practice managers	£2484
	Time spent assembling and mailing the guidelines	24 hours	£265
	Postage of guidelines and letters/reminders	Resource use not detailed but guideline was disseminated to approximately 300 GPs and 74 practice managers	£430.85

continued

TABLE 21c *Resources used in the dissemination and implementation stage (cont'd)*

Study	Cost-generating event	Quantity of resources used	Cost
	Educational meetings		
	GP/nurse/clinician researcher time spent at meetings (including travelling time and postgraduate education allowance payment for GPs	111 hours, all of which were outside regular working hours. This cost was spread over 74 general practices and approximately 300 GPs	£0 as no work time forgone, but £7024 if leisure time valued
	Research staff preparing for meetings	40 hours	£517
	Travel costs	Mileage or unit costs not reported, but 1998 UK Automobile Association cost per mile	£304
Rosser [A179][a]	*Letter reminder to patients*		
	Clerical time to prepare letter	1.7–5.77 minutes (depending on clinical area)	$0.28–0.96
	Physician time to sign letters	10 seconds	$0.16
	Stamps	Not specified	$0.32
	Stationery	Not specified	$0.06
	Repeated costs of above for 2nd reminder letter	84% patients required repeated mailing	
	Telephone reminder to patients		
	Clerical time to prepare patient documentation	0.33–2.8 minutes	$0.06–0.48
	Nurse time contacting patient	2.8 minutes	$0.70
	Cost of telephone calls	Not specified	Not specified
	Repeat calls to patients	1.7–2.4 calls per patient	Not specified
	Physician time to explain need for test/procedure	0.25–1.70 minutes	$0.25–1.70
Winickoff [A229]	Computer time used to create reports from routine data for 2216 patients for 16 physician/nurse provider teams	90 hours	$2300
	Staff time to produce reports	5–10 hours	$27.50–55.00

[a] Quantities of resource use and costs expressed are per patient.

criterion for inclusion stipulated that to be included a study had to present costing methodology in the methods section and results of the economic evaluation in the results section.[22] Had this criterion been applied in this review it would have undoubtedly reduced the number of included studies, but it is unlikely that the overall conclusion of generally poor methodology would have changed.

It is perhaps surprising in a review of guideline implementation studies that relatively few studies considered any costs other than those of treatment and its consequences. In several cases this limitation in scope would not be expected to change the conclusions as the magnitude of cost savings provided by adopting recommended practice was so large (A52, A101, A143, A189, A203). Therefore, in such cases, other methodological weaknesses aside, the evaluation would have been fit for the purpose for which it was designed. However, the results of such evaluations are context specific and have limited transferability. Furthermore, in many cases where it was concluded that implementation was cost saving, the inclusion of the costs of guideline development, dissemination and actual implementation could reverse the results.

The search strategy used to identify the included studies was devised with the intention of identifying those studies based on robust study designs. The study designs were chosen as it was felt that the data they could provide, if analysed appropriately, would provide the most robust data on effectiveness. As such data often help to determine total costs and are integral to estimates of efficiency, it is implicit that such studies should provide the best data on which to base at least some parts of an economic analysis. It is possible that the search strategy did not identify some studies that reported economic data separately or used data from robust study designs to model efficiency.

Conclusions

The paucity of data on the cost and efficiency of guideline implementation strategies has been shown in this review. In general, studies were of poor methodological quality and did not cover all stages of guideline introduction that may be relevant. While it is tempting to recommend further large-scale studies that can give unbiased estimates of costs and effect for all stage of guideline implementation, it is unlikely that in many cases such studies will be practical, owing to statistical issues relating to sample sizes.[23] More realistic is the approach outlined by Mason and colleagues.[24] With this approach primary studies should concentrate on evaluating behavioural change and costs of development, dissemination and implementation of the guideline, while modelling exercises should use these data to determine whether the guideline is efficient at the level of behaviour achieved or desired.

Chapter 6

Estimating the feasibility and likely resource requirements of guideline dissemination and implementation strategies in UK settings

Methods

A semistructured telephone survey of key informants in primary and secondary care settings was conducted to assess the feasibility and possible resource requirements of guideline dissemination and implementation strategies in the UK. The original intention had been to undertake focus groups with key informants; however, it proved impossible to arrange these within the study timescale.

Informants were selected to represent a range of viewpoints relevant to guideline implementation across primary care, district general hospital and teaching hospital settings. They were selected from a convenience sample of medical practitioners with experience of guideline implementation based predominantly in the Northern and Yorkshire region of England. Eight individuals were initially approached. One medical practitioner was a non-responder (a GP). One invitee was unable to carry out the interview owing to other commitments, but he was replaced by a colleague meeting the same criteria. The final seven informants were two full-time GPs, one of whom was a chair of a Primary Care Group (PCG) (GP1, GP2), one professor of general practice and part-time GP (GP3), one district general hospital (DGH) consultant physician (DGH1), one senior lecturer in medicine with beds in a district general hospital consultant (DGH2), one district general hospital associate medical director (DGH3) and one teaching hospital associate medical director/clinical director (TH1).

The semistructured telephone interviews were carried out by PW and lasted between 1 and 1.5 hours. Concurrent notes were made during the interviews, which were also audiotaped and transcribed. Before the interview, respondents were sent an interview brief (see Appendix 7) that outlined the interview questions and provided descriptions of guideline dissemination and implementation strategies to be considered (based on the definitions of the Cochrane EPOC group; see *Box 1*). It also gave examples of resource use reported in the systematic review of guideline dissemination and implementation strategies. Views were sought on the factors that might influence the resource use of each intervention strategy, how far the respondent thought the interventions were feasible within current resources in their own setting, and respondents' views on feasibility throughout the UK. Respondents were also asked to rank the interventions in order of how intensive resource use would be. As the interviews proceeded, it became clear that a wide definition of 'resource use' needed to be adopted and the full range of issues related to the feasibility of the interventions was explored.

The interviews were analysed thematically (from the written notes). Themes are backed up by a small number of relevant quotes taken from the interviews. The draft chapter was sent to respondents for amendment of their contributions and as validation of the findings presented. The presentation of the results is organised by the six broad dissemination/implementation interventions considered.

Results

Distribution of educational materials

Respondents were asked to consider the distribution of three different types of educational material:

- a laminated, single-sheet guideline sent to all relevant practitioners with an introductory letter
- guidelines mailed to all members of relevant specialist society and consumers associations; also published in national medical journals and speciality bulletins
- educational video given to relevant sites/specialists.

The GP respondents did not envisage any problems with resources for distributing either

laminated A4 sheets or copies of full guidelines (although all respondents commented that they would not choose to distribute full guidelines). Purchasing videos was thought to be possible by all if drug company sponsorship were available. One GP (GP2) had experience of making a video locally, funded by a research grant, but otherwise this was thought to be neither possible (no access to expertise) nor affordable.

At least two of the four secondary care respondents considered that every strategy would need 'soft' money (pharmaceutical company or health charity) to support it.

> "District general hospitals just don't have budgets for this sort of thing" (DGH1).

One secondary care respondent (DGH2) had experience of producing a video locally funded by research money, through collaboration with a nearby university, but acknowledged that most DGHs would be unlikely to have access to such facilities.

The main resources identified by respondents were postage costs if sent by external mail. Respondents identified a number of factors that could also influence the resources required (*Box 4*).

BOX 4 *Factors influencing resource requirements for distribution of educational materials*

Printed materials
- Number of pages (e.g. full guidelines versus a single two-sided A4 sheet)
- Number of copies
- Frequency of revision of guidelines; therefore, the number of times they need to be distributed
- Labour involved in photocopying, assembling, and stuffing and labelling envelopes
- Opportunity costs of reading time
- Availability and cost of a laminating machine (approximately £50) for sending laminated two-sided sheets
- Availability and cost of suitable software to produce professional, eye-catching single-sheet materials, and the skills to use the software

Purchased educational videos
- Cost (typical examples approximately £12)
- Number required (distribution to individuals or groups)
- Availability of audiovisual facilities for viewing the video
- Opportunity costs of group viewing

Local development of educational videos
- Clinical expertise
- Scripting expertise
- Filming expertise
- Technical editing expertise

Educational meetings

Respondents were asked to consider four different types of educational meeting:

- a series of 1-hour departmental lectures for relevant practitioners
- a half-day conference for relevant practitioners, hosted by a local 'expert'
- several intensive group educational sessions (over 2 hours), for small groups of relevant practitioners
- one didactic 2-hour meeting for relevant practitioners (given a choice of several sessions being run at different venues) with presentations from a local expert and peers involved in developing the guidelines of interest.

A key issue identified by primary care respondents concerned the time of day and length of meeting. They suggested that lunchtime and evening meetings would be much less costly as locum fees to release GPs are not required. However, locum fees would be required for meetings at other times, and for half-day or longer meetings. Where locum fees are required, the number of GPs attending the event will be the key determinant of the resources required. GP2 illustrated this with an estimate that a single half-day event for 38 GPs would cost his PCG £4000 out of an annual GP education budget of £20,000. One respondent commented that the traditional approach of 1-hour didactic lectures given by a local 'expert' was often considered relatively inexpensive but that this was:

> "only true for the provider of the lecture, not for the audience. The 'opportunity costs' of large numbers of GPs attending a lecture [during daytime hours] are enormous" (GP1).

Time of day, length of the meeting and time available are also issues in secondary care. Respondents thought that it would be possible for at least some staff to attend lunchtime meetings, but that it is more difficult to free staff for a half-day meeting (which might require cancellation of outpatient clinics or theatre lists). Depending on the service involved, there may also be problems with providing emergency cover. Where guidelines potentially affect a large number of staff in a range of professions and specialities (e.g. prevention and management of deep vein thrombosis), these problems are compounded by the need to arrange several sessions at different times to accommodate shift patterns. One respondent noted:

> "there is no protected time available for us …. And there is no such thing as an audit half-day. The only way to organise sessions of any length is to cancel clinics" DGH1.

For all settings, the availability of 'expert' speakers with adequate presentational and educational skills, or of trained facilitators could be problematic. GP2 noted that there were only three GPs who could facilitate such work in his PCG. For ongoing local educational programmes, the costs of training facilitators in these skills and the availability of facilitators in local networks may be problematic.

A key difference for specialists working in secondary care is the relatively small number of people who would be participants in a local educational event targeted at their specialist area. Respondents thought that regional or national events may be more appropriate. However, respondents noted that secondary care clinicians tend to work in multidisciplinary clinical teams, where there would be an expectation to hold regular educational meetings in-house.

In summary, 1-hour lectures (either as a single didactic meeting or in a series) were considered feasible in both primary and secondary care settings. In primary care, other educational activities would incur additional expenses, most notably costs for locums to cover GPs' surgeries.

GP3 held the view that all of these activities are feasible if available education budgets are tapped into. One of the secondary care respondents who also works with primary care (DGH3) felt that resources are more readily available in primary care for these activities:

> "badged under clinical governance, lifelong learning, etc. …. Whereas it is much more difficult in secondary care, particularly to persuade consultants to free up a half-day."

Respondents identified a number of factors common to both primary and secondary care that influence the resources required (*Box 5*).

BOX 5 *Factors influencing resource requirements for educational meetings*

- Location: in-house versus external
- Distance from participants' workplaces; cost of the venue, including catering (for external venues)
- Frequency of meetings
- Length of meetings
- Expert speaker fees
- Group facilitator fees
- Local availability of experienced group facilitators
- Administration costs

Educational outreach visits

Respondents were asked to consider three different types of educational outreach:

- one-to-one visits by pharmacists (community or hospital as relevant to the setting) who have been trained as educators (> 2 days' training). No more than two brief (less than half an hour) visits to all relevant practitioners
- visits to clinical or practice teams by a nurse trained as an educator (> 2 days' training). Meetings last for at least an hour and occur several times
- a doctor trained as an educator (> 2 days' training). The doctor presents guidelines and specific educational messages at 1-hour presentations to services or groups of practices. In in-patient settings, this could be supplemented by participation of the doctor-educator on ward rounds.

Respondents recognised that educational outreach was likely to be an expensive strategy and thought that a key issue would be whether the topic was considered to be a priority where additional resources might be available; for example, money made available for the implementation of a National Service Framework topic in England.

Primary care respondents recognised that educational outreach was a less costly strategy for primary care staff as it would be delivered in the practice setting. Some suggested that the new PCG liaison pharmacists might be able to undertake outreach visiting as part of their role, although one respondent (DGH3) questioned the feasibility of doing this if they could only manage infrequent visits to the practices in their area. One respondent noted that the local PCG had decided they would need additional pharmacist posts to undertake this sort of work, estimating a need for up to three posts at approximately £40,000 per post per annum. Using doctors in this role was judged much less feasible and would require payment for a fully trained GP to be released (full-time costs of up to £60,000 per annum). The academic GP respondent (GP3) who had researched the cost-effectiveness of educational outreach visiting by doctors, costed this as £583 per practice in 1995 to implement a full set of guidelines, using doctors paid on a clinical lecturer's salary and including the costs of Postgraduate Education Allowance.

There were two significant areas of discrepancy in respondents' views about the availability of nurses or doctors in hospital to undertake such roles. Two hospital respondents (DGH3, TH1) saw ward nurses, in particular, as too valuable to be spared for such activities, especially given recruitment problems for many acute inpatient nursing specialities. DGH3 also disputed the feasibility of the strategy, questioning the credibility of nurses in this role with some doctors. Others (DGH1, DGH2) pointed out that this role is routinely undertaken by specialist nurses (e.g. in diabetes, renal care, respiratory liaison, stroke), who are usually credible with doctors. DGH3 described doctors taking on this role as "*cloud cuckoo land*", whereas other hospital respondents saw releasing junior doctors as feasible (and more feasible than nursing staff).

The primary care view appeared to be that resources for educational outreach visits are not routinely available, but could be accessed via 'targeted' funding (e.g. to implement a National Service Framework) or otherwise from Health Authority-type funding (if successful in a competitive bidding process). This was one area where possibly resources are more readily available in secondary care, at least for pharmacists. This may be an example of redeployment of existing resources as DGH1, for example, commented that they had not thought of using clinical pharmacists in this role.

One of the secondary care respondents (DGH2) felt that, for this strategy and for the next strategy considered (use of opinion leaders), the resources not only were currently available but were being deployed. In DGH2's hospital, lead consultants for each speciality (stroke medicine, respiratory medicine, diabetes, etc.) were identified as 'leads' in these areas and expected to provide educational input for colleagues. However, DGH2 acknowledged that this model may not be generalisable to other DGHs:

> "… you need to be able to take advice from colleagues for this sort of model to work …."

DGH3 thought it unlikely that educational outreach could be a cost-effective strategy:

> "… if educational outreach is to be used at all, it should be for the 'laggards' who haven't attended the other educational events you've provided."

Factors influencing resources required for educational outreach are summarised in *Box 6*.

BOX 6 *Factors influencing resource requirements for educational outreach visits*

- Location: in most cases, visits would be expected to be to practitioners in their own setting
- Frequency of visits
- Length of visits (time implications of this for both the receivers and the deliverers)
- Delivered to individuals or groups
- Number of outreach workers
- Profession of outreach workers (availability of staff and salary costs)
- Costs of training the outreach workers
- Costs of providing 'freebies' similar to those provided by drug company representatives (lunch, pens, Post-it pads, etc.)
- Administrative time for setting up appointments

Local opinion leaders

Respondents were asked to consider two scenarios, both involving colleagues nominating peers as OLs:

- small number of OLs identified who undergo a short period of training (< 1 day). Activities they carry out may include providing covering letters for guidelines mailed to colleagues, hosting an educational meeting, enhancing their ongoing educational contacts with colleagues
- large number of OLs identified who undergo a substantial period of training (minimum 2 days). Activities they carry out may include establishing and leading task forces including educational activities and outreach programmes.

Respondents thought that there may be a 'threshold effect', according to the number of OLs required and the amount of time that they would be expected to contribute to the strategy: "by definition, opinion leaders are few and far between" (GP1). Primary care respondents noted that one individual may be identified as an OL for several clinical areas, which would have a major impact on their time commitment (GP2). This is also true for secondary care, where there may be only one 'specialist lead' for a subject (e.g. diabetes), especially in DGHs. Identifying several OLs may only be possible over a region, rather than locally. The time commitment could dictate whether the OLs require paying: one respondent felt strongly that being an OL for your colleagues is an integral part of being a 'specialist lead' (DGH2) and one of the GP respondents (GP3) could see this as an informal activity not requiring payment. However, significant time commitment

and extension of activity beyond local colleagues and local area would, in both of these respondents' views, increase the likelihood that payment for time would be required.

Although the first scenario was seen as more feasible than the second owing to the numbers required, the primary care respondents appeared to see use of OLs as a less feasible strategy overall. A particular issue was the absence of easily identifiable OLs in primary care.

DGH1 identified a practical problem with these strategies in his setting: where a clinician has a special interest, a relevant OL needs to be brought in from elsewhere. Alternatively, when that same clinician acts as an OL for generalist colleagues (e.g. a diabetologist for other general physicians), then their time availability for the preparation and delivery of educational activities becomes a problem and would frequently be done out of hours. The other DGH respondent (DGH2) took a different view, stating that all the 'specialist leads' act as OLs for their colleagues, with all of the physicians acting as OLs to each other. But this is only possible when there is "… an open atmosphere in terms of asking for and receiving advice."

There were contrasting views on the identification and training of OLs. TH1 commented that "… it would be good to have a systematic way of identifying and using opinion leaders … but it isn't a sufficient priority." DGH3 suggested that it might be possible to identify large numbers of OLs in primary care, using clinical governance leads from practices, but countered this by questioning whether such a person, who is accountable for clinical governance within a practice, could ever meet the definition of an OL. GP3 pointed out that, even though it may possible to identify a larger number of OLs at regional level, it would be difficult to justify the time to train them: "By definition, these are busy people with little spare capacity". DGH2, in contrast, suggested that training for local OL activity should not be necessary: "Aren't opinion leaders by necessity good communicators, otherwise how can they be recognised as influential by their peers?"

Factors influencing resources required for local OLs are summarised in *Box 7*.

BOX 7 *Factors influencing resource requirements for local opinion leaders*

- Number of OLs required
- Costs of identifying and training the OLs
- Payment for the OLs' time
- Opportunity costs of OLs' time
- Investment of time by the OL (i.e. preparation time out of hours)
- Costs of educational strategies used by the OLs (educational meetings, etc.)

Audit and feedback

Respondents were asked to consider four scenarios:

- regular, frequent (minimum weekly) electronic mail messages to practitioners, containing computer-generated reports on compliance with guidelines over the previous recent weeks
- monthly paper reports to practitioners containing data on compliance with guidelines and a comparison of performance with anonymous peers; data generated from electronic medical records
- monthly seminars where individual practitioners are given paper reports containing personal performance in complying with guidelines, comparison of performance with anonymous peers and a commentary from a 'local expert'; data obtained from (manual) medical record audits
- quarterly departmental meetings where departmental compliance with guidelines is presented to the department, with commentary from an external expert; data obtained from (manual) medical record audits.

The main resource factors identified related to the methods of information gathering (in particular whether data had to be abstracted from manual records or routine computerised information systems). All agreed that data abstraction from manual records is very resource intensive in terms of people required and time. Further data handling and preparation of feedback are also time consuming and may require training for the staff involved. Respondents recognised that there may be staff available who can do all or some of these activities without needing additional payment (e.g. in secondary care, junior doctors carrying out their required audit projects), but such staff may not be the best suited to carrying out these activities, especially without training (DGH2).

Where audit data are to be abstracted from a computerised information system, the key factor influencing resources is whether a system is already in place that can fulfil some or all of the requirements. Where it is, there are very few additional costs of providing feedback of audit data. If not, the costs of purchase and installation

of a new system are substantial. The training implications to ensure standardised data entry and appropriate use of a new system are also substantial. Respondents suggested that currently there were few computerised information systems that could provide reliable and relevant feedback. Secondary care respondents considered manual data abstraction to be the only feasible method of producing audit data currently.

Primary care respondents, while recognising the greater availability of computerised systems, highlighted the limitations of current systems to provide audit and feedback. DGH3 observed that "… though GPs usually have access to computers, they would not usually have systems to generate automated reports". GP1 and GP2 also expressed frustration at the failure to achieve a minimum standard of systems that could routinely provide clinical audit data. GP3 identified some exceptions in primary care based around MIQUEST (which enables interrogation and extraction of data from different types of GP practice systems using a common query language),[25] but noted that area-wide audits using MIQUEST software are uncommon and frequently funded from 'soft' money.

Other resource factors were the frequency of reports (although more so for manual record-based audits and where feedback is through a group meeting) and whether feedback is through reports to individual clinicians or requires a group meeting. There are the 'opportunity costs' of the time commitment required by the meeting. Where an 'expert' is expected to provide commentary on the feedback, there may be a particular problem in sustaining his or her time commitment and enthusiasm if this is expected over a series of meetings (DGH2).

Factors influencing resources required for audit and feedback are summarised in *Box 8*.

Reminders

Respondents were asked to consider three scenarios:

BOX 8 *Factors influencing resource requirements for audit and feedback*

- Costs of data abstraction
- Costs of data preparation and handling
- Costs of preparing feedback reports
- Costs of disseminating reports
- Frequency of reports
- Additional associated activities, e.g. educational meetings

- stickers (with spaces for recording appropriate clinical management actions) placed on medical records by administrative staff. Additional brightly coloured 'spots' added by administrative staff to records of patients meeting criteria that indicate a specific intervention is due or required
- computer-generated reports sent annually to clinicians by a central administrative system, detailing interventions/procedures undertaken during the previous year and those apparently overdue. Space for clinicians to complete missing information in procedures done during the year, to be returned to the administrator
- in the context of a computerised tracking and/or electronic record system, computer-generated 'alerts' and 'messages' to clinicians, derived from management guidelines. 'Alerts' would be sent to clinicians every time a relevant event happened to one of their patients; a 'message' would be a prompt to appropriate management action when the clinician opens that patient's electronic record.

Respondents raised issues relating to resources required to put stickers on notes relating to identification of relevant patients and clerical time. They also raised issues relating to the feasibility of this strategy, especially if multiple stickers were used: "You forget what they're for or start ignoring them" (GP1). GP3 made the general point that each general practice can manage a maximum of two guidelines per year, in terms of both the educational input required and the numbers of reminders being generated, which has clear implications for the widespread introduction of such reminder strategies.

Secondary care respondents suggested that stickers on manual records had not been fully explored in their setting. DGH1 commented that "it would not be possible to use medical record staff in our DGH for such an activity. It might be possible to carry out for a small number of people in selected clinical areas". DGH2 also highlighted that care pathways frequently use reminders in the form of structured manual records containing prompts and tick-boxes.

Respondents raised similar issues about the use and functionality of information systems for reminders to those raised in the discussions about audit and feedback. Primary care respondents thought that computer generation of annual or low-frequency reports containing reminders would not necessarily need additional resources, as systems exist for selected conditions (e.g. diabetes

registers, cervical screening, immunisations). Such reports could be generated from centrally held manual records. However, GP3 suggested that the time for practices to consider and respond to lists generated by such systems should not be underestimated, although administrative staff may welcome the assistance provided by externally generated lists. The extra complexity of a prompt to specific clinical action on a summary-type report, rather than a 'yes/no' to whether an intervention has been completed, was highlighted by a secondary care respondent (DGH2).

All respondents agreed that more sophisticated systems as described in the third scenario would need very high levels of capital and software IT investment in both primary and secondary care. In addition, resources would be needed to train staff to input data and routinely use the system. Several respondents suggested that there would need to be a major cultural shift to make routine use of such a system viable (GP3, DGH3, TH1). However, once such systems were implemented, the resource implication of this reminder strategy would be fairly low. Secondary care respondents were concerned about the impact of these systems on clinician time (in particular the differences in clinician time required to enter data onto computer compared with manual notes or dictation) (DGH1, DGH2).

Factors influencing resources required for reminders (for manual stickers) are summarised in *Box 9*.

BOX 9 *Factors influencing resource requirements for reminders*

Manual sticker resource factors
- Staff time identifying patients meeting the set criteria and placing the stickers on records
- Availability of staff
- Training required by administrative staff to identify patients [the more complex the clinical issue, the more training is required (GP2)]

Relative use of resources for the different strategies

Respondents were asked to rank the relative resources required for the different strategies (*Table 22*). Educational outreach was considered the most resource intensive and distribution of educational materials the least resource intensive by most respondents. There was greater variation in rankings across the other interventions, which appeared to be explained in part by the recognition that resource implications vary within individual strategies and different ways for accounting for IT costs.

Discussion

Respondents considered distribution of educational materials to be the least resource-intensive strategy and educational outreach the most resource-intensive strategy. They suggested that the availability (now or in the future) of suitable IT systems is a fundamental issue if either 'audit and feedback' or 'reminder' strategies are to be used widely. This was particularly problematic in secondary care: none of the respondents worked in a hospital with routine computer information systems that could provide routine feedback or generate reminders. Even with the better

TABLE 22 *Respondents' ranking of how resource intensive the guideline implementation interventions would be (1 = most resource intensive; 6 = least resource intensive)*

Respondent	Distribution educational materials	Educational meetings	Educational outreach visits	Local opinion leaders	Audit and feedback	Reminders
GP1	4	3	1	2	5	6
GP2	6	5	2	4	1	3[a]
GP3	6	3	1	2	5	4
DGH1	6	4	1	2	3	5
DGH2	6	1	3	4	2	5
DGH3	6	5	1	4[b]	3[c]	1[d]
TH1	6	4	3	2	1	5[d]

[a] Would be ranked '1' if manual records were used to extract data.
[b] Would have ranked '1' for scenario 2.
[c] Scenarios 1 and 2 for 'audit and feedback' and 2 and 3 for 'reminders' would be the highest cost because of the need for IT.
[d] Except for scenario 3, which would rank as 1.

developed systems in primary care, none appeared to have the capabilities to run the more sophisticated reminder strategies.

Respondents (especially in secondary care) found it difficult to identify routinely available resources to fund guideline implementation strategies, but recognised that resources are more likely to be available for national priority areas, for example, those covered by National Service Frameworks. There also seem to be differences for the management of chronic disease areas, for example, diabetes, where the existence of computerised disease registers (if available) makes some of the IT-based strategies possible when they are infeasible for the rest of that healthcare setting.

Practical feasibility issues may be more important in some cases than finance; for example, the problems of bringing GPs from a wide geographical area together for an educational meeting. Locum costs may be a significant expense for GP educational meetings, but increasing use of locums in general practice may create additional problems for continuity of patient care. In secondary care, a key feasibility issue is the opportunity cost of cancelling outpatient clinics or operating lists for group activities that cannot be contained within a lunch hour.

In general terms, there appear to be significantly higher resources available for all of the strategies in primary care. In secondary care, even distribution of educational materials may require external sponsorship to make it happen. However, there appeared to be some cultural differences between the secondary care settings, which suggested that strategies might be more or less feasible depending on the enthusiasm of lead clinicians, and how receptive their colleagues were.

Clearly, these findings are limited by the size of sample, the analysis to date and the selected nature of the convenience sample. As relative enthusiasts for the strategies under discussion, they may be thought to underestimate the difficulties inherent in implementation. For pragmatic reasons, other community and mental healthcare services were not represented in the survey and therefore no comment can be made on how generalisable any of the issues raised might be to these settings.

There did not appear to be any new themes being identified from primary care respondents by the end of their interviews. In secondary care, the variation in environment and culture identified (between specialities, teaching hospitals versus DGHs, etc.) suggested that a larger sample would be needed to obtain a comprehensive picture from these different settings. Nevertheless, the issues raised by the secondary care respondents still have important implications for clinical governance in many hospital settings.

The findings from this survey need to be placed in the context of developments in the NHS information strategy for England, particularly the publication of 'Information for Health' in 1998.[26] This laid out targets for achieving lifelong electronic health records for patients, held in primary care, and electronic patient records within hospital settings, before 2005. The respondents in this survey held contradictory views on whether these targets are achievable and whether, even if achieved, the resulting information systems would be able to support the computerised reminder strategies discussed here. The GPs with experience or knowledge of MIQUEST software for extracting data from primary care systems[25] or of using the PRODIGY[27] computerised reminder system were the most optimistic about the potential. Two of the secondary care respondents also thought that their own Trusts' plans for implementing the electronic patient record would be able to deliver the computerised reminder strategies. Views from the more pessimistic respondents ranged from, in primary care, the likelihood of failure while national strategy continues to encourage a 'mixed economy' of system suppliers rather than an NHS-wide information system; and, in secondary care, serious concerns about the feasibility of computerised medical records without widespread cultural change, training and, particularly, increases in medical staffing levels.

One of the notable features of the interviews was the creativity with which the respondents had pursued non-routine sources of funding for clinical effectiveness implementation initiatives; for example, bidding for Health Authority funds, accessing educational money and bidding for research and development (R&D) funding. The respondents may not be typical in this and the funding situation in most circumstances may actually be worse than described here. Even so, the non-recurrent nature of most funding earmarked for specific circumstances was apparent, and several respondents commented on the need to increase the general capacity of the system to allow a consistent expansion of educational activity in routine practice.

Chapter 7
Discussion

Principal findings of the study

Systematic review of guideline dissemination and implementation strategies

Overall, the majority of comparisons reporting dichotomous process data (86.6%) observed improvements in care suggesting that dissemination and implementation of guidelines can promote compliance with recommended practices. However, there was considerable variation in the observed effects both within and across interventions, although the majority of interventions observed modest to moderate improvements in care. The reviewers looked at the consistency of the direction and effect size of the results across the different study designs and across studies where statistical significance could be determined and studies where it could not. The conclusions are therefore tentative and need to be explored in future well-designed, robust evaluations.

Reminders are the single intervention that have been evaluated the most frequently. These results suggest that reminders are a potentially effective intervention and are likely to result in moderate improvements in process of care. The majority of comparisons evaluated reminders across a wide range of settings and targeted behaviours.

Educational outreach was the next most commonly evaluated intervention. It was often a component of a multifaceted intervention and might be considered to be inherently multifaceted. There were modest effects across the majority of studies. Combinations of educational materials and educational outreach appeared to be relatively ineffective. The results suggest that educational outreach may result in modest improvements in process of care, but this needs to be offset against both the resources required to achieve this change and practical considerations.

The evidence about the effectiveness of educational materials, audit and feedback and patient-directed interventions was less robust as there were fewer evaluations of these interventions. Educational materials and audit and feedback appeared to result in modest effects, whereas patient-directed interventions appeared to result in moderate effects. Nevertheless, if the median effect observed for educational materials as a single intervention could be achieved in routine practice, it would be important (especially given the feasibility and relatively low cost of disseminating educational materials; see below). However, the addition of educational materials to other interventions did not seem to increase effectiveness.

Although the majority of comparisons evaluated multifaceted interventions, there were few replications of specific multifaceted interventions either against a no-intervention control or against a specific control group. This raised considerable problems for the interpretation and analysis of the comparisons. The original plan was to undertake a meta-regression, but this was not possible. Although it would have been possible to undertake an analysis assuming that the effects of the various interventions were additive, this was judged to be unrealistic (see Appendix 1 for further details). Instead, all the results of the comparisons were described and the results summarised for multifaceted interventions including educational outreach and other multifaceted interventions with at least four comparisons. It was difficult to draw generalisable conclusions from comparisons of multifaceted interventions owing to the large number of different combinations evaluated. However, across all combinations, multifaceted interventions did not appear to be more effective than single interventions and the effects of multifaceted interventions did not appear to increase with the number of component interventions.

The majority of studies were conducted in the USA and the applicability of the results to other settings is uncertain. In general, there is poor understanding of the factors that may influence the applicability of implementation research findings.

The review highlights many of the methodological and reporting weaknesses of existing studies and highlights the importance of using well-applied robust designs when evaluating guideline dissemination and implementation strategies.

Reported details of the study interventions and contextual factors were poor and it was often difficult to assess the rationale for the choice of intervention. There was little description of the potential barriers and facilitators to practice. Many studies suffered from methodological weaknesses (e.g. unit of analysis errors) that created considerable uncertainties in synthesising and interpreting the results of the review.

Systematic review of economic evaluations and cost analyses in evaluations of guideline dissemination and implementation strategies

Policy makers need to have information about the likely benefits and costs of different guideline dissemination and implementation strategies if they are to make informed decisions about whether it is worthwhile to introduce guidelines. Despite this, only 29% (63) of studies reported any economic data. Eleven reported cost-effectiveness analyses, 38 reported cost–consequence analyses (where differences in cost were set against differences in several measures of effectiveness) and 14 reported cost analyses (where some aspect of cost was reported but not related to benefits). The majority of studies only reported costs of treatment; only 25 studies reported data on the costs of guideline development or guideline dissemination and implementation. The majority of studies used process measures for their primary end-point, despite the fact that only three guidelines were explicitly evidence based (and may not have been efficient).

Overall, the methods of the economic evaluations and cost analyses were poor. The viewpoint adopted in economic evaluations was stated in only ten studies. The methods to estimate costs were comprehensive in about half of the studies, and few studies reported details of resource use. Owing to the poor quality of reporting the studies, data on resource use and cost of guideline development, dissemination and implementation were not available for most of them. Only four studies provided sufficiently robust data for consideration. These studies demonstrated that the costs of local guideline development are not insubstantial if explicitly costed, although they recognised that the time needed for many activities is frequently not made explicit and such activities are often undertaken outside work time. Further estimates of the resources and costs of different methods of dissemination and implementation are needed before the generalisability of the reported costs can be determined. This is a relatively new area of work, with few publications discussing the methods of economic evaluation alongside implementation research trials.[18,19,28] Nevertheless, future studies need to undertake such analyses if they are to maximise their benefits to policy makers.

Key informant survey

The key informant survey, although limited in scope, raised many issues about the practicability and feasibility of using dissemination and implementation strategies in NHS settings. Respondents rarely identified existing budgets to support guideline dissemination and implementation strategies and made frequent comments about using 'soft' money or resources for specific initiatives to support such activities. Some respondents observed that there were non-monetary resources in the system that could be used. In general, the respondents thought that dissemination of educational materials and short (lunchtime) educational meetings were generally feasible within current resources (although primary care respondents were aware of the opportunity costs of inviting primary care staff to educational meetings). Respondents raised concerns about the feasibility of audit and feedback and reminders relating to the resources required to deliver these interventions through manual systems, and the poor availability and functionality of current IT systems required to deliver these interventions through computerised systems (especially in secondary care settings). Educational outreach was seen as attractive in primary care settings as it allowed primary care staff to remain in their practices, but both primary and secondary care respondents were unsure how this could be resourced in current NHS settings.

Strengths and weaknesses

The systematic review is the most extensive, methodologically sophisticated and up-to-date review of clinical guideline implementation strategies. The reviewers developed highly sensitive search strategies and consider the included studies to be a relatively complete set of studies for the period from 1966 to mid-1998. The searches detected more studies than previous reviews covering similar periods. For example, the *Effective Health Care* bulletin on implementing clinical guidelines[4] identified 91 studies from 1976 to 1994, whereas over the same period the current searches identified 165 studies. Potentially relevant studies that did not meet the inclusion criteria for this specific review have now been screened for the Cochrane EPOC register. The search strategy has

since been used to develop the EPOC strategy and future reviewers should be able to reduce search time by searching the EPOC register for relevant studies (available by contacting epoc@uottawa.ca).

Detailed data abstraction was conducted on the quality of the studies, characteristics of the studies and interventions. As a result, the methodological weaknesses of the primary studies have been made more explicit within this review. Where possible, attempts were made to reanalyse the results of comparisons with common methodological errors; for example, 27 ITS comparisons were reanalysed. In addition, the interventions were characterised in greater detail than in previous reviews; this revealed the multifaceted nature of the majority of evaluated interventions. Previous reviews have tended to describe interventions based on authors' main description of the intervention, which often ignores co-interventions.

Previous reviews have tended to use vote-counting methods of studies of statistically significant comparisons, often ignoring unit of analysis errors. These are potentially misleading owing to the uncertainty of statistical significance of comparisons with unit of analysis errors and the lack of statistical power of many comparisons. They also provide little information about the potential effect sizes of different interventions. In the present review, a more explicit framework was used that allowed the methodological quality of studies (based on design and presence of unit of analysis errors when interpreting the results of comparisons) to be considered. It also allowed for the provision of information about the potential effect size of interventions. In addition, the methods, assumptions and results of the review have been reported in great detail to allow readers to explore these further.

The review has a number of weaknesses. It is inevitable that some studies will have been missed; the search strategy was estimated to be around 93% sensitive, suggesting that around 17 studies may have been missed. The authors believe that the searches are complete up to mid-1998. Current estimates suggest that around 35–40 studies per year are being published, suggesting that the review is already out of date. There is also some evidence that newer studies had fewer methodological weaknesses, for example, potential unit of analysis errors. Studies that the reviewers considered to meet the Institute of Medicine definition of clinical guidelines[11] were included, irrespective of whether the term guideline was explicitly mentioned in the study. This was largely done for pragmatic reasons. Over time the use of the term guideline has changed, and many studies undertaken in the 1970s and 1980s targeted recommendations that would probably have been called guidelines if published today. These judgements inevitably included a degree of subjectivity. Attempts were made to minimise this by having two reviewers independently assess studies against the inclusion criteria and by resolving any disagreements in discussion with a third reviewer. The Cochrane EPOC taxonomy of interventions was used; this classification is based upon pragmatic broad descriptions of interventions, was relatively easy to apply and has been used in a wide range of reviews. However, it is possible that different ways of classifying interventions may have been more discriminatory (although the authors are unaware of any comparably developed taxonomies). The breadth of the review also mitigated against detailed examination of effect modifiers of individual interventions, and it was not possible to undertake a meta-regression as planned. It was originally planned to contact authors for additional information, but this was not possible within the resources of the review, given the unexpected size of the task. Finally, perhaps the most important limitation of the review is the quality of the primary studies, many of which had common methodological weaknesses that diminish the certainty of the conclusions.

This investigation also included the first systematic review of economic evaluations and cost analyses in evaluations of guideline dissemination and implementation strategies, using a standard checklist for appraising economic evaluations. The review only considered the studies identified for the review of effectiveness. It is possible that the reviewers missed economic evaluations published in secondary publications (even though the searches identified many secondary publications) or economic evaluations based upon modelling. The most striking observation from this review was the paucity of high-quality economic evaluations, which limits the ability to draw firm conclusions about the resources required by and the efficiency of different dissemination and implementation strategies.

To supplement the findings of both reviews, a small key informant survey was conducted to explore the feasibility and resource requirements to deliver interventions within the NHS. Although this was a small survey it demonstrated many of the practical constraints that decision-makers in the NHS face when considering different

dissemination and implementation strategies. Further research could usefully expand this survey to explore the generalisability of its findings.

Meaning of the study: possible mechanisms and implications for clinicians or policy makers

Comparison with previous systematic reviews

An overview of systematic reviews of professional behavioural change strategies published in 1999 identified 44 reviews covering a wide range of activities and interventions.[2] On the basis of these reviews, it was concluded that "Passive dissemination (for example, mailing educational materials to targeted clinicians) is generally ineffective and is unlikely to result in behaviour change when used alone; however this approach may be useful for raising awareness of the desired behaviour change. Active approaches are more likely to be effective but also likely to be more costly. Interventions of variable effectiveness include audit and feedback, and use of local opinion leaders. Generally effective strategies include educational outreach (for prescribing behaviour) and reminders. Multifaceted interventions based on assessment of potential barriers to change are more likely to be effective than single interventions."

This review has arrived at conclusions that may appear at odds with the overview and previous systematic reviews. For example, Freemantle and colleagues[29] reviewed 11 studies evaluating the effects of the dissemination of educational materials. They used vote-counting methods and excluded studies with unit of analysis errors. None of the comparisons using appropriate statistical techniques found statistically significant improvements in practice. In the current review, a more explicit analytical framework was used to explore the median and range of effect sizes in studies with and without unit of analysis errors. Four C-RCT comparisons reporting dichotomous process measures were identified. All of these observed improvements in care; the median effect was +8.1% (range +3.6 to +17%) absolute improvement in performance. Two comparisons had potential unit of analysis errors and the significance of one comparison could not be determined. The remaining comparison without a potential unit of analysis error observed an effect of +6% (NS). If the same analytical approach had been used here, the reviewers would have reached similar conclusions to Freemantle and colleagues. However, such an approach would have failed to detect that printed educational materials led to improvements in care across the majority of studies (albeit that the statistical significance of the majority of comparisons is uncertain). Based upon this review the authors would not conclude that printed educational materials are effective, given the methodological weaknesses of the primary studies. Instead, it was concluded that printed educational materials may lead to improvements in care and that policy makers should not dismiss printed educational materials given their possible effect, relative low cost and feasibility within NHS settings.

Thomson O'Brien and colleagues[30] undertook a systematic review of educational outreach using vote-counting methods, and concluded that there was support for "educational outreach visits combined with additional interventions to reduce inappropriate prescribing by physicians", but that the cost-effectiveness of this approach remained unclear. In the present review, the majority of studies evaluating multifaceted interventions including educational outreach detected positive directions of effect. Again, if the same methods of review had been used here, the reviewers would have come to similar conclusions to Thomson O'Brien and colleagues. However, the analytical approach adopted in the current review highlighted the modest effect sizes observed in the primary studies (median effect size of +6%, range –4 to 17.4% across 11 C-RCTs). Based upon this finding, the potential high cost of the intervention and concerns about the feasibility of this intervention in routine NHS settings, the authors are more cautious in their conclusions about the efficiency of this intervention. Since this review was carried out, two well-conducted, UK-based studies of educational outreach have been published. Hall and colleagues[31] observed no effect of educational outreach on prescribing. Freemantle and colleagues observed an effect of +3% improvement across four areas of prescribing.[32] Interestingly, despite the small effect size, an economic analysis[24] suggested that educational outreach was cost-effective in small general practices for certain drugs.

Previous reviews have also suggested that multifaceted interventions are more effective than single interventions, on the basis that they address multiple barriers to implementation. Davis and colleagues' review of continuing medical education strategies[13] concluded that multifaceted interventions were more likely to be effective. Wensing and colleagues[8] undertook a review of

the effectiveness of introducing guidelines in primary care settings; they concluded that multifaceted interventions combining more than one intervention tended to be more effective but might be more expensive. The specific details about how interventions were coded and the analytical method of these two reviews are unclear. In this review, all intervention components were coded and explicit methods were used to determine a single effect size for each study. The analysis suggested that the effectiveness of multifaceted interventions did not increase incrementally with the number of components. Few studies provided any explicit rationale or theoretical base for the choice of intervention. As a result, it was unclear whether researchers had an a priori rationale for the choice of components in multifaceted interventions based upon possible causal mechanisms, or whether the choice was based on a 'kitchen sink' approach. It is plausible that multifaceted interventions built upon a careful assessment of barriers and coherent theoretical base may be more effective than single interventions.

It is important not to overinterpret these results, as there were few head-to-head comparisons of the different interventions. In particular, it would be inappropriate to conclude that one intervention is likely to be more effective than another based upon indirect comparisons. There is probably a natural confounder, in that researchers chose interventions that they thought were likely to be effective within their study context; thus, there may be important differences in the context, barriers or targeted behaviour between studies that assessed, for example, printed educational materials and educational outreach.

Implications for policy makers

Guidelines are increasingly common in healthcare settings. Within the UK, local decision-makers in clinical governance structures need to consider how to respond to the increasing number of guidelines developed by national agencies such as NICE and SIGN. Ideally, the decision of whether and how to develop, disseminate and implement a clinical guideline should be based upon careful estimates of the costs and benefits of the dissemination and implementation strategy, and the costs and benefits of the resulting changes in patient care. Mason and colleagues[24] coined the term policy cost-effectiveness to describe this combination of treatment cost-effectiveness with the cost and magnitude of change achieved by an implementation method. They elegantly demonstrated that educational outreach visits were cost-effective to promote the uptake of angiotensin-converting enzyme inhibitors in heart failure, but were not cost-effective to promote the substitution of tricyclic antidepressants for selective serotonin reuptake inhibitors in UK primary care.

This review has demonstrated that there is not a robust, generalisable evidence base to provide estimates of the likely costs and benefits of different strategies. Further, for many areas of clinical care, robust evidence on the cost-effectiveness of clinical interventions is not available and few guidelines have incorporated economic considerations when deriving recommendations.[33] As a result, it is unlikely that decision-makers will be able to replicate the sophisticated modelling approach demonstrated by Mason and colleagues.[24] Nevertheless, decision-makers can use a similar framework to guide their decisions about how best to use the limited resources they have for clinical governance and related activities to maximise population benefits.

The following conclusions were drawn for policy makers.

- Given the limited resources available, it seems likely that decision-makers will need to prioritise which guidelines to disseminate and implement actively based upon considerations of local burden of disease, availability of effective and efficient healthcare interventions, and local evidence of current suboptimal performance. This will clearly be influenced by national priorities and initiatives such as National Service Frameworks (several key informant respondents noted that it is often easier to secure additional funds around national priorities).
- Decision-makers should also consider the availability (or otherwise) of high-quality, cost-conscious guidelines. If such a guideline is unavailable, the resources required to develop a robust guideline *de novo* are substantial.
- Decision-makers need to identify what resources are available to them to support dissemination and implementation strategies. While there may only be limited direct resources available within clinical governance budgets, there may be substantial additional resources that can be mobilised through partnership and coordination with other stakeholders (e.g. coordinated educational approaches with continuing medical education providers) or within the environment of the organisation (e.g.

information systems that can support the generation of reminders).
- The choice of guideline dissemination and implementation strategies should be based upon considerations of the barriers to and facilitators of change, and the likely costs and benefits of different strategies. Several authors have proposed rationales for choosing different interventions in the presence of different types of barriers and faciliatators (e.g. Grol,[34] Mittman and colleagues,[35] Grimshaw and Eccles[36] and Moulding and colleagues[37]). However, the empirical basis for these rationales is not strong. Further, the lack of explicit rationale for interventions and poor reporting of potential barriers and facilitators in current studies limits our understanding of the applicability of study results to other settings and problems.
- Decision-makers need to use considerable judgement about which interventions are most likely to be effective in any given circumstance and choose intervention(s) based upon consideration of the feasibility, costs and benefits that potentially effective interventions are likely to yield.
  - Wherever possible, interventions should include paper-based or computerised reminders.
  - It may be more efficient to use a cheaper, more feasible, but less effective intervention (e.g. passive dissemination of guidelines) than a more expensive but potentially more effective intervention.
  - Decision-makers also need to consider the resource implications associated with consequent changes in clinical practice and to assess their likely impact on different budgets.
- Decision-makers should assess the effects of any interventions, preferably using rigorous evaluative designs.

Unanswered questions and future research

This review highlights the fact that despite 30 years of research in this area, we still lack a robust, generalisable evidence base to inform decisions about strategies to promote the introduction of guidelines or other evidence-based messages into practice.

The UK Medical Research Council (MRC) recently proposed a framework for evaluating complex interventions (such as guideline dissemination and implementation strategies).[38] This suggests that the evaluation of complex interventions should follow a sequential approach, involving:

1. development of the theoretical basis for an intervention
2. definition of components of the intervention (using modelling or simulated techniques and qualitative methods)
3. exploratory studies to develop further the intervention and plan a definitive evaluative study (using a variety of methods)
4. a definitive evaluative study (using quantitative evaluative methods, predominantly randomised designs).

This recognises the benefits of establishing the theoretical basis of interventions and conducting exploratory studies to choose and refine interventions to minimise the number of costly 'definitive' RCTs. Most of the studies included within this systematic review could be considered definitive evaluations. Relatively little research has been conducted to develop the theoretical basis of healthcare professional and organisational behaviour and behavioural change. Few studies appear to have undertaken preparatory studies to define the intervention or demonstrate the feasibility of the planned definitive study.

The lack of a coherent theoretical basis for understanding professional and organisational behavioural change limits our ability to formulate hypotheses about which interventions are likely to be effective under different circumstances and hampers our understanding of the likely generalisability of the results of definitive trials. An important focus for future research should be to develop a better theoretical understanding of professional and organisational behavioural change. Ferlie and Shortell[39] suggested four levels at which interventions to improve the quality of healthcare might operate: the individual health professional, healthcare groups or teams, organisations providing healthcare (e.g. NHS trusts) and the larger healthcare system or environment in which individual organisations are embedded. A full scientific rationale for guideline dissemination and implementation interventions requires exploration of theories relevant to each of these four levels. A large number of educational, behavioural, social and organisational theories has proved to be useful to understand behaviour in a wide range of contexts. However, the applicability of these theories to healthcare professional and organisational behaviour has rarely been tested. Thus, an initial step in the development of the

theoretical basis of professional and organisational behaviour could involve testing the applicability of such theories in healthcare settings. Different types of theory will be relevant to interventions at different levels; for example, psychological theories will be more relevant to interventions directed at individuals and teams, and theories of organisational change will be more relevant to interventions directed at hospitals or trusts. In addition, further research is required into methods to optimise interventions before conducting definitive trials (equivalent to the modelling and exploratory trial phases of the MRC framework).

As in other areas of medical and social research, rigorous evaluations (mainly RCTs) will provide the most robust evidence about the effectiveness of guideline dissemination and implementation strategies because effects are likely to be modest, there is substantial risk of bias and we have poor understanding of the likely effect modifiers and confounders.[40,41] However, despite the considerable number of evaluations that have been undertaken in this area, we still lack a coherent evidence base to support decisions about which dissemination and implementation strategies to use. This is partly due to the poor methodological quality of the existing studies; for example, the statistical significance of many comparisons could not be determined from the published studies owing to common methodological weaknesses. Further rigorous evaluations of different dissemination and implementation strategies need to take place. These evaluations need to be methodologically robust (addressing the common methodological errors identified in the systematic review), incorporate economic evaluations and, wherever possible, explicitly test behavioural theories to develop further our theoretical understanding of factors influencing the effectiveness of guideline dissemination and implementation strategies.

The MRC framework represents an idealised sequential framework for developing and evaluating interventions. However, it is modelled on the stages of development of pharmaceuticals, which typically takes 15–20 years to work through each stage, requiring substantial investment of resources. It is unlikely that implementation researchers will have the time and resources required to work through each phase sequentially when evaluating guideline dissemination and implementation strategies. Given this, future research needs to adopt a variety of approaches to developing and testing theory and evaluating interventions.

Specific recommendations for further research

Further research is required in the theoretical, substantive and methodological domains.

Theoretical

- Further research is required to develop and validate a coherent theoretical framework of health professional and organisational behaviour and behavioural change to inform better the choice of interventions in research and service settings. Initially, research should evaluate the applicability of existing predictive theories relevant to a range of levels (e.g. individual, team, organisation, system) rather than attempt to develop theory *de novo*.

Substantive

- Further rigorous evaluations of different dissemination and implementation strategies are required that:
  - are methodologically robust (addressing the common methodological errors identified in the systematic review)
  - incorporate economic evaluations
  - wherever possible, explicitly test behavioural theories to develop further our theoretical understanding of factors influencing the effectiveness of guideline dissemination and implementation strategies. If this is not possible, then trials should include integrated process evaluations better to describe and illuminate their findings
  - evaluate a range of (theoretically based) interventions in head-to-head comparisons.
- Further rigorous evaluations are required of relatively cheap interventions such as printed educational materials that have the potential to be efficient.

Methodological

- Further research is required to develop methods to optimise interventions before definitive trials.
- Further research is required to improve the methods of conducting C-RCTs of complex interventions such as guideline dissemination and implementation strategies. The MRC has published a monograph[42] on the conduct of C-RCTs; however, there needs to be a wider understanding of these issues and their application within the research community.
- Although not strictly a recommendation for further research, there is a clear need for improved reporting of studies evaluating complex interventions such as guideline dissemination and implementation strategies. This would include areas such as how best to

provide detail on the choice and construction of interventions and the evaluation of effect modifiers.
- The results of this systematic review differ from previous reviews as a result of the methods used. Further research is required to establish optimal review methods for complex interventions such as guideline dissemination and implementation strategies. This should include issues such as:
  – how best to aggregate results across studies that have unit of analysis errors
  – how best to deal with categorising interventions.

Other issues

The research agenda is substantial and will require an interdisciplinary perspective using a variety of methodological approaches. This remains a relatively new area of health services research, with a distinct perspective and specific methodological challenges. It is important to develop expertise in this area through interdisciplinary programmatic activities. Further progress in this area will require ongoing support by a range of funding agencies. For example, the theoretical developments needed (equivalent to establishing 'the basic science of implementation') may best be supported by fundamental, as opposed to applied, health research funders. Further, these research issues cut across national and cultural differences in the practice and financing of healthcare, and their scope is such that no individual country's health services research programme or individual research group alone can systematically examine them in a comprehensive way. There are substantial opportunities for international collaboration to enhance our understanding in this area.

Acknowledgements

We are extremely grateful to Liz Grant for her excellent work and support in the preparation of this report.

We also wish to thank the key informants who responded to our semistructured telephone survey in Chapter 6 (Drs Phil Ayres, John Parr, Helen Rodgers, Peter Sanderson, Professors Colin Bradshaw, Gene Feder and John Wright).

The Health Services Research Unit is funded by the Chief Scientist Office of the Scottish Executive Health Department. The views expressed are those of the authors.

References

1. Woolf SH, Grol R, Hutchison A, Eccles M, Grimshaw J. Clinical practice guidelines. The potential benefits, limitations, and harms of clinical guidelines. *BMJ* 1999;**318**:527–30.
2. NHS Centre for Reviews and Dissemination. Getting evidence into practice. *Effective Health Care* 1999;**5**:1–16.
3. Lomas J. Words without action? The production, dissemination, and impact of consensus recommendations. *Annu Rev Public Health* 1991; **12**:41–65.
4. NHS Centre for Reviews and Dissemination and Nuffield Institute for Health. Implementing clinical guidelines. Can guidelines be used to improve clinical practice? *Effective Health Care* 1994;**1**:1–12.
5. Oxman A, Thomson MA, Davis DA, Haynes RB. No magic bullets: a systematic review of 102 trials of interventions to improve professional practice. *CMAJ* 1995;**153**:1423–31.
6. Davis DA, Taylor-Vaisey A. Translating guidelines into practice. A systematic review of theoretic concepts, practical experience and research evidence in the adoption of clinical practice guidelines. *CMAJ* 1997;**157**:408–16.
7. Worrall G, Chaulk P, Freake D. The effects of clinical practice guidelines on patient outcomes in primary care: a systematic review. *CMAJ* 1997;**156**:1705–12.
8. Wensing M, Van der Weijden T, Grol R. Implementing guidelines and innovations in general practice: which interventions are effective? *J Gen Pract* 1998;**48**:991–7.
9. Thomas L, Cullum N, McColl E, Rousseau N, Soutter J, Steen N. Clinical guidelines in nursing, midwifery and other professions allied to medicine (Cochrane Review). The Cochrane Library (Issue 1). Oxford: Update Software; 1999.
10. Bero L, Grilli R, Grimshaw JM, Mowatt G, Oxman A, Zwarenstein M. Cochrane Effective Practice and Organisation of Care Review Group (Cochrane Group Module). In Bero L, Grilli R, Grimshaw JM, Mowatt G, Oxman A, Zwarenstein M, editors. The Cochrane Library (Issue 3). Oxford: Update Software; 2001.
11. Institute of Medicine Committee on Clinical Practice Guidelines. Guidelines for clinical practice: from development to use. Washington, DC: National Academy Press; 1992.
12. Cook TD, Campbell DT. Quasi-experimentation: design and analysis issues for field settings. Chicago, IL: Rand McNally; 1979.
13. Davis DA, Thomson MA, Oxman AD, Haynes RB. Changing physician performance. A systematic review of the effect of continuing medical education strategies. *JAMA* 1995;**274**:700–5.
14. Whiting-O'Keefe QE, Henke C, Simborg DW. Choosing the correct unit of analysis in medical care experiments. *Med Care* 1984;**22**:1101–14.
15. Ukoumunne OC, Gulliford MC, Chinn S, Sterne JAC, Burney PGJ. Methods for evaluating area-wide and organisation based interventions in health and health care: a systematic review. *Health Technol Assess* 1999;**3**(5).
16. Bushman BJ. Vote counting methods in meta-analysis. In Cooper H, Hedges L, editors. The handbook of research synthesis. New York: Russell Sage Foundation; 1994; pp. 193–213.
17. Drummond M, O'Brien B, Stoddart G, Torrance GW. Methods of economic evaluation of health care programmes. Oxford: Oxford University Press; 1997.
18. McIntosh E. Economic evaluation of guideline implementation studies. In Makela M, Thorsen T, editors. Changing professional practice. Theory and practice of clinical guidelines implementation. Copenhagen: Danish Institute for Health Services Research and Development; 1999, pp. 77–98.
19. Mason J, Wood J, Freemantle N. Designing evaluations of interventions to change professional practice. *J Health Serv Res Policy* 1999;**4**:106–11.
20. Drummond MF, Jefferson TO, for the BMJ Working Party on guidelines for authors and peer-reviewers of economic submissions to the *British Medical Journal*. Guidelines for authors and peer-reviewers of economic submissions to the *British Medical Journal*. *BMJ* 1996;**313**:275–83.
21. Jefferson TO, Demicheli V, Vale L. Methodological reviews of economic evaluations in health care. *JAMA* 2002;**287**:2809–12.
22. Fergusson D, van Walraven C, Coyle D, Laupacis A. Economic evaluations of technologies to minimize perioperative transfusion: a systematic review of published studies. International Study of Perioperative Transfusion (ISPOT) investigators. *Transfus Med Rev* 1999;**13**:106–17.
23. Campbell MK, Mollison J, Grimshaw JM. Cluster trials in implementation research: estimation of intracluster correlation coefficients and sample size. *Stat Med* 2001;**20**:391–9.
24. Mason J, Freemantle N, Nazereth I, Eccles M, Haines A, Drummond M. When is it cost effective

to change the behaviour of health professionals? *JAMA* 2001;**286**:2988–92.

25. Neal RD, Heywood PL, Morley S. Real world data – retrieval and validation of consultation data from four general practices. *Fam Pract* 1996;**13**: 455–61.
26. Department of Health. An information strategy for the modern NHS 1998–2005. A national strategy for local implementation. London: Department of Health; 1998.
27. Purves IN. PRODIGY: implementing clinical guidelines using computers. *Br J Gen Pract* 1998; **48**:1552–3.
28. Sculpher M. Evaluating the cost-effectiveness of interventions designed to increase the utilisation of evidence-based guidelines. *Fam Pract* 2000;**17**:S26–31.
29. Freemantle N, Harvey EL, Wolf F, Grimshaw JM, Grilli R, Bero L. Printed educational materials: effects on professional practice and health care outcomes (Cochrane Review). *Cochrane Database of Systematic Reviews* 2003;**1**.
30. Thomson O'Brien MA, Oxman AD, Davis DA, Haynes RB, Freemantle N, Harvey EL. Educational outreach visits: effects on professional practice and health care outcomes (Cochrane Review). *Cochrane Database of Systematic Reviews* 2003;**1**.
31. Hall L, Eccles MP, Barton R, Steen N, Campbell M. Is untargeted outreach visiting in primary care effective? A pragmatic randomized controlled trial. *J Public Health Med* 2001;**23**:109–13.
32. Freemantle N, Nazereth I, Eccles M, Wood J, Haines A. Evidence-based OutReach trialists. A randomised controlled trial of the effect of educational outreach by community pharmacists on prescribing in UK general practice. *Br J Gen Pract* 2002;**52**:290–5.
33. Eccles M, Mason J. How to develop cost-conscious guidelines. *Health Technol Assess* 2001;**5**(16).
34. Grol R. Beliefs and evidence in changing clinical practice. *BMJ* 1997;**315**:418–21.
35. Mittman BS, Tonesk X, Jacobsen PD. Implementing clinical practice guidelines: social influence strategies and practitioner behavior change. *Quality Review Bulletin* 1992;**18**:413–22.
36. Grimshaw JM, Eccles MP. Clinical practice guidelines. In Haines A, Silagy C, editors. Evidence-based practice in primary care. London: BMJ Publishing; 2001, pp. 120–34.
37. Moulding NT, Silagy CA, Weller DP. A framework for effective management of change in clinical practice: dissemination and implementation of clinical practice guidelines. *Quality in Health Care* 1999;**8**:177–83.
38. Medical Research Council. A framework for development and evaluation of RCTs for complex interventions to improve health. MRC Health Services and Public Health Board discussion document. London: MRC; 2000.
39. Ferlie EB, Shortell SM. Improving the quality of health care in the United Kingdom and the United States: a framework for change. *Milbank Q* 2001;**79**:281–315.
40. Cochrane AL. Effectiveness and efficiency: random reflections on health services. London: Nuffield Provincial Hospitals Trust; 1979.
41. Campbell DT. Methodology and epistemology for social science: selected papers. In Overman ES, editor. The experimenting society. Chicago, IL: University of Chicago Press; 1988, pp. 290–314.
42. Medical Research Council. Cluster randomised trials: methodological and ethical considerations. MRC Clinical Trials Series. London: MRC; 2002.
43. Grimshaw J, Campbell M, Eccles M, Steen I. Experimental and quasi-experimental designs for evaluating guideline implementation strategies. *Fam Pract* 2000;**17**:S11–18.
44. Donner A, Birkett N, Buck C. Randomization by cluster. Sample size requirements and analysis. *Am J Epidemiol* 1981;**114**:906–14.
45. Mollison J, Simpson J, Campbell MK, Grimshaw JM. Comparison of analytical methods for cluster randomised trials: an example from a primary care setting. *J Epidemiol Biostat* 2000;**5**:339–46.
46. Draper N, Smith H. Applied regression analysis. New York: Wiley; 1981.
47. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ* 1994;**309**:1286–91.

This version of HTA monograph volume 8, number 6 does not include the 269 pages of appendices. This is to save download time from the HTA website.

The printed version of this monograph also excludes the appendices.

View/download the appendices (936 kbytes).

Health Technology Assessment Programme

Prioritisation Strategy Group

Members

Chair,
Professor Tom Walley, Director, NHS HTA Programme & Professor of Clinical Pharmacology, 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 & Professor of Clinical Pharmacology, University of Liverpool

Chair,
Professor Shah Ebrahim, Professor in Epidemiology of Ageing, Department of Social Medicine, University of Bristol, Canynge Hall, Whiteladies Road, Bristol

Deputy Chair,
Professor Jenny Hewison, Professor of Health Care Psychology, Academic Unit of Psychiatry and Behavioural Sciences, University of Leeds School of Medicine, Leeds

Professor Douglas Altman, Professor of Statistics in Medicine, Centre for Statistics in Medicine, Oxford University, Institute of Health Sciences, Cancer Research UK Medical Statistics Group, Headington, Oxford

Professor John Bond, Professor of Health Services Research, Centre for Health Services Research, University of Newcastle, School of Health Sciences, Newcastle upon Tyne

Professor John Brazier, Director of Health Economics, Sheffield Health Economics Group, School of Health & Related Research, University of Sheffield, ScHARR Regent Court, Sheffield

Dr Andrew Briggs, Public Health Career Scientist, Health Economics Research Centre, University of Oxford, Institute of Health Sciences, Oxford

Dr Christine Clark, Medical Writer & Consultant Pharmacist, Cloudside, Rossendale, Lancs
and
Principal Research Fellow, Clinical Therapeutics in the School of Pharmacy, Bradford University, Bradford

Professor Nicky Cullum, Director of Centre for Evidence Based Nursing, Department of Health Sciences, University of York, Research Section, Seebohm Rowntree Building, Heslington, York

Dr Andrew Farmer, Senior Lecturer in General Practice, Department of Primary Health Care, University of Oxford, Institute of Health Sciences, Headington, Oxford

Professor Fiona J Gilbert, Professor of Radiology, Department of Radiology, University of Aberdeen, Lilian Sutton Building, Foresterhill, Aberdeen

Professor Adrian Grant, Director, Health Services Research Unit, University of Aberdeen, Drew Kay Wing, Polwarth Building, Foresterhill, Aberdeen

Professor Alastair Gray, Director, Health Economics Research Centre, University of Oxford, Institute of Health Sciences, Headington, Oxford

Professor Mark Haggard, Director, MRC ESS Team, CBU Elsworth House, Addenbrooke's Hospital, Cambridge

Professor F D Richard Hobbs, Professor of Primary Care & General Practice, Department of Primary Care & General Practice, University of Birmingham, Primary Care and Clinical Sciences Building, Edgbaston, Birmingham

Professor Peter Jones, Head of Department, University Department of Psychiatry, University of Cambridge, Addenbrooke's Hospital, Cambridge

Professor Sallie Lamb, Research Professor in Physiotherapy/Co-Director, Interdisciplinary Research Centre in Health, Coventry University, Coventry

Dr Donna Lamping, Senior Lecturer, Health Services Research Unit, Public Health and Policy, London School of Hygiene and Tropical Medicine, London

Professor David Neal, Professor of Surgical Oncology, Oncology Centre, Addenbrooke's Hospital, Cambridge

Professor Tim Peters, Professor of Primary Care Health Services Research, Division of Primary Health Care, University of Bristol, Cotham House, Cotham Hill, Bristol

Professor Ian Roberts, Professor of Epidemiology & Public Health, Intervention Research Unit, London School of Hygiene and Tropical Medicine, London

Professor Peter Sandercock, Professor of Medical Neurology, Department of Clinical Neurosciences, University of Edinburgh, Western General Hospital NHS Trust, Bramwell Dott Building, Edinburgh

Professor Martin Severs, Professor in Elderly Health Care, Portsmouth Institute of Medicine, Health & Social Care, St George's Building, Portsmouth

Dr Jonathan Shapiro, Senior Fellow, Health Services Management Centre, Park House, Birmingham

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

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

Dr Andrew Farmer, Senior Lecturer in General Practice, Institute of Health Sciences, University of Oxford

Dr Karen N Foster, Clinical Lecturer, Dept of General Practice & Primary Care, University of Aberdeen

Professor Jane Franklyn, Professor of Medicine, University of Birmingham

Professor Antony J Franks, Deputy Medical Director, The Leeds Teaching Hospitals NHS Trust

Mr Tam Fry, Honorary Chairman, Child Growth Foundation, London

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, Devon

Mr Tony Tester, Chief Officer, South Bedfordshire Community Health Council, Luton

Dr Andrew Walker, Senior Lecturer in Health Economics, University of Glasgow

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 Iain T Cameron, Professor of Obstetrics & Gynaecology, University of Southampton

Mr Peter Cardy, Chief Executive, Macmillan Cancer Relief, London

Dr Christopher Cates, GP and Cochrane Editor, Bushey Health Centre, Bushey, Herts.

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

Professor Alastair Gray, Professor of Health Economics, Institute of Health Sciences, University of Oxford

Mrs Sharon Hart, Managing Editor, *Drug & Therapeutics Bulletin*, London

Dr Christine Hine, Consultant in Public Health Medicine, Bristol South & West Primary Care Trust

Professor Robert Peveler, Professor of Liaison Psychiatry, Royal South Hants Hospital, Southampton

Dr Frances Rotblat, CPMP Delegate, Medicines Control Agency, London

Mrs Katrina Simister, New Products Manager, National Prescribing Centre, Liverpool

Dr Ken Stein, Senior Lecturer in Public Health, University of Exeter

Professor Terence Stephenson, Professor of Child Health, University of Nottingham

Dr Richard Tiner, Medical Director, Association of the British Pharmaceutical Industry, London

Professor Dame Jenifer Wilson-Barnett, Head of Florence Nightingale School of Nursing & Midwifery, King's College, London

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

Professor John Bond, Head of Centre for Health Services Research, University of Newcastle upon Tyne

Mr Michael Clancy, Consultant in A & E Medicine, Southampton General Hospital

Dr Carl E Counsell, Senior Lecturer in Neurology, University of Aberdeen

Dr Keith Dodd, Consultant Paediatrician, Derbyshire Children's Hospital, Derby

Professor Gene Feder, Professor of Primary Care R&D, Barts & the London, Queen Mary's School of Medicine and Dentistry, University of London

Ms Bec Hanley, Freelance Consumer Advocate, Hurstpierpoint, West Sussex

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

Mr George Levvy, Chief Executive, Motor Neurone Disease Association, Northampton

Professor James Lindesay, Professor of Psychiatry for the Elderly, University of Leicester

Dr Mike McGovern, Senior Medical Officer, Heart Team, Department of Health, London

Dr John C Pounsford, Consultant Physician, North Bristol NHS Trust

Professor Mark Sculpher, Professor of Health Economics, Institute for Research in the Social Services, University of York

Dr L David Smith, Consultant Cardiologist, Royal Devon & Exeter Hospital

Professor Norman Waugh, Professor of Public Health, University of Aberdeen

Expert Advisory Network

Members

Mr Gordon Aylward,
Chief Executive,
Association of British Health-Care Industries, London

Ms Judith Brodie,
Head of Cancer Support Service, Cancer BACUP, London

Mr Shaun Brogan,
Chief Executive, Ridgeway Primary Care Group, Aylesbury, Bucks

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 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

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, West Sussex

Ms Grace Gibbs,
Deputy Chief Executive, Director for Nursing, Midwifery & Clinical Support Servs., West Middlesex University Hospital, Isleworth, Middlesex

Dr Neville Goodman,
Consultant Anaesthetist, Southmead Hospital, Bristol

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

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, Dorset

Professor Alistair McGuire,
Professor of Health Economics, London School of Economics

Dr Peter Moore,
Freelance Science Writer, Ashtead, Surrey

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, Surrey

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 Chris Price,
Visiting Chair – Oxford, Clinical Research, Bayer Diagnostics Europe, Cirencester

Ms Marianne Rigge,
Director, College of Health, London

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