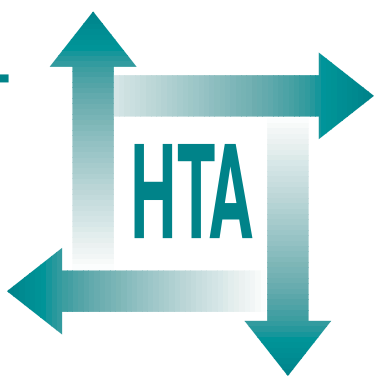
Health Technology Assessment 2000; Vol. 4: No. 22

Methodology

Using routine data to complement and enhance the results of randomised controlled trials

JD Lewsey
AH Leyland
GD Murray
FA Boddy

Health Technology Assessment
NHS R&D HTA Programme



Standing Group on Health Technology

Current members

Chair:
Professor Kent Woods
Professor of Therapeutics,
University of Leicester

Professor Martin Buxton
Director & Professor of
Health Economics,
Health Economics
Research Group,
Brunel University

Professor Shah Ebrahim
Professor of Epidemiology
of Ageing, University of Bristol

Professor Francis H Creed
Professor of
Psychological Medicine,
Manchester Royal Infirmary

Professor John Gabbay
Director, Wessex Institute
for Health Research
& Development

Professor Sir John
Grimley Evans
Professor of Clinical Geratology,
Radcliffe Infirmary,
Oxford

Dr Tony Hope
Clinical Reader in Medicine,
Nuffield Department of
Clinical Medicine,
University of Oxford

Professor Richard Lilford
Regional Director of R&D,
NHS Executive West Midlands

Dr Jeremy Metters
Deputy Chief Medical Officer,
Department of Health

Professor Maggie Pearson
Regional Director of R&D,
NHS Executive North West

Mr Hugh Ross
Chief Executive,
The United Bristol
Healthcare NHS Trust

Professor Trevor Sheldon
Joint Director, York Health
Policy Group, University of York

Professor Mike Smith
Faculty Dean of Research
for Medicine, Dentistry,
Psychology & Health,
University of Leeds

Dr John Tripp
Senior Lecturer in Child
Health, Royal Devon and Exeter
Healthcare NHS Trust

Professor Tom Walley
Director,
Prescribing Research Group,
University of Liverpool

Dr Julie Woodin
Chief Executive,
Nottingham Health Authority

Past members

Professor Sir Miles Irving*
Professor of Surgery,
University of Manchester,
Hope Hospital, Salford

Dr Sheila Adam
Department of Health

Professor Angela Coulter
Director, King's Fund,
London

Professor Anthony Culyer
Deputy Vice-Chancellor,
University of York

Dr Peter Doyle
Executive Director, Zeneca Ltd,
ACOST Committee on Medical
Research & Health

Professor John Farndon
Professor of Surgery,
University of Bristol

Professor Charles Florey
Department of Epidemiology
& Public Health, Ninewells
Hospital & Medical School,
University of Dundee

Professor Howard
Glennester
Professor of Social Science
& Administration, London
School of Economics &
Political Science

Mr John H James
Chief Executive,
Kensington, Chelsea &
Westminster Health Authority

Professor Michael Maisey
Professor of
Radiological Sciences,
Guy's, King's & St Thomas's
School of Medicine & Dentistry,
London

Mrs Gloria Oates
Chief Executive,
Oldham NHS Trust

Dr George Poste
Chief Science & Technology
Officer, SmithKline Beecham

Professor Michael Rawlins
Wolfson Unit of
Clinical Pharmacology,
University of Newcastle-
upon-Tyne

Professor Martin Roland
Professor of General Practice,
University of Manchester

Professor Ian Russell
Department of Health Sciences
& Clinical Evaluation,
University of York

Dr Charles Swan
Consultant Gastroenterologist,
North Staffordshire
Royal Infirmary

* Previous Chair

Details of the membership of the HTA panels, the NCCHTA Advisory Group and the HTA Commissioning Board are given at the end of this report.

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.

Using routine data to complement and enhance the results of randomised controlled trials

JD Lewsey[1]
AH Leyland[1*]
GD Murray[2]
FA Boddy[1]

[1] Public Health Research Unit, University of Glasgow, UK
[2] Medical Statistics Unit, University of Edinburgh, UK

[*] Corresponding author

Competing interests: none declared.

Published October 2000

This report should be referenced as follows:

Lewsey JD, Leyland AH, Murray GD, Boddy FA. Using routine data to complement and enhance the results of randomised controlled trials. *Health Technol Assess* 2000;**4**(22).

Health Technology Assessment is indexed in *Index Medicus*/MEDLINE and *Excerpta Medica*/EMBASE. Copies of the Executive Summaries are available from the NCCHTA website (see overleaf).

NHS R&D HTA Programme

The overall aim of the NHS R&D Health Technology Assessment (HTA) programme is to ensure that high-quality research information on the costs, effectiveness and broader impact of health technologies is produced in the most efficient way for those who use, manage and work in the NHS. Research is undertaken in those areas where the evidence will lead to the greatest benefits to patients, either through improved patient outcomes or the most efficient use of NHS resources.

The Standing Group on Health Technology advises on national priorities for health technology assessment. Six advisory panels assist the Standing Group in identifying and prioritising projects. These priorities are then considered by the HTA Commissioning Board supported by the National Coordinating Centre for HTA (NCCHTA).

This report is one of a series covering acute care, diagnostics and imaging, methodology, pharmaceuticals, population screening, and primary and community care. It was identified as a priority by the Methodology Group and funded as project number 94/06/01.

The views expressed in this publication are those of the authors and not necessarily those of the Standing Group, the Commissioning Board, the Panel members 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 the recommendations for policy contained herein. In particular, policy options in the area of screening will be considered by the National Screening Committee. This Committee, chaired by the Chief Medical Officer, will take into account the views expressed here, further available evidence and other relevant considerations.

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.

Criteria for inclusion in the HTA monograph series
Reports are published in the HTA monograph series if (1) they have resulted from work either prioritised by the Standing Group on Health Technology, or otherwise commissioned for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.

Series Editors: Andrew Stevens, Ken Stein and John Gabbay
Monograph Editorial Manager: Melanie Corris

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 review. They would like to thank the referees for their constructive comments on the draft document.

ISSN 1366-5278

Published by Core Research, Alton, on behalf of the NCCHTA.
Printed on acid-free paper in the UK by The Basingstoke Press, Basingstoke.

Copies of this report can be obtained from:

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

Contents

List of abbreviations

CABG	coronary artery bypass grafting
CI	confidence interval
CT	computed tomography
GCS	Glasgow Coma Score
ICD	International Classification of Diseases
MI	myocardial infarction
OPCS	Office of Population Censuses and Surveys
OPEN	open prostatectomy
OR	odds ratio*
PTCA	percutaneous transluminal coronary angioplasty
RCT	randomised controlled trial
SAH	subarachnoid haemorrhage
SMR1	Scottish Morbidity Record Form 1
TURP	transurethral prostatectomy
WFNS	World Federation of Neurosurgeons

* Used only in tables

Executive summary

Background

Randomised controlled trials (RCTs) are widely accepted as the best way to assess the outcomes and safety of medical interventions, but are sometimes not ethical, not feasible, or limited in the generalisability of their results. In such circumstances, routinely available data could help in several ways. Routine data could be used, for example, to conduct 'pseudo-trials', to estimate likely outcomes and required sample size to help design and conduct trials, or to examine whether the expected outcomes observed in an RCT will be realised in the general population.

Objectives

The project was undertaken to explore how routinely assembled hospital data might complement or supplement RCTs to evaluate medical interventions:

- in contexts where RCTs are not feasible for defining the context and design of an RCT
- for assessing whether the benefits indicated by RCTs are achieved in wider clinical practice.

Methods

The project was based on the system of linked Scottish morbidity records, which cover 100% of acute hospital care episodes and statutory death records from 1981 to 1995. Three case studies were undertaken as a way of investigating the utility of these records in different applications.

First, an attempt was made to analyse the link between the timing of surgery for subarachnoid haemorrhage (SAH) and subsequent outcomes (a question not easily susceptible to RCT design). A subsample was derived by excluding patients for which a diagnosis of SAH may not have been established or that may not have been admitted to a neurosurgical unit, and the data were assessed to attempt to inform the design of a trial of early versus late surgery.

Transurethral prostatectomy (TURP), the second case study, has become the surgery of choice for benign prostatic hyperplasia without systematic assessment of its effectiveness and safety, and an RCT would now be considered unethical. However, there is a need to investigate long-term effects and the influence of co-morbidities on outcomes. A retrospective comparison of mortality and re-operation following either open prostatectomy (OPEN) or TURP was, therefore, undertaken. Patients for whom it was not possible to establish the initial procedure were excluded.

The third case study compared coronary artery bypass grafting (CABG) with percutaneous transluminal angioplasty (PTCA) for coronary revascularisation. RCTs have been conducted in limited patient subgroups with short follow-up periods. A meta-analysis of RCTs could be augmented by routine data, which are available for large populations. This would allow assessment of subgroup effects, and outcomes over a long period. A subgroup of patients was therefore constructed for whom relevant routine data were available and who reflected the entry criteria for major RCTs, thus enabling a comparison between the results expected from this subgroup and those of the general population.

Results and conclusions

The uses of routine data in these contexts had strengths and weaknesses. The SAH study suggested a means of assessing outcomes and survival rates following haemorrhage, which could have value in informing the design of more precise trials and in evaluating changes in outcome following the introduction of new treatments such as embolisation. However, the potential of the data was not realised because their scope and content were insufficient. For example, lack of data on the time of onset of symptoms and patients' conditions at hospital admission made it difficult to establish the link between timing of surgery and the outcome, and there was insufficient information on patients' conditions at discharge to enable a comparison of outcomes.

The prostatectomy study was able to address questions not answered by RCT literature because the large number of cases it included allowed exploration of subgroup effects. The data indicated that younger patients and those with previous hospital admissions for cardiovascular, respiratory or ischaemic disease were more likely to have TURP, suggesting that these may influence treatment decisions. However, the risk of re-operation was higher in patients who initially underwent TURP, and, although mortality at 90 days was higher in patients who had OPEN initially, the difference seen from the routine data was not significant at 1 and 5 years. The records for this study were more satisfactory than for the SAH study. However, lack of data on the severity and complexity of patients' conditions limited the potential of the data.

The study of coronary revascularisation supported findings of the earlier meta-analysis, but with more prolonged follow-up and a broader population. Of the three studies, the data for this study were the most satisfactory, although lack of precise information on the complexity and severity of patients' conditions made it difficult to establish the full extent of subgroups. Patients who had an initial PTCA were more likely to require re-intervention than those who had CABG and, as expected, there was a lower rate of death and myocardial infarction (MI) in the RCT-like subgroup than in patients excluded from this sample. Using the routine data, the rates of death and MI at 1 year were significantly higher in patients who had an initial CABG, whereas this difference was not significant in the RCTs, but the difference was not significant in both at 5 years. A Bayesian comparison of the two interventions illustrated that Bayesian analyses can provide a link between RCTs, which are unbiased by design but may not reflect real populations, and routine data, which reflect reality but may be biased. This can facilitate better evaluations of outcomes associated with new technologies.

In general, linked data have value in two main ways. First, they relate to complete populations of cases and might thus clarify issues relating to patient selection. Second, by linking episodes of care to each other and to deaths, it is possible to gain information about prior medical histories, longer-term outcomes and the place of treatment, providing a context for more focused RCTs and multicentre comparisons of new techniques and their outcomes. Both of these are probably a prerequisite for comparable work; however, the shortcomings in the content and quality of current data limit these applications. Indeed, three further intended studies – laparoscopic versus open cholecystectomy, modes of treatment for breast cancer, and colorectal surgery by specialist versus generalist surgeons – proved impossible to undertake because of inadequacies in the routine data.

Implications for healthcare and recommendations for research

The shortfalls in the available data appear to be related to the largely administrative uses of the data at present. As the NHS moves closer to clinical governance, and as clinical audit develops, there are strong arguments for increasing the potential of routine data systems to complement information provided by RCTs. Ways in which the data systems might be improved include:

- publication of audits and feedback to those recording the data
- a stronger link between providers of care and those who generate and use routine records to monitor the extent to which the data reflect clinical practices
- expanding and updating coding systems to reflect new procedures and treatments
- adding more focused data collection to records for current clinical interests as a way of answering predetermined questions.

Chapter 1

Introduction

The randomised controlled trial (RCT) is widely acknowledged as the best available tool for evaluating the risks and benefits of medical interventions. There are many situations, however, where ethical or other more pragmatic reasons mean that it is not feasible to conduct an RCT: despite largely unexplained differences in rates between consultants or obstetric units, a randomised trial of delivery by Caesarean section would be considered unethical.[1,2] Similarly, a trial investigating an uncommon outcome or side-effect (such as common bile duct injury in laparoscopic cholecystectomy, or deep vein thrombosis in low-risk surgical groups) would require very large studies to detect clinically relevant differences, and their costs would be difficult to justify in terms of their potential benefits.

In these circumstances, it can be argued that routinely assembled data – usually those abstracted from patient records – could be used either to complement or supplement the RCT as a way of evaluating medical interventions. In brief, these uses have three main applications:

(i) in contexts where RCTs are not feasible
(ii) for defining the context and design of an RCT
(iii) for assessing whether the benefits indicated by RCT findings are achieved in later clinical practice.

When trials are feasible, their design is generally based on specifying their desired power and significance levels. Routine data may allow more reliable estimates of the treatment effect likely to be observed in a clinical trial, with the consequence that power calculations will be more accurate or realistic. Estimating the 'patient horizon' (the number of present and future patients, including those taking part in the trial) can only be done realistically by using routine data; doing so has implications for the estimation of sample size and for the method of treatment assignment.[3]

When trials have been completed, important questions are often left unanswered. The use of angiotensin converting enzyme inhibitors following an acute myocardial infarction (MI) is arguably one of the most comprehensively researched questions in clinical medicine, and yet it is still unclear whether the use of these drugs should be restricted to 'high-risk' patients. Equally, it is not clear how soon treatment should be initiated after an infarct, or for how long it should be continued.[4,5] There are other data which show that, in spite of overwhelming evidence of their efficacy, the uptake of these drugs is very low.[6,7]

There is thus a need to augment information from clinical trials: this project has explored the uses of routine NHS data for these purposes. We first examined the use of routine data as a way of 'sharpening' the treatment comparisons derived from RCTs, either by informing the trial design, or – more directly – by using routine data to estimate treatment effects. Second, we investigated the use of routine data to estimate the impact of RCTs on clinical practice and patient outcomes.

Sharpening treatment comparisons

Although RCTs may be the only unambiguous method of equalising risk among groups of patients receiving different treatments,[8] ethical or financial constraints may mean that RCTs are not feasible. In these situations, forming pseudo-trials from routine data is a possible alternative. Such an approach has its limitations and difficulties: routine data sets have generally been assembled for other purposes and may omit potentially confounding variables, such as earlier medical history or the severity of disease. The 'allocation' of patients to treatments will tend to reflect other factors whose effects cannot be fully captured in a covariate adjustment.[9]

Routine data can, however, be used as a supplement to clinical trials in subsequent Bayesian analyses. This may be for a specific trial when the data provide prior information regarding expected outcomes for one or more of the treatment groups, or during subsequent meta-analyses when the routine data can be regarded as comprising a large single or

multicentre trial. Such analyses take the form of a weighted average of the information provided by the routine data set and that provided by the trial. The weighting then depends on both the numbers involved (favouring routine data) and the likelihood of bias that will favour RCT data.[10]

The use of routine data before a trial begins can provide estimates of likely event rates or even of treatment effects. They thus have value when performing power calculations, estimating sample sizes or deciding when to stop a trial. The decision to stop or continue a trial is made by weighing the consequences of the possible actions, averaging over the distribution of future observations.[11] Observational data can then be used in conjunction with trial data to improve estimates of the probability that an apparent difference occurred by chance, leading to the trial either being stopped early or continued until decisive evidence has been accumulated.

The impact and uptake of new technologies

After the initial introduction of a new technology, there is a period before it becomes a recommended treatment. Its use may be modified or refined as clinicians gain experience with it. Little is known about the rate of uptake of new technologies and the factors that influence this process. Fewer than 10% of hospitals in Pennsylvania were using laparoscopic cholecystectomy 1 year after the introduction of the technique; after a further year, the proportion had increased to 80%.[12]

When an RCT recommends the uptake of a new technology, it anticipates an expected improvement in outcomes. Routine data make it possible to examine observed outcomes and assess whether the expected benefits have been realised. These may be in terms of the difference between alternative treatments (current and historical), as differences between treatment centres, or between subgroups of patients. If a treatment is recommended for a particular subpopulation then it is possible to assess the extent to which appropriate patients have been targeted, whether the benefits are also seen in other subpopulations and whether there has been a change in trends in outcomes (such as survival rates). Large routine data sets also permit assessments of the importance of an RCT result, particularly for rare conditions or when rare outcomes are being considered.

Case studies

The three case studies included in this report illustrate the potential of routine data as an alternative to RCTs in the evaluation of a new technology when an RCT was not feasible (the example of subarachnoid haemorrhage (SAH)), for a comparison of two established treatments (the example of prostatic surgery) and for comparing trial results with those achieved in clinical practice (the example of surgery for coronary artery disease). In a rather different sense, the three studies provide a basis for assessing the utility of routine data, first, as a supplement to evaluations employing RCTs and, second, as the means of understanding RCT findings in the context of their impact on outcomes in the populations to whom they are applied. It will be evident that a central feature of this assessment concerns the adequacy and competence of routine data when used for these purposes, and the improvements in them that would be necessary if they were to be employed in this way in the future. Specific aspects of this issue are described in the account of the case studies themselves; our more general conclusions are summarised in the discussion section of the report.

Chapter 2

Linked Scottish morbidity records

The data for the case studies were taken from the linked hospital discharge and mortality data set created by the Information and Statistics Division of the NHS in Scotland.[13] Summaries of hospital inpatient episodes are completed on Scottish Morbidity Record Form 1 (SMR1) for 100% of non-psychiatric and non-obstetric inpatients and day cases in Scotland, and provide information about the background of the patient as well as detailing case management, diagnoses and surgical procedures. The data are linked by individuals and to statutory death records from the Registrar General for Scotland and registrations in the Scottish Cancer Register. At present, they provide information about hospital care and mortality for individuals in the whole Scottish population for the period 1981–1995 and are unique in the UK. Further linkage (up to 1998) is in progress. Because the SMR1 and death records both include the individual's postcode sector of residence, they can be related to population denominators and to sociodemographic area descriptions derived from the 1991 census. The content of these records is described in *Box 1*.[14,15]

The selection of case studies

The scope of the data meant that we were restricted to considering surgical interventions because other interventions (such as drug therapies) are not recorded. Although desirable in the light of current debate, it was not possible, for example, to use the linked SMR1 database

BOX 1 The content of Scottish Morbidity Record 1 (SMR1)

Data item	Comment
Patient 'identifier'	Does not permit identification of individual patients; a unique number allowing linkage of hospital discharge, death and cancer registry records
Continuous inpatient stay marker	SMR1 forms are completed for each episode of care; a patient may have several episodes within a period of hospital stay
Hospital code	The identifier for the hospital attended
Area of residence	Health Board, Local Government District, postcode sector and census enumeration district
Age	The patient's age in months
Sex	
Marital status	
Admitted from	Home, other NHS hospital, other unit in this hospital, other
Type of admission	Deferred, waiting list/diary/booked, repeat admission, transfer, emergency
Wait	Days on the waiting list
Discharge code	Irregular (e.g. self discharge), home, convalescent hospital or nursing home, other hospital, local authority care, transfer to other specialty in the same hospital, died
Date of admission	
Date of discharge	
Date of operation	
Length of stay	
Specialty	
Diagnoses	Up to six diagnoses, coded to four-digit ICD-9 (1981—1995)[14]
Operation	Up to four procedures, coded to OPCS4[15]

ICD: International Classification of Diseases; OPCS: Office of Population Censuses and Surveys

to investigate the uptake of thrombolysis for the treatment of acute MI.

Given this restriction we chose, first, to study the timing of surgery following aneurysmal SAH as a way of informing the design of a trial of early versus late surgery. The second study compared open prostatectomy (OPEN) with transurethral prostatectomy (TURP) for the treatment of benign prostatic hyperplasia. The third investigated percutaneous coronary angioplasty (PTCA) with coronary artery bypass grafting (CABG) as approaches to coronary revascularisation. This study illustrates the use of routine data to augment a meta-analysis of RCT data, and explored the uptake and impact of the two procedures. Between them the three studies illustrate aspects of the uses of routine data sketched above. A number of other studies were proposed in the original grant application, namely a comparison of laparoscopic with open surgery for cholecystectomy and hernia repair; a comparison of modes of treatment for female breast cancer; and a comparison of outcomes following surgery for colorectal cancer performed by general surgeons and by specialists. For reasons noted at the end of the case study reports, it was not possible to undertake this last group of investigations.

Chapter 3

Case study 1: The timing of surgery after SAH

Spontaneous aneurysmal SAH is a common and often devastating event: there are approximately 6000 cases each year in the UK with a case fatality of the order of 40%, and a significant residual morbidity in approximately 50% of survivors.[16–19] In about two thirds of cases the haemorrhage results from the rupture of an aneurysm on a major cerebral blood vessel; and in 5% from an arteriovenous malformation. Even after complete angiography, no source of bleeding can be identified in about 25% of cases.[17]

Epidemiology

Epidemiological reports have suggested an annual incidence of the order of 10 to 15 cases per 100,000 of the population, although more recent reports have estimated an annual incidence in the range of 6–8 cases per 100,000. A recent meta-analysis[18] has reviewed this literature and has shown that the apparent fall in incidence can be explained by more precise diagnosis, following the wider availability of computed tomography (CT) scanning. SAH is more common in women (unselected series are typically 60% female). There is strong evidence that genetic factors predispose individuals to the risk of SAH, and some populations have much higher incidence rates: Finland, for example, has an incidence rate that is almost three times that observed in other parts of the world.[18]

Clinical presentation

SAH can occur at any time, with no obvious association with physical exertion. With a major bleed, consciousness is lost rapidly, and death can follow within minutes or hours. Smaller leaks lead to a severe headache with a characteristic rapid onset, and are almost always followed by vomiting. Neck stiffness can take 24–48 hours to appear. Consciousness is often impaired, but the illness can be minor and mistaken for influenza or migraine. More severe cases with headache, vomiting and neck stiffness can be misdiagnosed as meningitis, resulting in admission to an infectious disease unit.

A number of different grading systems have been devised to measure a patient's clinical state following SAH; the most widely accepted of which is now the World Federation of Neurosurgeons (WFNS) Grade.[20] This is based on the patient's Glasgow Coma Score[21] (GCS) and the presence or absence of a motor deficit (*Table 1*).

TABLE 1 *The WFNS subarachnoid scale*

WFNS grade	GCS score	Motor deficit
I	15	Absent
II	13–14	Absent
III	13–14	Present
IV	7–12	Present or absent
V	3–6	Present or absent

Despite the characteristic history, SAH is commonly misdiagnosed.[22–24] This can have serious consequences for the patient, because a minor 'warning leak' often precedes a potentially devastating rupture which could be prevented by early intervention.[17]

Diagnosis

CT scanning is the key initial investigation: if the scan is performed within 24 hours of the onset of symptoms then a high density clot in the subarachnoid space will be identified in over 90% of cases.[17] Diagnostic sensitivity declines progressively after the first day. A negative CT scan should be followed by a diagnostic lumbar puncture; if this is also negative then the diagnosis of a warning leak is effectively excluded and the prognosis is excellent. If a positive diagnosis of SAH is reached then cerebral angiography is the next standard investigation to identify the source of the haemorrhage. However, some 20–25% of such angiograms will fail to reveal any source of bleeding.

Complications

Further bleeding

Without surgical intervention, approximately 20% of patients will have further haemorrhage within 2 weeks, and 60% within 6 months. Of those who bleed again, some two-thirds die as a consequence, therefore, reducing this risk is a high priority in the management of SAH.

Cerebral infarction

Cerebral infarction is the other major cause of mortality and morbidity following SAH, apart from the local damage associated with the haemorrhage itself.

Delayed cerebral ischaemia

A late deterioration, occurring at any time from 3 days to 3 weeks after the initial bleed, is observed in up to one-third of survivors. This is thought to be a consequence of reduced perfusion of areas of the brain supplied by arteries that are affected by vasospasm.

Treatment

The objectives of treatment are to reduce the risk of the two complications of subsequent haemorrhage and cerebral ischaemia. A comprehensive review of the evidence supporting the use of different treatments following SAH has been published on behalf of the American Heart Association.[19]

A number of well-controlled trials have investigated the use of **antifibrinolytic drugs** to prevent rebleeding. Unfortunately, typical findings are that a substantial reduction in further bleeding is balanced by an increase in cerebral infarction, with no overall change in outcomes.[19,25]

The mainstay of treatment to prevent further haemorrhage, therefore, is direct surgical intervention to close the ruptured aneurysm. Since this risk peaks in the few days following SAH, it is natural to wish to operate as soon as possible. Intervention at this stage must be balanced by the fact that it can be technically more difficult to operate on a brain still suffering from the acute effects of initial haemorrhage, and the patient is more likely to be in a poorer clinical state and less able to withstand the trauma of surgery. For these reasons, the optimal timing of surgery is controversial.

The timing of surgery

The only published RCT on the timing of surgery is that of Ohman and Heiskanen[26] that randomised a total of 216 patients to acute surgery (on days 0–3); intermediate surgery (on days 4–7); or late surgery (on day 8 or later). Only 'good' grade patients (Grades I to III on the Hunt and Hess classification[27] were recruited, but in this selected subgroup and in what was an underpowered trial, there was some evidence to support acute surgery. This conclusion appears to reflect current practice, where there has been a drift towards early surgery, especially for patients in a good clinical condition.[17]

Endovascular techniques as an alternative to clipping intracranial aneurysms have evolved over the last 10–15 years. The aneurysm is occluded either by inflating a detachable balloon or by releasing coils of platinum wire to induce thrombosis. The 'state of the art' is arguably the Guglielmi detachable coil device, which has been available in North America since 1991 and in Europe since 1992. Uncontrolled series have provided encouraging results, especially with patients in poor clinical condition or with difficult surgical aneurysms. Formal evaluation of the Guglielmi detachable coil device continues, most notably in a UK Medical Research Council-funded trial of coiling versus conventional surgery.[28]

Treatment to prevent ischaemia

The strongest scientific evidence for the efficacy of a treatment following SAH is for the calcium antagonist nimodipine, which has been shown to reduce the incidence of cerebral infarction and to improve management outcomes.[29] A popular treatment used to control vasospasm is 'Triple H' therapy, namely hypertension, hypervolaemia and haemodilution. This treatment is recommended in the American Heart Association Guidelines,[19] although there is little hard evidence from clinical trials to support its use.

SMR1 data in the study of SAH

Against this background, it seemed possible to explore a sequence of questions using the linked SMR1 data.

(i) Was it possible to describe the epidemiology of the condition in the whole population and thus establish a context for its management in specialist centres and the outcomes they achieved?

(ii) Given the problem of misdiagnosis and the sequence of events needed to establish a definitive diagnosis, was it possible to describe the patterns of care on which the practice of specialised centres was actually based?
(iii) Acknowledging the difficulty of conducting formal trials of one form of intervention or another, could the SMR1 data be used to compare patients treated by the clipping of aneurysms with those treated by embolisation?
(iv) What had been the pattern of uptake of the newer embolisation treatment?
(v) Finally, providing that satisfactory answers could be found for these questions, was it possible to devise a pseudo-trial of early or late interventions?

It quickly became evident that the SMR1 data presented a number of difficulties in addressing these questions. One was that of identifying 'true' cases of SAH amongst multiple patient transfers either between hospitals or between specialties within a hospital. It was not uncommon, for example, to have an initial diagnosis of meningitis that changed to SAH in a subsequent episode record, or an initial diagnosis of SAH meriting transfer to a neurosurgical unit where the diagnosis of SAH was subsequently refuted. What were the criteria for accepting a diagnosis of SAH and at what stage in a patient's history should they be applied? The question was complicated further by the lack of information about the time of onset of symptoms or of the patient's condition on admission to hospital or later deterioration. This meant that it was not possible to discern what had been the **intended** management of the patient: was early surgery planned or the consequence of a further haemorrhage? Although death was available as an outcome measure, there was no other information about the patient's condition on discharge from hospital or subsequently.

There were also other technical problems. With the SMR1 data, surgical interventions are coded according to the Office of Population Censuses and Surveys (OPCS) codes which changed from OPCS3 to OPCS4 from 1988 to 1989;[15] codes for embolisation are included only in OPCS4 and so it was not possible to analyse the data for earlier periods. The format of the record made it difficult to reconstruct sequences of operative events and so it was not always possible to distinguish between a definitive procedure to repair a ruptured aneurysm and the surgical treatment of a later complication.

With these qualifications, the potential of the SMR1 data for studying SAH appeared to be epidemiology, admissions to neurosurgical units, the timing of surgery, embolisations, and other treatments.

Epidemiology

The linked data and death records provided an opportunity for estimating the incidence of SAH, and the case fatality rate. The inclusion of place of residence allowed the use of census-derived population denominators.

Admissions to neurosurgical units

The linkage of 'episode' data permitted an account of the sequences of patient care – for example, the numbers of patients who died before being admitted to hospital, the numbers admitted to a neurosurgical unit, and their prior inpatient care. Doing so can provide a useful account of admission policies, although the absence of data on the patients' clinical state limited this capability.

Timing of surgery

The data report the numbers of patients undergoing clipping of an aneurysm, and the timing of surgery could be related to the date of first hospital admission. Linking this information to the date of death allowed the construction of best and worst case scenarios for early surgery versus late surgery.

Embolisation

The data from different neurosurgical units provided information about the patterns of uptake of a new, unproven, technology. These data would be relevant when considering potential recruitment rates into (say) trials of embolisation versus other surgery.

Other treatments

The potential of SMR1 data to investigate other treatments was limited because information about drugs is not included. The data could have other uses, however, these might include exploring changes in case fatality rates following the publication (in the late 1980s) of evidence about the value of nimodipine in reducing cerebral infarction.[29]

The study we describe focused the use of these data as a means of assessing the optimal timing of surgical interventions.

Methods

For the reasons we describe above, subgroups reflecting different degrees of confidence in the diagnosis of SAH were extracted from the complete SMR1 data set (see *Figure 1*). The main analysis was restricted to records from 1989–1995, when surgical interventions were coded according to OPCS4. The data were for the whole of Scotland, with no attempt to compare the different neurosurgical units contributing to the data.

The first sample (sample I) comprised all cases of **suspected SAH**, defined as all cases with a primary diagnosis of SAH not otherwise specified (ICD-9 430.9) in any SMR1 record included in a continuous inpatient study or as the principal cause of death in a death record. A total of 4903 cases were identified, comprising 1796 men (37%) and 3107 women (63%); this total provides an annual incidence of 13.7 per 100,000 of the population.

Sample I consisted of three disjoint subsamples:

(i) individuals with a death record but no record of hospital admission for SAH
(ii) individuals admitted to hospital but not to a neurosurgical unit during the relevant continuous inpatient stay
(iii) patients who were admitted to a neurosurgical unit during the relevant continuous inpatient stay.

We then excluded patients for whom it might be argued that the diagnosis of SAH had not been established. These were cases admitted to a neurosurgical unit or where the final SMR1 record did not include SAH, or in cases with no primary SAH diagnosis in SMR1 records for the period of neurosurgical care. This provided sample II, which can be considered as cases of '**presumed SAH**'. For patients who were hospitalised but not admitted to a neurosurgical unit, we excluded cases where the final SMR1 record did not mention SAH as any of the recorded diagnoses. When patients had been admitted to a neurosurgical unit, we excluded cases where there was no primary diagnosis of SAH for records relating to the period of neurosurgical care. These two rules excluded

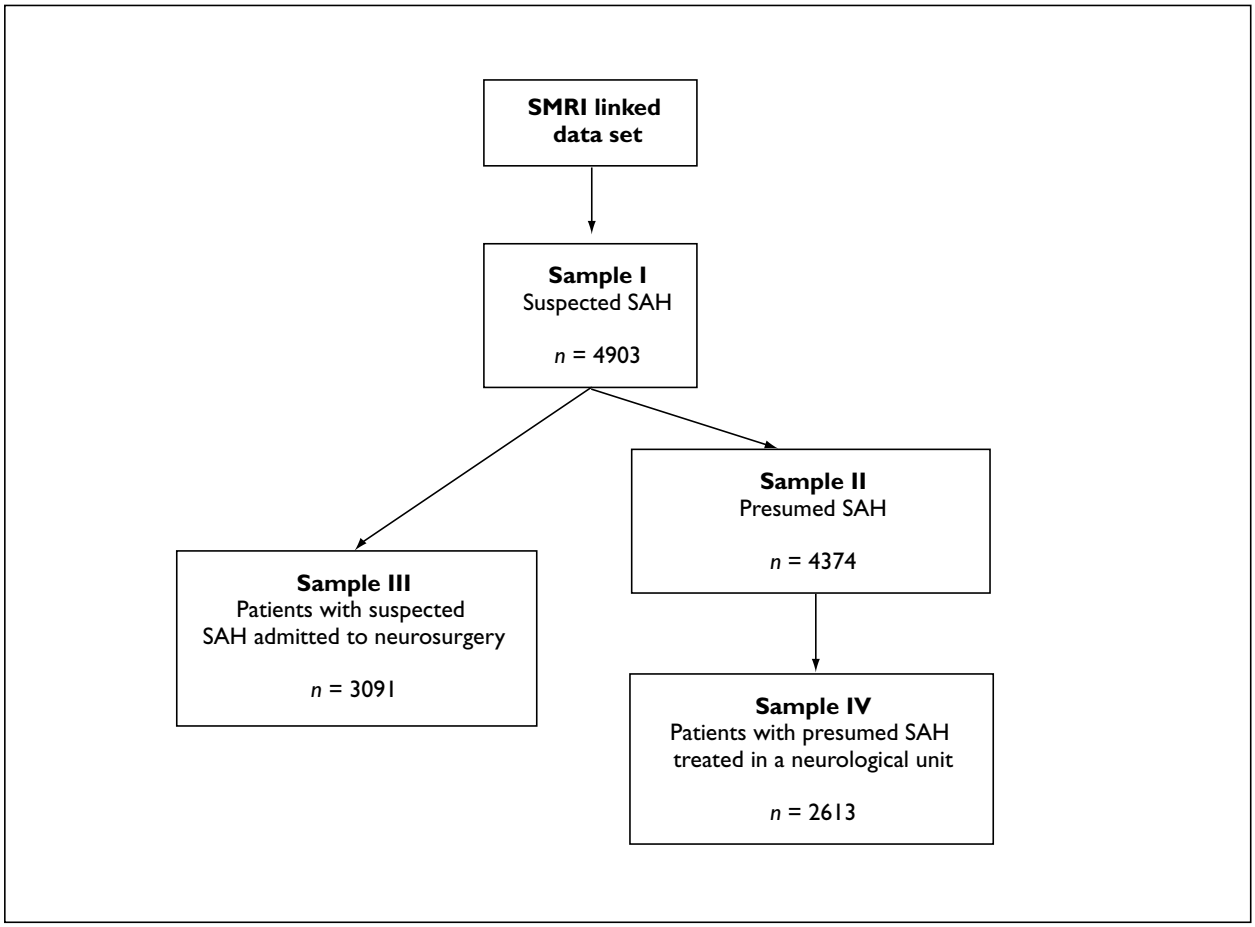


FIGURE 1 *Derivation of the four samples used in the SAH study*

51 male and 478 female cases, leaving sample II with 4374 cases of presumed SAH. This total provides an annual incidence of 12.3 cases per 100,000 of the population.

Sample III comprised all individuals in sample I who were admitted to a neurosurgical unit – that is, **patients with suspected SAH admitted to neurosurgery**. This sample reduced the initial population to 3091 individuals comprising 1163 males (38%) and 1928 females (62%). It provided a means of exploring the admission policies of the different neurosurgical units.

The final sample (sample IV) included individuals from sample II who were admitted to a neurosurgical unit – that is, **patients with a presumed SAH who were treated in a neurosurgical unit**. This sample comprised 2613 individuals (949 males and 1664 females) and is the relevant sample for exploring the potential for trials of the timing of surgery or embolisation. One might, however, take a wider view and regard the admission policies of neurosurgical units as needing review if trial populations are to be optimised.

Results

The timing of surgery

Patients who were treated within a neurosurgical unit for presumed SAH (sample IV) comprised 2613 individuals, of whom 1252 (47.9%) underwent surgery (defined as clipping of the aneurysm). *Table 2* shows the distribution of the timing of surgery, in days from first hospital admission, for three age groups: under 45, 45 to 64 and 65 and over. The percentages undergoing surgery for all ages are virtually the same in this study as the corresponding results from the International Cooperative Study on the Timing of Aneurysm Surgery,[30] although, in the latter study, the timing was from onset of symptoms rather than admission to hospital.

Table 3 reports the numbers of patients undergoing clipping in each of the three age groups: about half of the patients aged 64 or less underwent definitive procedures but this proportion was only 32% for those at older ages. It is clear from *Table 2* that the tendency towards earlier surgical intervention decreases with increasing age; for example, 15% of those aged 65 and over who underwent surgery were not operated on within 2 weeks of admission. The equivalent proportions are 8.9% for patients of age 45 to 64 years and just 4.3% for those under 45 years of age.

One-fifth of the sample died from any cause within 90 days of their first admission. *Table 4* details 90-day survival for the three age groups, when mortality was highest (32%) in the oldest group. There were only 61 deaths (5%) in the 1252 operated cases compared with 441 deaths

TABLE 3 *The proportion of patients receiving surgery in different age groups*

Age group	Number of patients	Proportion receiving surgery
< 45	902	55%
45–64	1273	49%
≥ 65	438	32%

TABLE 4 *Proportion of patients in different age groups dying within 90 days of their first hospital admission*

Age group	Number of patients	Proportion dying within 90 days of first hospital admission
< 45	902	14%
45–64	1273	19%
≥ 65	438	32%

TABLE 2 *Timing of surgery: days after first admission to hospital*

	Timing of surgery (days from first admission)								
Age group	**0**	**1**	**2**	**3**	**4–7**	**8–10**	**11–14**	**≥ 15**	**Total**
< 45	29	117	102	56	94	47	27	21	493
45–64	36	125	106	79	118	44	56	55	619
≥ 65	7	12	18	15	32	14	21	21	140
All ages	72 (6%)	254 (20%)	226 (18%)	150 (12%)	244 (20%)	105 (8%)	104 (8%)	97 (8%)	1252 (100%)

TABLE 5 *Interval before death within 90 days (patients who died) following first admission to hospital, according to operational status*

Days to death	0–1	2–3	4–7	8–10	11–14	15–28	29–90	Total
Operated cases	3% (2)	8% (5)	21% (13)	25% (15)	13% (8)	7% (4)	23% (14)	100% (61)
Cases without operation	22.5% (100)	10.5% (47)	14% (62)	15% (65)	10% (44)	15% (66)	13% (57)	100% (441)

TABLE 6 *Interval between surgery and death for cases of SAH who died following surgery and by interval between first hospital admission and surgery*

Days to death	0–1	2–3	4–7	8–10	11–14	15–28	29–90	Total
Days between first admission and surgery								
0–1	2	5	7	5	4	1	4	28
2–3	–	0	5	6	2	2	2	17
4–7	–	–	1	4	1	0	2	8
8–10	–	–	–	0	0	0	3	3
11–14	–	–	–	–	1	0	2	3
≥ 15	–	–	–	–	–	1	1	2
All cases with operation	3% (2)	8% (5)	21% (13)	25% (15)	13% (8)	7% (4)	23% (14)	100% (61)

(32%) amongst the 1361 patients who did not receive clipping. *Table 5* describes the survival time within 90 days for patients who died, following first hospital admission. About one-third of cases with no operation compared to only 11% of those who did have operations died within 3 days of admission. Approximately one-third of deaths in the operated cases and 28% in the group who did not have operations occurred between 15 and 90 days after their first admission to a neurosurgical unit. *Table 6* shows the timing of the 61 deaths following operation distributed according to the timing of surgery. Collectively, *Tables 4, 5* and *6* allow one to perform various 'what if ...?' calculations. Taking a very pessimistic view of early surgery, we could assume that all deaths following surgery performed within 3 days of admission could have been prevented if surgery had been delayed. This would have prevented 45 deaths, reducing the overall mortality rate for the whole sample of 2613 patients from 19.2% to 17.5%. In contrast, by taking an optimistic view of early surgery, we could assume that all deaths in patients without operations occurring between days 2 to 10 could have been prevented by early surgery, and that deaths in patients who had operations after day 3 could have been prevented by earlier surgery. This would have prevented 190 deaths, reducing the overall mortality rate from 19.2% to 11.9%.

These are extreme scenarios. In practical terms, they mean that a comparison between early and late surgery in an unselected neurosurgical population might require a trial with sufficient power to detect a reduction in mortality of the order of 1.7%. This implies a trial in which an unfeasibly large sample size of about 720,000 would be needed in order to achieve 90% power to detect a reduction in mortality from 19.2% to 17.5% at the 5% significance level. The SMR1 data could be used to refine these calculations – for example, by imposing an upper age limit on the trial population – but they would not permit more precise inclusion/exclusion criteria based on the patient's clinical state, which would be important considerations in the design of an actual trial.

The uptake of embolisation

We had a particular interest in the analysis of data relating to embolisation of aneurysms as a way of illustrating the uptake of a new technology. The data were divided by years from 1989 to 1995 (that is, from the introduction of OPCS4 coding), and examined separately for the four Scottish neurosurgical units (in

Aberdeen, Dundee, Edinburgh and Glasgow). When considered in the context of their catchment populations, there were clear differences between them in the timing and extent of take-up of embolisation. It was also apparent, however, that there were inconsistent patterns in the data; in particular, many patients were recorded as receiving **both** embolisation and clipping of their aneurysm. Presentation of these data at a neurosurgical meeting resulted in the clear consensus that the data on clipping were credible, but those on embolisation were not.

Having encountered this difficulty, we assessed the validity of the SMR1 data in one of the four Scottish centres (the Institute of Neurological Sciences in Glasgow) using two additional sources of data. These were the computerised log of neurosurgical procedures undertaken at the Institute of Neurological Sciences and an examination of individual (non-anonymised) SMR1 records for all neurosurgical admissions to the Institute during 1989–1995. These data were provided with the consent of the neurosurgeons concerned. The two data sets were then linked and crosschecked.

The linkage was on the basis of name, age and date of admission and/or operation. This was not straightforward; the inconsistent spelling of names and the fact that age is recorded only to the nearest month caused many problems. With a great deal of manual intervention, it was possible to create an almost perfect match at the gross level of identifying patients in the different data sets: all but four of the 643 patients recorded on the theatre log as undergoing clipping did have an SMR1 record; similarly, almost all patients with aneurysm clipping recorded on the SMR1 database did appear in the theatre log. There were, however, very obvious inconsistencies in the finer detail of the records. Seven categories of problem were identified:

(i) according to SMR1, 249 patients underwent both clipping and embolisation (39% of clipped patients according to the theatre log)
(ii) in SMR1, 36 (6%) patients were recorded as being clipped, but not recorded in the theatre log
(iii) 29 (5%) patients were clipped according to the theatre log but their SMR1 records did not mention clipping or a diagnosis of SAH
(iv) 55 (9%) patients were clipped according to the theatre log and did have a diagnosis of SAH on SMR1, but no SMR1 record of clipping
(v) 18 (3%) patients were clipped according to both the theatre log and the SMR1 data, but without an SMR1 diagnosis of SAH or non-ruptured aneurysm
(vi) 148 (23%) patients had discrepant dates between the theatre log and the SMR1 record
(vii) 184 patients were discharged alive from the Institute of Neurological Sciences with a diagnosis of SAH, but with no SMR1 record of an angiogram, clipping, or embolisation.

One of the consultants at the Institute of Neurological Sciences keeps a comprehensive file of discharge letters, and used these to check the data for 42 patients from categories (i) to (v) and (vii). In 41 instances, the SMR1 record was incorrect. One patient from group (vii) did indeed have a diagnosis of SAH but was discharged without an angiogram or any operation. By definition, the discrepant dates (category (vi)) were also errors in the creation of the SMR1 data, or arose from the structure of the 'procedures' section of the record. We discuss the more general implications of these findings at a later stage of this report.

In summary, the SMR1 data on clipping are relatively reliable, although the timing of surgery relative to admission is not always recorded reliably. The data on embolisation appear to be largely unreliable. On further examination, it was found that the specific code for angiography (L35.3 – 'arteriography of cerebral artery') did not appear in the SMR1 records, although the majority of patients with SAH will undergo this procedure. The specific code for embolisation (L35.1 – 'percutaneous transluminal embolisation of cerebral artery') was used only very occasionally. The code L35.2 – 'embolisation of cerebral artery (not elsewhere classified)' was more widely used and it is conceivable that L35.2 was used in error for L35.3. L35.2 did not appear frequently enough for it to have been systematically misused as a code for angiography.

These coding difficulties are being pursued further, but the data on 'embolisation' are clearly far too untrustworthy for reporting at this stage.

Conclusions

Using routine data for the study of SAH

Earlier in this account, we listed a number of ways in which routine data appeared to have potential for activities of the kind we describe.

An assessment of their adequacy for this purpose has two main dimensions: first, whether the scope and content of the data were sufficient and, second, whether their quality was adequate for judgements about clinical interventions and outcomes.

With qualifications, it was possible to gain insights into the contemporary epidemiology of SAH and, perhaps as usefully, to locate the patients who are treated in specialised neurosurgical units in the larger context of patients admitted to hospital with this diagnosis. The need to proceed through four stages of defining a final sample for analysis is at least a cautionary indication that the patients identified as appropriate for inclusion in trials undertaken in specialist centres may not represent the complete spectrum of a given condition. More in the sense of audit rather than evaluation, the analysis provided a means of assessing outcomes and determining (for example) survival rates over the weeks following haemorrhage. These observations should have value for informing the design of more precise trials and should have had value in determining changes in outcome following the introduction of embolisation – an application of routine data that was entirely feasible. It was regrettable, therefore, that inadequacies in the quality of the data (as we identified them in one centre) largely nullified these applications.

In summary, and for this application, the data fell short in terms of their scope and content chiefly in terms of their precision and their account of the clinical state of patients both at the time of admission and in measures of subsequent outcomes other than death. Information about the complexity and severity of individual cases were likely to have informed clinical judgements about early or late interventions so that its lack was an important deficiency. This general observation has relevance for comparable attempts to employ routine data as an alternative to RCTs in other circumstances where the latter are not thought feasible. It is worth noting, however, that these shortcomings could be overcome with relatively simple (although focused) alternatives to the usual practice of 'standard' record extraction for routine data collection. The issue of quality is of equal importance although – perhaps – one that concerns local practices rather than the inadequacies inherent in routine data systems as such. Our experiences illustrate the further difficulty that the quality of data collection is likely to depend on the historical uses for which the data have been employed. If routine data are to be employed for new purposes (such as the one we describe), then it follows that new standards of abstraction and quality control are also needed. This requirement takes in the other questions we identify about the content of these records.

Chapter 4

Case study 2: Prostatectomy for the treatment of benign prostatic hyperplasia

Introduction

The prostate is a wedge-shaped gland that surrounds the urethra as it emerges from the bladder. Benign prostatic hyperplasia is a now general term covering a wide range of urinary symptoms but was originally a precise term to describe the age-related growth of the prostate. Symptoms of benign prostatic hyperplasia include hesitant, slow, erratic, and frequent urination, inability to empty the bladder fully and frequent night time urination.[31] The epidemiology of these problems is poorly described because benign prostatic hyperplasia can be defined in many ways, but it has been estimated that 75% of men will develop benign prostatic hyperplasia by the time they are 80 years of age.[32]

Treatment

Treatment is indicated by the severity of a patient's symptoms, which are often measured by questionnaire although no specific degree of difficulty can be taken as an indication for a particular treatment.[33] One option is the use of drugs, such as α-blockers and finasteride. Many RCTs have compared these drugs with placebo and have shown them to lead to substantial reductions in the symptoms of a proportion of men.[31] The alternative is surgery, in particular prostatectomy which involves the removal of the prostate gland.

History of prostatectomy

Until 35 years ago, OPEN was the standard method of prostatectomy in the UK.[34] Since then, TURP has gradually replaced OPEN as the surgical treatment of choice.[35] TURP is an operation that reduces the bulk of the prostate and relieves the static obstruction that the gland's enlargement has created; it was the first major endoscopic operative procedure in medicine. In 1996, TURP accounted for 94% of all surgery for benign prostatic hyperplasia in the US Medicare program.[36] OPEN is still the surgery of choice for patients with particularly large prostate glands:[31] a prospective study undertaken by the American Urological Association found that all patients with a prostate larger than 80 g were subjected to OPEN.[32]

A retrospective study of TURP and OPEN has been performed in Scotland for the period 1968 to 1989.[37] The study was based on the linked data set used for this case study – that is, the linked SMR1 data. The Scottish results showed that there were more OPENs than TURPs until 1976 with a dramatic increase in TURPs and a gradual decline in the OPEN procedure since that date.

TURP versus OPEN procedures

This change of surgical practice occurred without a systematic assessment of whether TURP was preferable in terms of effectiveness and safety to OPEN.[35] Initially, it seemed that TURP had the advantages of lower operative mortality and shorter hospital stay. Its disadvantages were that it was difficult to learn, required special training and often needed revision.[34]

A number of more recent retrospective studies have compared the merits of TURP and OPEN. One included data from England, Denmark and Canada with a total recruitment of more than 54,000 patients.[35] The results provided evidence for increased rates of re-operation and reduced survival after TURP compared with OPEN and these were consistent over time and location. They also suggested an excess number of deaths due to cardiovascular disease, especially MI, in patients who underwent TURP. On the other hand, it has been argued that the presence of cardiovascular disease may influence the decision to perform TURP rather than OPEN, thus explaining the higher mortality associated with the TURP patients.[38]

A further study conducted in Denmark also found elevated mortality following TURP.[39] In contrast to the findings of Roos and colleagues,[35]

this study found that cardiovascular disease, and in particular MI, were not especially important causes of subsequent death in the TURP patients but the Danish investigators did report that respiratory disease (especially chronic bronchitis) accounted for a disproportionate number of deaths. These findings were mirrored by the Scottish study, which also reported that the risk of mortality from cancer (especially prostate and bladder cancer) was significantly increased after TURP.[37]

SMR1 data in the study of prostatectomy

The potential of the data

Genuinely comparable groups of patients, necessary for an assessment of the merits of TURP and OPEN can only be obtained from RCTs. In the present climate, however, it would be difficult to conduct RCTs on ethical grounds because the majority opinion is that TURP is superior to OPEN.[39]

The SMR1 data set allowed a retrospective, non-randomised comparison of TURP and OPEN. The study was representative of a general population – that is, all men in Scotland receiving surgical treatment for benign prostatic hyperplasia – rather than the selected patient group of an RCT. Large patient numbers meant that it was possible to explore subgroup effects (such as mortality amongst patients with a history of cardiovascular disease). By defining a suitable length of study period, it was possible to assess differences between TURP and OPEN for longer-term outcome measures.

The limitations of the data

Although TURP has become the surgery of choice worldwide, OPEN is still the preferred method for patients with particularly large prostate glands. It was reasonable to assume that TURP surgery is not performed in Scotland for prostates larger than a certain size, but a disadvantage of the SMR1 system is that neither the size of the prostate nor other indications for surgical management are recorded. This information would have been necessary to obtain truly comparable groups of patients for an assessment of the relative merits of TURP and OPEN, and it would have been preferable to omit patients in whom TURP was contraindicated. In an ideal world, it would have been desirable to perform analyses of the outcomes of TURP and OPEN surgery for different degrees of prostatic enlargement.

Co-morbidities, such as the presence of cardiovascular disease, are included in the SMR1 data but the severity of either prostatic enlargement or of other morbidities is not. Given the findings of earlier studies, these considerations are obviously relevant to a comparison of the two procedures and the indications for one or the other in individual patients.

Methods

The basic study sample consisted of all patients who received their first prostatectomy (TURP or OPEN) between 1989 and 1995 with a principal diagnosis of benign prostatic hyperplasia (ICD-9 code 600). Patients whose first hospital admission (episode) for prostatic hyperplasia included a record of both TURP and OPEN prior to 1989 were excluded because it was not possible to ascertain the initial procedure. Historical variables were derived from the linked data in order to describe a patient's previous hospital admissions. These included codes for diabetes at any time prior to hospital admission for prostatic disease, a hospital admission with a diagnosis of circulatory disease in the previous 5 years, and an admission with a diagnosis of acute MI in the previous 28 days. These variables are listed in *Table 7*. Each patient's subsequent hospital and death records were abstracted in order to measure mortality and re-operation rates.

Results

The initial comparison

There were 26,225 first prostatectomies in Scotland between 1989 and 1995; of these, 97% (25,345) were TURP and only 3% (880) were OPEN. *Figure 2* plots the frequency of the two procedures: the number of OPEN procedures has gradually declined and, except for a peak in 1994, the number of TURP operations has remained fairly constant.

The patient group who underwent OPEN (mean age 72 years) were slightly older than TURP patients (mean age 70 years); this finding has been reported in other retrospective studies.[39,40] The mean age of TURP patients remained fairly constant between 1989 and 1995, but the mean age of OPEN patients has been more variable. Part of this observation can be attributed to the small numbers involved,

TABLE 7 *Variables describing a patient's hospital admissions prior to prostatic surgery abstracted from the linked SMR1 data set*

Variable	ICD-9 code(s)	Variable name
Acute MI coded in any diagnostic position at any previous time	410	AMI
Acute MI coded in any diagnostic position in the previous 28 days	410	AMI4
Angina coded in any diagnostic position at any previous time	413	ANGI
Cancer coded in any diagnostic position in the previous 5 years	140–239	CAN5
Condition of the circulatory system coded in any diagnostic position in the previous 5 years	390–405; 430–459	CIR5
Diabetes coded in any diagnostic position at any previous time	250	DIAB
Heart disease (not ischaemic heart disease) coded in any diagnostic position in the previous 5 years	415–417; 420–429	HTD5
Other ischaemic heart disease coded in any diagnosis column at any previous time	411–412; 414	OIHD
Conditions of respiratory system coded in any diagnostic position in the previous 5 years	460–519	RES5

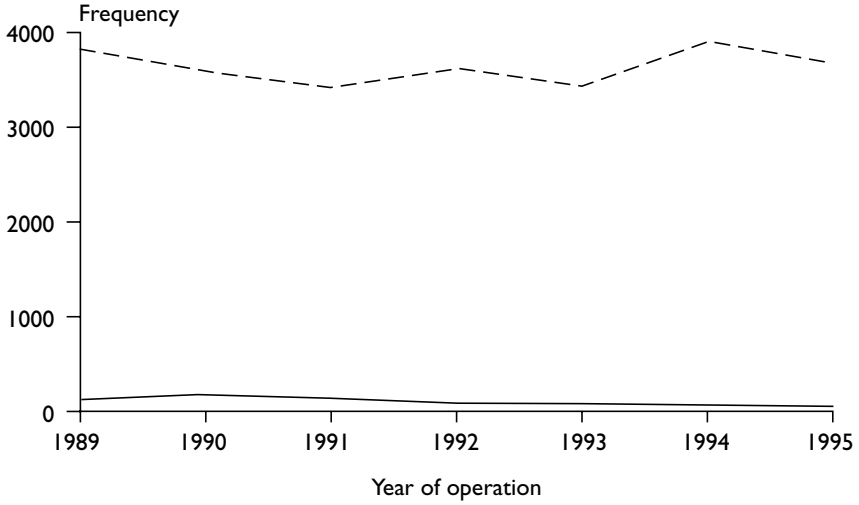


FIGURE 2 *The frequency of TURP (– – –) and OPEN (——) in Scotland: 1989–1995*

but there is a case for suggesting the mean age for this procedure has been falling (*Figure 3*).

Characteristics of patients undergoing TURP and OPEN

In a retrospective, non-randomised study of this kind, it is impossible to guarantee that the characteristics of patients in different treatment groups will be similar. Indeed, it is unlikely that this will be so because the data will reflect clinical grounds for patient selection.

An early analysis, based on logistic regression, explored differences in case mix between TURP and OPEN patients. The significant covariates and their associated odds ratios from the final regression model are listed in *Table 8*. As might have been expected, younger patients (< 60 years) had an increased chance of receiving TURP. Patients with previous hospital admissions for circulatory or respiratory disease also had an increased chance of receiving TURP, as did patients with earlier ischaemic heart disease,

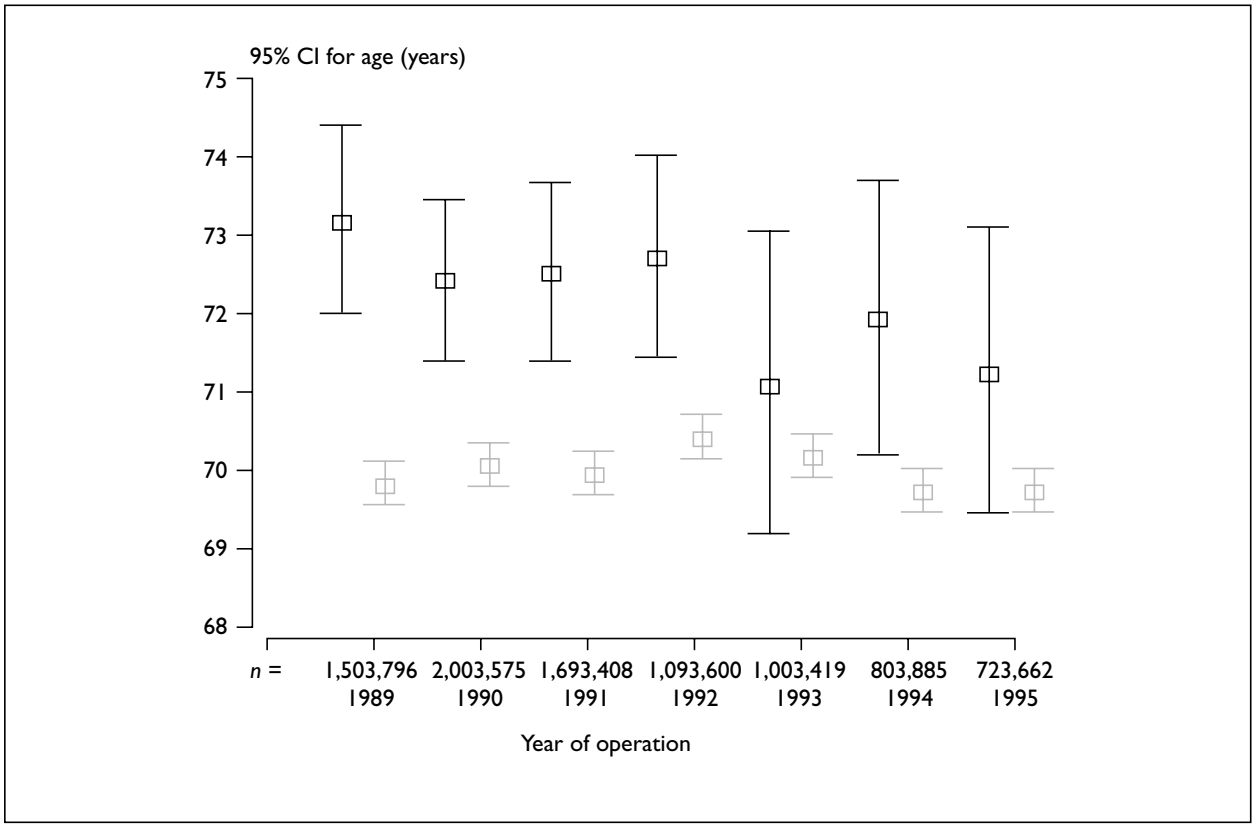


FIGURE 3 *Mean age and 95% confidence intervals (CIs) for OPEN (——) and TURP (——) patients: Scotland: 1989–1995*

TABLE 8 *Significant covariates and their associated odds ratios from a logistic regression model investigating treatment selection bias. Variable definitions are those in* Table 7

Significant covariates	*n*	% TURP (*n*)	Odds ratio (95% CI)
Age 70*	1204	95.8% (1153)	1
Age 60	652	98.0% (639)	1.41 (1.40 to 1.42)
No prior admission for circulatory system	23,455	96.5% (22,640)	1
Prior record of CIR5	2770	97.7% (2705)	1.42 (1.09 to 1.83)
No prior admission for respiratory system	24,308	96.6% (23,472)	1
Prior record of RES5	1917	97.7% (1873)	1.46 (1.07 to 1.99)
No prior admission for other ischaemic heart disease	24,494	96.5% (23,643)	1
Prior record of OIHD	1731	98.3% (1702)	1.90 (1.30 to 2.76)

**Age was fitted as a continuous variable. Ages 60 and 70 are included as illustrations*

which included conditions such as acute MI (ICD-9 410), other ischaemic heart disease (ICD-9 411–412; 414) and pre-infarction syndrome (ICD-9 411). This finding supports the argument that the presence of cardiovascular disease may influence the decision to perform TURP rather than OPEN.[38] Somewhat surprisingly, therefore, there appeared to be no selection bias for patients who had been admitted to hospital for acute MI (ICD-9 410) in the previous 4 weeks.

Survival without re-operation

Figure 4, derived from a Cox regression model, shows survival free from re-operation in the TURP and OPEN patient groups. The survival curves in this figure describe the time interval to **first** re-operation: some patients had more than one subsequent operation. About 12% of the TURP patients had at least one more prostatic operation over the 7 years of the analysis with a fairly smooth decline in the proportion who

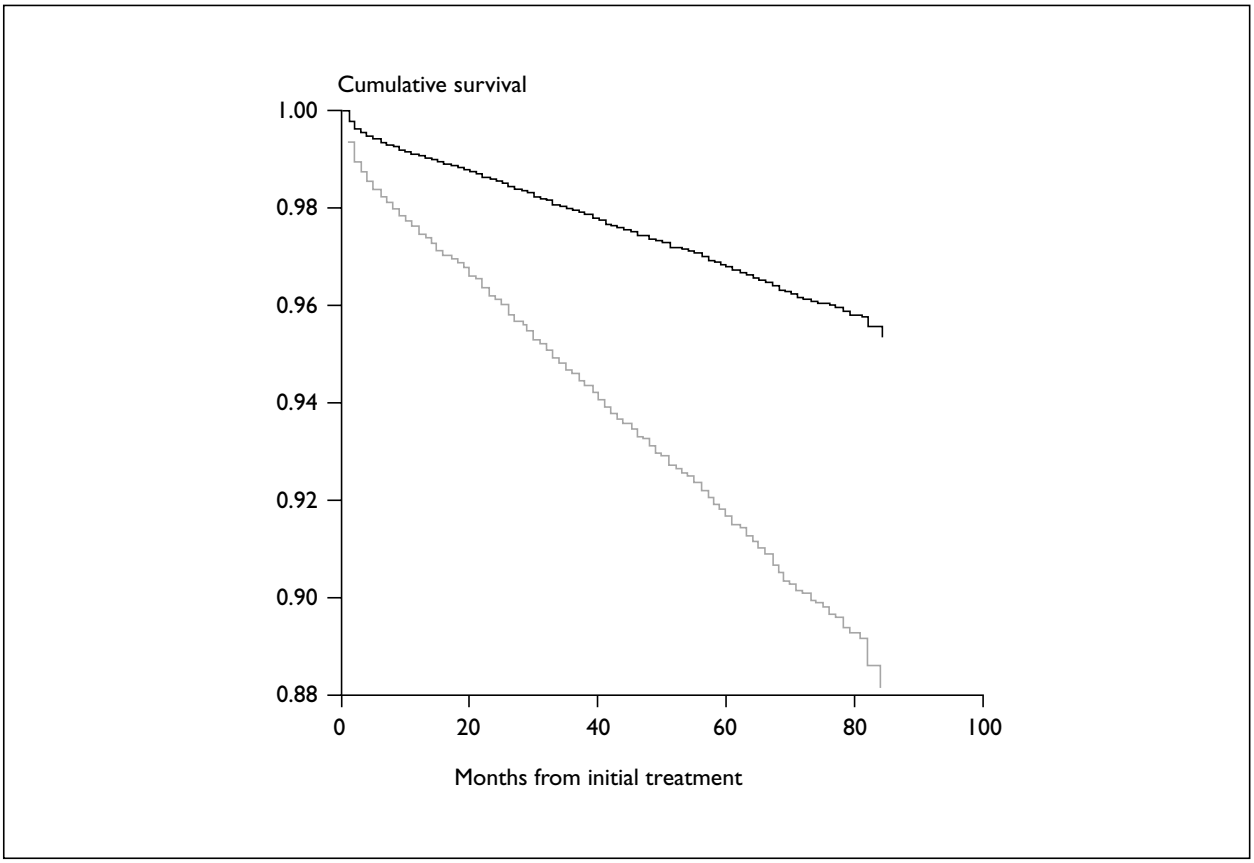


FIGURE 4 *Survival free from further prostatic surgery for OPEN (——) and TURP (——) patients (cumulative survival in months from first operation)*

had not had a further operation over this period. Perhaps because of the nature of the procedure, the re-operation rate for patients with an initial OPEN operation was very much smaller at about 4%. Younger patients (odds ratio 0.94) were less likely to undergo further operation and patients with cancers (odds ratio 1.26) were more likely to experience further surgery.

There was a highly significant interaction between diabetes as a co-morbidity and the type of prostatectomy ($p < 0.0001$) when diabetic patients who received OPEN were much more likely (odds ratio 16.94) to require further surgery than diabetic patients who received TURP (odds ratio 0.87). Initially, this result appeared to illustrate the potential of routine data for exploring subgroup effects that might be difficult to identify in RCTs. Only 17 diabetic patients were treated with OPEN surgery, however, and thus, in statistical terms, it is likely that this observed interaction was due to the effect of small numbers rather than a true difference between TURP and OPEN for patients with diabetes. It is also possible that the choice between TURP and OPEN reflected differences in the severity of the patients' diabetes (or the presence of diabetic complications) that were not included in the SMR1 data.

In total, 96% of the 1440 re-operations were for TURP, regardless of the initial operation performed. As well as TURP being the surgery of choice for a patient's first prostatectomy, it was also preferred if a patient's condition worsened, if initial surgery failed, or if the surgery was associated with complications.

A drawback of the analysis was that only 24 OPEN patients were re-operated on compared to 1416 TURP patients. This discrepancy made it difficult to verify the results of the study as this imbalance had a considerable influence on the study as a whole.

Mortality rates for TURP and OPEN

Crude mortality rates at 90 days, 1 year and 5 years are shown in *Table 9*, which also includes adjusted odds ratios of TURP compared to OPEN. These were adjusted for case mix using logistic regression, in which those variables listed in *Table 7* found to be significantly related to mortality were included. The mortality rates are biased because, in order to allow each

TABLE 9 *Crude mortality rates after 90 days, 1 year and 5 years for TURP and OPEN in Scotland: 1989–1995. The odds ratios are adjusted for case mix using logistic regression*

Time	Crude mortality rate	Unadjusted odds ratio	Adjusted odds ratio
90 days	TURP 1.4%; OPEN 2.7%	OR 0.52 (95% CI, 0.34 to 0.78)	OR 0.56 (95% CI, 0.36 to 0.86)
1 year	TURP 5.6%; OPEN 6.7%	OR 0.84 (95% CI, 0.64 to 1.09)	OR 0.95 (95% CI, 0.71 to 1.27)
5 years	TURP 28.0%; OPEN 29.1%	OR 0.96 (95% CI, 0.81 to 1.14)	OR 1.15 (95% CI, 0.89 to 1.48)

OR: odds ratio

patient a suitable length of follow-up, data from later years (or in the case of 90 day rates, year and months) have been omitted.

Both the unadjusted and adjusted odds ratios at 90 days show that mortality for the OPEN patients is significantly worse than mortality for the TURP patients. This result contradicts the overall finding of the study reported by Roos and co-workers,[35] which described an increased risk of death at 90 days after TURP when compared with OPEN. In that study, however, (and in Denmark, which accounted for the majority of study patients) 90-day mortality was slightly lower among patients undergoing TURP who were 75 to 84 years of age.

Mortality rates, both unadjusted and adjusted, were not significantly different after 1 and 5 years. Higher mortality amongst TURP patients was reported at these time points by Roos and colleagues.[35] It is worth noting that the period of the present study is 1989–1995 whereas the Danish, Canadian and English study employed data ranging from 1963 to 1985. The results of the two studies are thus not strictly comparable because different time periods (and different stages of the adoption of the TURP procedure) were investigated.

Significant predictors of mortality at 90 days, 1 year and 5 years were increasing age, cancer, earlier circulatory disease, other heart disease, and respiratory disease as co-morbidities. As we note above, exploration of subgroup effects was not pursued because of the effect of small numbers on covariate interactions.

Survival following prostatectomy

Mortality following TURP and OPEN was assessed for three specific time intervals. A Cox regression analysis was employed as a way of evaluating cumulative survival following treatment between 1989 and 1995.

The significant covariates from the final regression model, and their associated relative risks, are listed in *Table 10*. As before, patients aged less than 60 years had a much reduced risk of death (odds ratio 0.44) and patients with co-morbidities – cancers, other circulatory disease, older records of ischaemic heart disease, other heart disease, respiratory disease, or recent MIs – had an increased risk of death with relative risks ranging between 1.64 (cancers) and 1.16 (previous ischaemic heart disease). There was no evidence of a treatment by covariate interaction. *Figure 5*, derived from the Cox regression model, shows survival in the TURP and OPEN patient groups. The similarity in shape of the two survival curves is striking – although TURP patients have a reduced chance of survival, the difference is not significant. *Figure 5* is virtually the same as the corresponding results from the Danish retrospective study, except that the death rates for the Scottish data are lower for both procedures.[39]

Death from MI following prostatectomy

Earlier, we noted evidence that suggested an excess number of deaths from cardiovascular disease, especially MI, in patients receiving TURP.[35] Patients whose principal diagnosis at death was MI (ICD-9 code 410) were identified and a Cox regression analysis performed to determine whether TURP patients had an increased risk of death from this cause. Death from MI accounted for 20% of all deaths.

Figure 6 shows mortality from MI in the TURP and OPEN patient groups. The shapes of the survival curves are almost identical to those for all deaths (*Figure 5*) except that OPEN patients have a reduced chance of survival compared to TURP patients. This difference is not significant, but it contradicts the findings of Roos and colleagues.[35] Not surprisingly, the best predictor of death from this cause was

TABLE 10 *Significant covariates and their associated odds ratios from a Cox regression model investigating survival free from mortality. Variable definitions are those in* Table 7

Significant covariates	*n*	Proportion who died (*n*)	Odds ratio (95% CI)
Age 70*	1204	14.9% (179)	1
Age 60	652	8.4% (55)	0.444 (0.442 to 0.446)
No prior admission for cancer in previous 5 years	24,023	18.3% (4394)	1
CAN5	2202	27.7% (610)	1.64 (1.51 to 1.78)
No prior admission for circulatory disease in previous 5 years	23,455	18.1% (4256)	1
CIR5	2770	27.0% (748)	1.46 (1.35 to 1.58)
No prior admission for heart disease in previous 5 years	24,758	18.0% (4454)	1
HTD5	1467	37.5% (550)	1.56 (1.41 to 1.72)
No prior admission for respiratory disease in previous 5 years	24,308	18.0% (4376)	1
RES5	1917	32.8% (628)	1.64 (1.50 to 1.79)
No prior admission for acute MI	24,487	18.6% (4565)	1
AMI	1738	25.3% (439)	1.27 (1.15 to 1.41)
No prior admission for diabetes	25,556	18.8% (4793)	1
DIAB	669	31.5% (211)	1.57 (1.36 to 1.80)
No prior admission for other ischaemic heart disease	24,494	18.7% (4582)	1
OIHD	1731	24.4% (422)	1.16 (1.04 to 1.29)

**Age was fitted as a continuous variable. Ages 60 and 70 are included as illustrations*

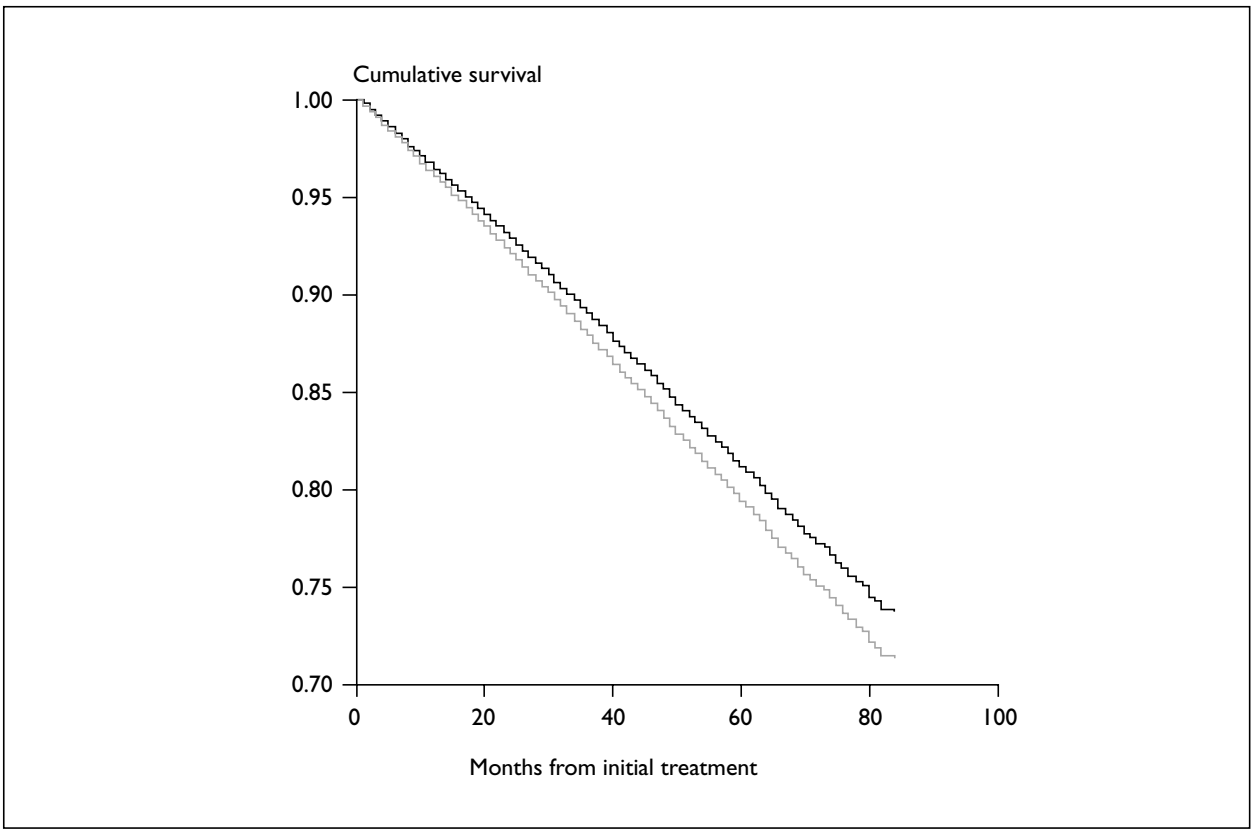


FIGURE 5 *Survival for OPEN (——) and TURP (——) patients in the 90 months following operation*

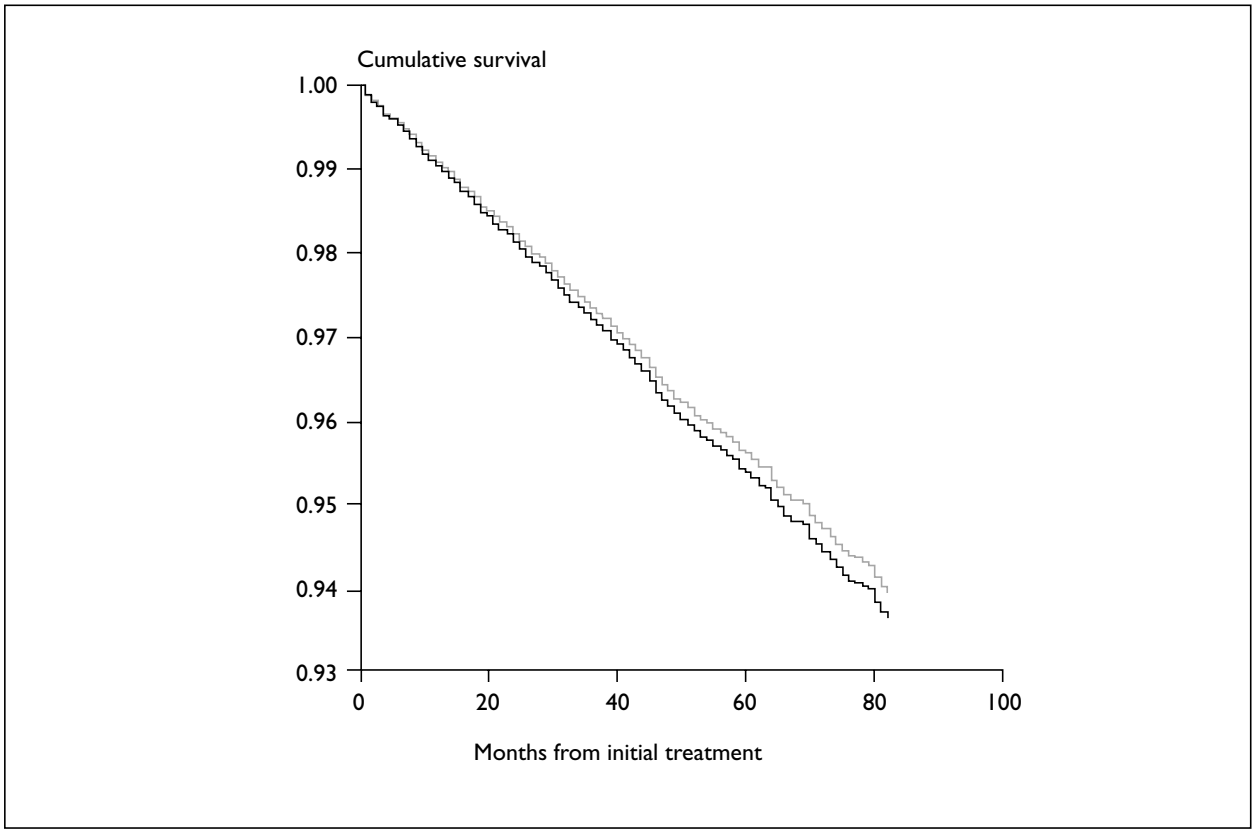


FIGURE 6 *Death from acute MI (ICD-9 code 410) for OPEN (—) and TURP (——) patients in the 90 months following operation*

whether the patient had a history of acute MI at the time of operation [relative risk (acute MI: no acute MI) = 1.73; 95% CI, 1.42 to 2.11].

Conclusions

TURP has become the surgery of choice for benign prostatic hyperplasia without a systematic assessment of its effectiveness and safety. In Scotland, as in the rest of the world, TURP is the surgery of choice with OPEN accounting for only 3% of first operations between 1989 and 1995. For this reason, RCTs in this area are likely to be both impractical and unethical because the majority of surgeons believe TURP to be superior to OPEN. In these circumstances, the objective of the present study was to compare TURP and OPEN using retrospective, non-randomised data.

As the literature had reported, the results of the study show that the risk of re-operation was substantially higher among patients who underwent TURP.[35] Mortality at 90 days was significantly worse for OPEN patients compared to TURP patients but, after 1 year and 5 years, there were no significant differences in death rates between the two groups. This finding contradicts earlier reports although it should be remembered that the study periods were different.[35] An analysis of cumulative survival failed to demonstrate significant differences between the two procedures: these survival curves were virtually identical to those from a comparable Danish study.[39]

In a somewhat negative sense, the study illustrates the need for timeliness in comparisons of this kind. As TURP is now the procedure of choice and is employed for a very high proportion of cases, a comparison of the two procedures may no longer be meaningful. The patient characteristics of the two treatment groups were different suggesting different (unrecorded) criteria in the choice of procedure. The substantial discrepancy in the size of the two groups introduced difficulties in the statistical analysis, especially in the assessment of potential subgroup effects. This was an important disadvantage when one considers the prevalence of co-morbidities in the relatively elderly population at risk of this procedure; questions such as re-operation or survival rates are likely to be important in light of these additional conditions and not simply as a characteristic of the procedures themselves.

Using routine data for the study of prostatectomy

It will be evident that the linked SMR1 data set was a more satisfactory framework for a study of this type than that we describe for SAH although some of the inadequacies of the data – notably the lack of data about complexity and severity – were common to both studies. In the context of what was virtually a single procedure for a given condition, the issue emerging from the analysis of the SMR1 data was that of the way in which co-morbidities influenced or modified outcomes. The prospect of a long series of RCTs investigating the implications of these many possibilities is difficult to imagine and so one use of routine data (in this and comparable circumstances) could be that of evaluating covariate effects at least to the stage of identifying relationships meriting more formal investigation. The odds ratios we report above go some of the way towards illustrating this application, but due to the lack of clinical detail in the records, they are not a satisfactory basis for study design. This is not to say, however, that routine record systems could not be adapted for these purposes.

There is a potential for bias if there are differential levels of the recording of secondary diagnoses between hospitals, as was noted in a recent study using routine data in England based on the English NHS Contract Minimum Data Set.[41] Whilst it is possible to take such bias into account during analyses,[42] this is of more importance if direct comparisons are being made between providers. In a similar vein it is possible that differences between hospitals in admission thresholds could lead to differences in recorded medical history (as reflected in prior hospital admissions).[42]

Chapter 5

Case study 3: Coronary revascularisation

Introduction

CABG as a technique for coronary revascularisation in patients with symptomatic coronary artery disease was first reported in 1968.[43] When the procedure is compared to medical therapies, there is strong evidence that the technique results in prolonged survival and an improved quality of life in specific patient subgroups.[44] The rate of coronary artery surgery in the UK has risen steadily from 212 per million in 1986 to 341 per million in 1993/1994.[45]

PTCA was described in 1979[46] as an alternative revascularisation technique. It was used initially in patients with single vessel disease but, as the technology advanced, the procedure has been employed for patients with multivessel disease. A more recent development has been that of placing coronary stents as part of the PTCA procedure: it has been estimated that more than 500,000 such stenting procedures would take place worldwide in 1998, and would comprise between 60% and 90% of all PTCA procedures.[47]

Clearly, there are fundamental differences between the two techniques. With PTCA, there is a risk of further stenosis as a result of elastic recoil, or from the progression of atherosclerosis. PTCA does, however, retain future options for revascularisation. CABG places venous conduits, which are largely immune from advancing disease[48] although they do tend to deteriorate over time. Repeating the bypass surgery can be compromised if the 'prime' veins have been used in an initial procedure.

The evidence from trials

There are several well-defined subgroups of patients where one or other procedure is clearly indicated, but there is a large 'grey area' where either procedure may be thought appropriate. Several clinical trials comparing CABG and PTCA as the first intervention in patients with symptomatic coronary artery disease have been conducted or are on-going. In 1995, a meta-analysis by Pocock and colleagues[49] summarised the results of eight trials in which a total of 3371 patients (1661 CABG and 1710 PTCA) had been followed for a mean of 2.7 years. The number of deaths was 73 in the CABG group and 79 in the PTCA group resulting in a relative risk for PTCA of 1.08 (95% CI, 0.79 to 1.50). The number of deaths is too small for this result to establish the equivalence of the two mortality rates: a more sensitive comparison of the two procedures employed the composite endpoint of cardiac death or non-fatal MI.

The results of the eight trials were relatively consistent, with a total of 127 endpoints in the CABG group during the first year compared to 135 endpoints in the PTCA group (relative risk 1.03; 95% CI, 0.84 to 1.27). Over the entire follow-up period there were 154 endpoints in the CABG group and 169 in the PTCA group (relative risk 1.10; 95% CI, 0.89 to 1.37). In contrast to these findings, there was evidence that CABG gave better control of angina compared with PTCA and required less frequent subsequent intervention. A subgroup analysis based on three of the eight trials suggested that CABG was to be preferred to PTCA in patients with single vessel disease in terms of the endpoint of cardiac death or non-fatal MI. This is, perhaps, a counter-intuitive finding to be interpreted with caution given the very small number of events on which the subgroup analysis was based.

These findings are limited by the short follow-up period, and so it is helpful to consider the results of the BARI trial, which had an average follow-up of 5.4 years.[50] The 5-year survival rate was 89.3% for the CABG group and 86.3% for the PTCA group ($p = 0.19$; 95% CI for difference –0.2% to 6.0%). Again, there is no strong evidence of a difference in outcomes, although the limited number of deaths is reflected in a wide CI that does not exclude the possibility of a clinically relevant difference. A post hoc subgroup analysis suggested that CABG is to be preferred to PTCA in patients with treated diabetes.

Limitations of the trial data

In a commentary accompanying the Pocock meta-analysis, White[51] pointed out that the trials of CABG versus PTCA leave many questions unanswered. There is, first, the problem of generalising these findings: the trials have randomised only a minority of the patients who were eligible for revascularisation. Second, even when taken together in a meta-analysis, trials up to the present have been underpowered, and their results do not exclude what may be clinically relevant differences in either mortality or later MI. Third, there is the question of the most appropriate outcome measures: one could, for example, consider PTCA as a 'package' or a holding procedure, accepting as a 'cost' the possibility of further intervention in return for the perceived benefit of delaying a CABG procedure that it might not be possible to repeat. For this argument, follow-up periods of the trials have been too short because CABG procedures would not be expected to fail for about 6 years. Fourth, there is the question of subgroup effects. Various trials have suggested that there are such effects, but none has had sufficient power to investigate this issue. This is a crucial difficulty because the clinical question is not so much a simple comparison of CABG with PTCA, but rather one of narrowing the indeterminate area between them by improving the identification of patients where one or other procedure is indicated. Finally, the trials report outcomes that may have been overtaken by advances in the procedures themselves. Anaesthetic and surgical techniques have improved, and the use of stents in conjunction with PTCA is becoming widespread.[47] How relevant are comparisons of yesterday's technology for today's practice?

Some of these problems are inherent in any approach to health technology assessment. By definition, evaluation must lag behind the development of new technologies – in particular, and for questions of this kind, it takes time to accumulate adequate long-term follow-up data. Routine data have the potential to address some of these problems. They provide a powerful means of assessing the relationship of trial populations to the 'real world' from which they are drawn, and they have the potential for assembling much larger numbers of cases than can be achieved in trials. Routine data could be used to augment a meta-analysis of trial data, and could be used to explore subgroup effects.

SMR1 data in the study of coronary revascularisation

The potential of the data

The two main ways in which trial data in this area may be insufficient are, first, that, even with meta-analysis, patient numbers are insufficient for detecting modest but relevant differences in mortality between the two procedures and are inadequate for exploring subgroup effects. Second, the length of follow-up is unlikely to detect late failures of CABG. A further problem is that trial subjects are highly selected and likely to be unrepresentative of the large majority of patients who undergo coronary revascularisation. Hundreds of thousands of these operations are conducted around the world each year, and the poverty of trial data is, therefore, an interesting commentary on the perceived importance of trials as a support for major invasive procedures when compared to those expected for pharmaceutical innovations. It is, perhaps, a reflection of differing regulatory standards and requirements.

Routine data can provide the potential to study very large numbers of patients, with prolonged follow-up. Unlike trial data, they also provide the opportunity to study epidemiological trends in the uptake of different procedures, and to characterise their use in the entire patient population. In this case study we describe these epidemiological aspects of their use briefly, but focus chiefly on trying to combine information from the routine data with the trial data in order to make formal comparisons of outcomes following the two procedures.

The limitations of the data

In practice, several problems limited the usefulness of the linked SMR1 data for analysing the outcomes of coronary artery revascularisation. The first was that there was no definitive way of identifying patients who were undergoing their first coronary revascularisation outwith the data themselves. The linked data had to be used to identify records of previous procedures; doing so was further complicated by the fact that an operative code for PTCA was not included in the OPCS3 operation codes used before 1989. Patients who underwent this procedure before this date could not be confidently identified. The records did not include a description of the patients' clinical state (specifically, the severity of angina). This information would have been invaluable as a way of stratifying case mix, and in interpreting outcome measures.

There was no record of the extent of the underlying pathology – particularly, whether patients had single or multiple vessel disease, used in many trials as an entry criterion, or as a stratifying factor. It was possible to approximate to this pathological classification by using procedure codes, which permitted a distinction between single and multiple vessel interventions whether for CABG or PTCA. Doing so introduced a degree of confounding, however, because it was possible for patients with multivessel disease to have a first procedure performed on only one of the affected vessels.

Methods

The initial sample comprised all patients with an SMR1 record of CABG or PTCA between 1989 and 1995. These records were then divided between patients who had procedures for single or multivessel disease on the basis of the OPCS4 operative codes. A subset of the sample was then generated to reflect as closely as possible the entry criteria for the major trials, with the idea that it may be possible to compare the results achieved following CABG or PTCA in the general populace with those that would have been expected given the trial results. Moreover, it was of interest to note whether the slightly higher relative risk estimated for PTCA on the basis of the meta-analysis would change in the light of additional data. To define the subset, patients were excluded if there was a record of an earlier CABG (from 1981 onwards) or PTCA (from 1989 onwards) or if any of the following diagnoses were recorded at the time of surgery: diseases of mitral valve (ICD-9 394); diseases of aortic valve (ICD-9 395); diseases of mitral and aortic valves (ICD-9 396); acute MI (ICD-9 410); other diseases of endocardium (ICD-9 424); cardiac dysrhythmias (ICD-9 427); heart failure (ICD-9 428); and ill-defined descriptions and complications of heart disease (ICD-9 429). The remaining sample within the limits of the SMR1 data were regarded as patients undergoing their first coronary revascularisation and was designated the 'pseudo-RCT sample'. It included 12,238 individuals, of whom 8524 (70%) underwent CABG. Those failing to meet these criteria are described as the 'excluded sample' and comprised 2359 individuals, of whom 1825 (77%) underwent CABG.

Results

Epidemiology: coronary revascularisation in Scotland between 1989 and 1995

Figure 7 plots the number of CABG and PTCA procedures undertaken in Scotland from 1989 to 1995 for patients in the pseudo-RCT sample. At 33 per 100,000, the rate of CABG operations was considerably greater in 1995 than the rate of 9 per 100,000 in 1989; the rate for PTCA was 14 per 100,000 in 1995 – an increase from 7 per 100,000 in 1989. In absolute numbers,

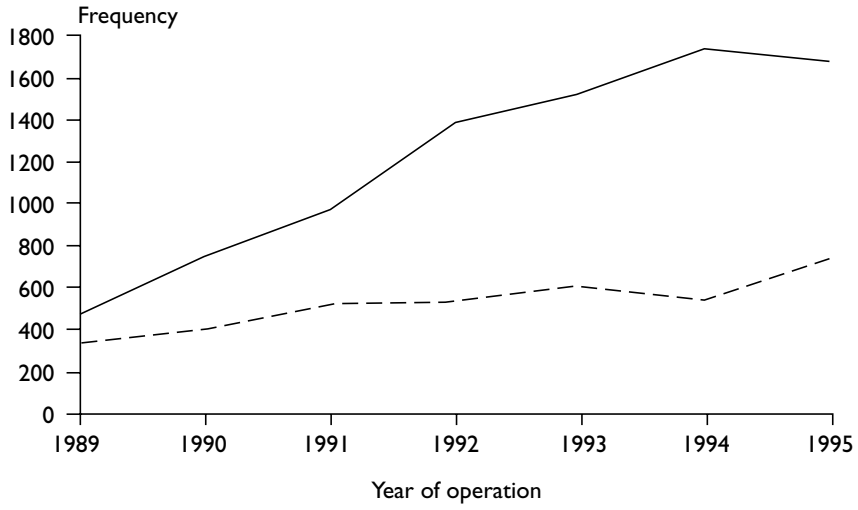


FIGURE 7 *The frequency of PTCA (– – –) and CABG (——) (numbers of cases) in Scotland: 1989–1995*

CABG was considerably more common than PTCA, and, until 1994, the rate of increase was also greater for CABG. This increase was mostly attributable to the largest of the three main Scottish centres performing these operations.

An initial logistic regression investigated case mix differences between patients undergoing the two procedures. As expected, there was strong evidence of selection bias: female patients, patients with multivessel disease, and patients with hypertension were more likely to receive CABG. (One referee pointed out that it was possible that the observation that female patients were more likely to undergo CABG could be explained by their older age at onset of angina. Unfortunately age at onset is not collected routinely and it is unlikely that a prior hospitalisation for angina, which would be available from linked records, would give a clear link to the time of onset.) Patients with a record of MI in the previous 4 weeks and those with angina as a recorded co-morbidity were more likely to receive PTCA. This pattern was clearly evident in the pseudo-RCT sample, and even more apparent in the excluded sample.

Comparing RCT subjects with 'routine' CABG/PTCA patients

Table 11 compares the baseline characteristics of the two samples from the SMR1 data with a major trial published after the Pocock meta-analysis,[49,50] and the three largest trials from the meta-analysis.[52–54] The pseudo-RCT patients are broadly similar to the trial populations, although a major difference between the trials is whether or not they focused on multivessel disease. The excluded sample had a higher proportion of females, but was otherwise broadly similar to the other populations.

Cardiac death and/or acute MI following CABG/PTCA

The primary endpoint for the Pocock meta-analysis[49] was cardiac death and/or acute MI within 1 year. *Table 12* shows the relevant event rates in each of the trials from the meta-analysis, together with those from the two routine data samples. In general, the pseudo-RCT sample had event rates for the two procedures that were less than those for most individual trials. One might note, however, that the range of these rates across the eight trials was between

TABLE 11 *The baseline characteristics of patient samples extracted from routine SMR1 data compared with those of four published trials*

	Patient sample					
	Pseudo-RCT	**Excluded**	**BARI[50]**	**CABRI[52]**	**RITA[53]**	**EAST[54]**
Number of patients	12,238	2359	1829	1054	1011	393
Mean age (years)	58	62	61	60	57	62
Female (%)	24	33	27	22	19	26
Diabetes mellitus (%)	7*	7*	25	12	6	23
Previous MI (%)	40*	41*	55	43	43	41
Single-vessel disease (%)	43†	48†	2	1	45	0

**Determined from previous hospital admission; † determined from operation code (not diagnosis)*

TABLE 12 *Event rates for death and/or acute MI within 1 year in samples extracted from routine SMR1 data compared with those in the eight RCTs included in the Pocock meta-analysis[49]*

	Patient sample									
	Pseudo-RCT	**Excluded**	**CABRI**	**RITA**	**EAST**	**GABI**	**Toulouse**	**MASS**	**Lausanne**	**ERACI**
Number of events	377	288	72	65	57	28	12	6	8	15
Event rates for death or acute MI (%):										
CABG	3.7	13.6	5.7	6.2	18.4	10.2	7.9	1.4	3.0	10.9
PTCA	4.1	27.6	7.9	6.7	13.7	5.5	7.9	6.9	8.8	12.7

1.4% and 18.4% for CABG and between 5.5% and 13.7% for PTCA with only very small numbers of events in some studies. As expected, the event rates for the excluded sample were considerably higher.

The meta-analysis combining the results from these eight trials reported a relative risk for PTCA compared to CABG of 1.03 (95% CI, 0.84 to 1.27). A logistic regression was used to estimate the same relative risk, controlled for case mix, for the pseudo-RCT sample. The point estimate was 1.15, with a 95% CI of 0.90 to 1.48. A more detailed analysis, which modelled potential interactions between covariates and the treatment effect, provided evidence supporting an interaction with patients who had had an MI within 4 weeks prior to their intervention. This approach, therefore, has the potential for exploring subgroup effects, which, in this context, are plausible *a priori*.[51]

Revascularisation following CABG/PTCA

Figure 8, derived from a Cox regression model, shows survival free from revascularisation in the CABG and PTCA groups, for the pseudo-RCT sample. It is clear that PTCA patients are far more likely to require reintervention than CABG patients. The BARI and EAST trials reported similar analyses and the similarity in the shape of the survival curves from the two sources is striking. For the PTCA patients, the rate of reintervention was higher for the pseudo-RCT sample than for the excluded sample.

Mortality following CABG/PTCA

Table 13 summarises mortality at 1 year for the pseudo-RCT and excluded groups, and *Table 14* reports the corresponding results for deaths within 5 years. The 5-year results are based on relatively small numbers from early in the study period. For the pseudo-RCT patients, both unadjusted and adjusted for case mix, 1-year outcomes demonstrate that deaths among the CABG patients were significantly more frequent than those in the PTCA patients. This finding contrasts with the findings of the randomised trials where, based on substantially smaller numbers, no significant difference in mortality has been reported. The 1-year mortality rate of 3.2% for CABG patients is comparable to

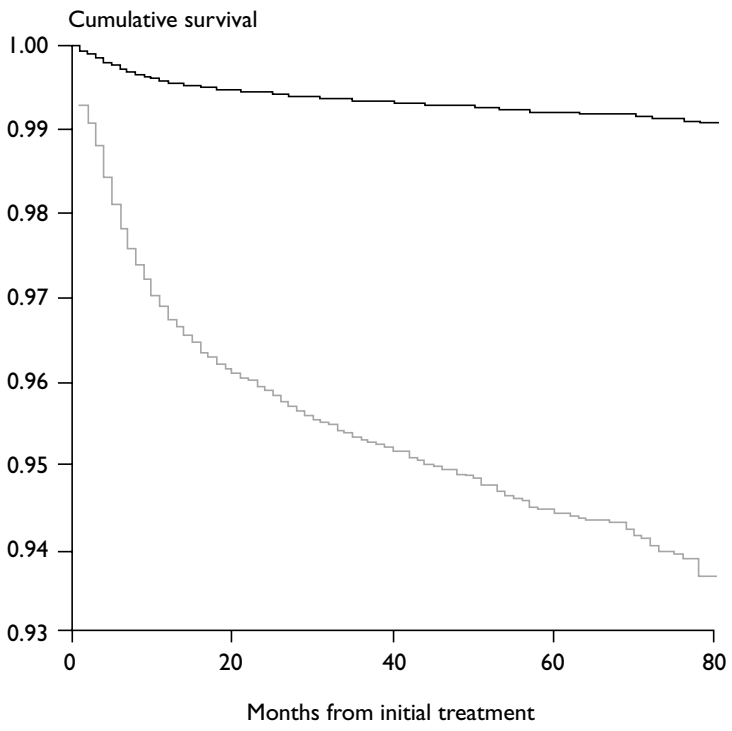


FIGURE 8 *Survival free from revascularisation for pseudo-RCT patients in the CABG (——) and PTCA (——) treatment groups*

TABLE 13 *Mortality rates and odds ratios between PTCA and CABG after 1 year in the two patient groups extracted from routine SMR1 data*

	Mortality rate		Odds ratio PTCA:CABG (95% CI)	Adjusted odds ratio PTCA:CABG (95% CI)
	CABG	PTCA		
Pseudo-RCT patients	3.2%	1.7%	0.54 (0.40 to 0.73)	0.68 (0.49 to 0.93)
Excluded patients	13.5%	7.1%	0.53 (0.36 to 0.78)	0.41 (0.25 to 0.68)

TABLE 14 *Mortality rates and odds ratios between PTCA and CABG after 5 years in the two patient groups extracted from routine SMR1 data*

	Mortality rate		Odds ratio PTCA:CABG (95% CI)	Adjusted odds ratio PTCA:CABG (95% CI)
	CABG	PTCA		
Pseudo-RCT patients	8.9%	7.2%	0.81 (0.59 to 1.11)	1.00 (0.70 to 1.43)
Excluded patients	27.0%	3.3%	0.12 (0.03 to 0.48)	0.12 (0.03 to 0.59)

the rate of 2.7% reported in the CABRI trial. This trial reported a rate of 3.9% for the PTCA patients, however, and this is more than twice the rate observed for the pseudo-RCT patients.

In the pseudo-RCT sample, neither the unadjusted nor the adjusted results provide evidence of a difference in mortality at 5 years between the CABG and PTCA groups. This is consistent with the results from the randomised trials, but there is a similar pattern to that observed at 1 year: the pseudo-RCT mortality rate for CABG was similar to that observed in the randomised trials, but the rate for PTCA (at 7.2%) was lower than that reported for the BARI and EAST trials.

The results for the excluded patients show the expected high mortality rates in the CABG group. The low 5-year mortality rate for the excluded PTCA patients could reflect the careful selection of patients with complicated coronary disease in the early years of developing this procedure.

Synthesising evidence from randomised and non-randomised studies using Bayesian methods

The odds ratio for mortality after 1 year in the pooled meta-analysis[49] was 1.03 [95% CI, 0.84 to 1.27]. For pseudo-RCT patients, the odds ratio was 1.15 [95% CI, 0.90 to 1.48] after case mix adjustment. These results appear to be in quite close agreement, but a more formal assessment is needed in order to have greater confidence in this conclusion. Two Bayesian approaches are used below as a way of synthesising the two sets of results. (Discussions on the use of Bayesian methods in health technology assessment[55] and RCTs[56] can be found elsewhere.)

Approach 1: forming a predictive distribution

For this approach, the eight trials included in Pocock's fixed-effects meta-analysis were re-analysed employing a Bayesian random-effects model[57] in which:

$$
\begin{aligned}
y_i &= \ln(\text{odds ratio}) \text{ for the } i\text{th study} \\
s_i &= \text{standard error of } \ln(\text{odds ratio}) \text{ for the } i\text{th study} \\
y_i &\sim N(\psi_i, s_i^2) \\
\psi_i &= \mu + \rho z_i \\
z_i &\sim N(0, 1)
\end{aligned}
\qquad [1]
$$

and where ψ_i denotes the true treatment effect on the ln(odds ratio) scale in the ith study; μ and ρ^2 denote the population mean and between-study variance, respectively; and z_i are random effects measuring the differences between the treatment effect in individual studies from the overall population effect.

The approach requires prior distributions to be placed on the unknown parameters μ and ρ. For a fully Bayesian approach, these prior distributions should reflect knowledge known about the parameters. This can take the form of a clinician's beliefs or information from previous

studies. For this synthesis, non-informative priors were used so that outcomes were driven from evidence provided by the data.

The random-effects model [1] was set up using the BUGS software.[58] The overall treatment effect was found to be $e^{\mu} = 1.03$ [95% CI, 0.78 to 1.37]. The point estimate was the same as that derived from the fixed-effects meta-analysis, but the random-effects CI was considerably wider than the fixed-effects CI. This is because the random-effects model accounts for heterogeneity between the studies.

A predictive distribution can be formed from this modelling procedure and can be regarded as the distribution of possible odds ratio values for a new study (randomised or non-randomised) deriving from the evidence of the RCTs. This is done by monitoring a new effect in the random-effects model [1], for example $\mu(pred)$, which incorporates the uncertainty associated with μ, the overall treatment effect, and ρ^2, the between-study variance.[57] The distribution of $\mu(pred)$ is shown in *Figure 9*, along with the odds ratio point estimate and CI for cardiac death and/or acute MI within 1 year formed from the routine data (pseudo-RCT) patients. It is clear that the pseudo-RCT estimate is wholly compatible with the RCT evidence.

Approach 2: accounting for different study types

For this approach an extension of the random-effects model [1] was used to account for the fact that some of the cardiac death or acute MI evidence stems from RCTs and some is attributable to the routine data. The model is detailed below:

$$
\begin{aligned}
y_{ij} &\sim N(\psi_{ij}, s_{ij}^2) \\
\psi_{ij} &= \theta_j + \sigma_j z_{ij} \\
\theta_j &= \mu + \tau\varepsilon_j \\
z_{ij} &\sim N(0, 1) \\
\varepsilon_j &\sim N(0, 1)
\end{aligned}
\qquad [2]
$$

where ψ_{ij} is the treatment effect of the ith study of type j; θ_j is the treatment effect for studies of type j; μ is the overall treatment effect; σ_j^2 is the variance between studies of type j; and τ^2 is the between-study type variance. The random effects z_{ij} and ε_j reflect the differences in the true intervention effect in individual studies from the overall effect for study type j, and

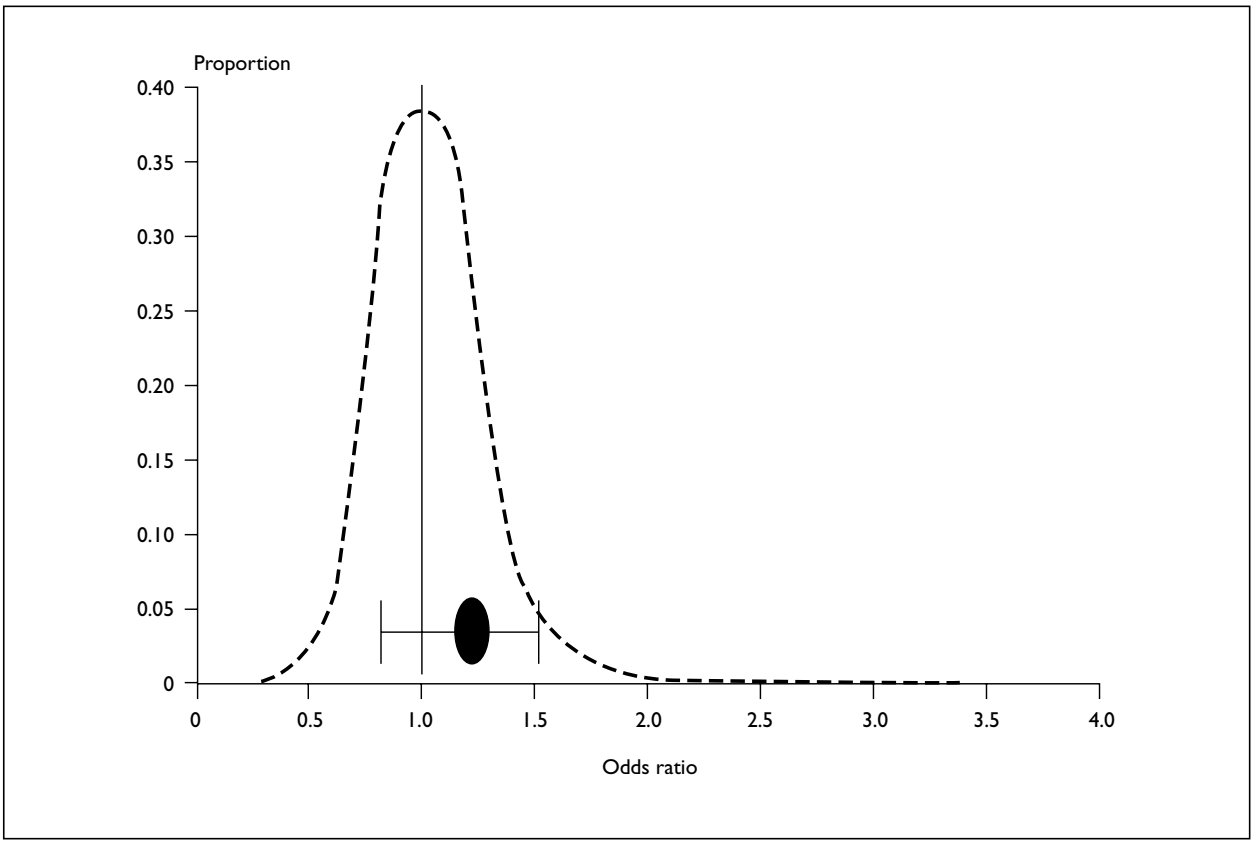


FIGURE 9 *Predictive distribution of possible odds ratio values for cardiac death or acute MI at 1 year derived from RCT evidence, and the corresponding odds ratio for pseudo-RCT patients*

the differences in the true intervention effect for study types *j* from the overall population effect, respectively. For this approach the pseudo-RCT patients were split according to the hospital in which they received their PTCA or CABG. This resulted in three centres; that is, Glasgow, Edinburgh and Aberdeen.

As for Approach 1, prior distributions need to be placed on the unknown model parameters, in this case, for σ_j, τ and μ. As before, non-informative priors were employed. The random-effects model accounting for study type [2] was set up using the BUGS software and the results are displayed in *Figure 10*. The estimate of the pooled odds ratio for the randomised studies is $e^{\theta}{}_1 = 1.08$ [95% CI, 0.83 to 1.42] and for the non-randomised (routine) studies is $e^{\theta}{}_2 = 1.13$ [95% CI, 0.86 to 1.52]. The estimate of the overall pooled odds ratio is $e^{\mu} = 1.11$ [95% CI, 0.81 to 1.55].

A comparison of the pairs of lines in *Figure 10* shows how model [2] has considerably reduced the width of the CIs for effects that are study specific and ignore study type. This is because model [2] allows the randomised evidence to contribute to the treatment effect in the routine evidence, and vice versa.[57]

One advantage of a Bayesian approach is that practitioners are able to incorporate their own beliefs by means of the prior distributions of the unknown parameters. It might be argued, for example, that the variability between RCTs will be less than the variability between data from centres providing routine services because of the homogeneous way in which RCTs are performed. *Table 15* shows the estimated overall, randomised and non-randomised treatment effects (μ, θ_1, and θ_2, respectively) that result from assuming different values for a, the ratio of $\sigma_2{}^2$ and $\sigma_1{}^2$.

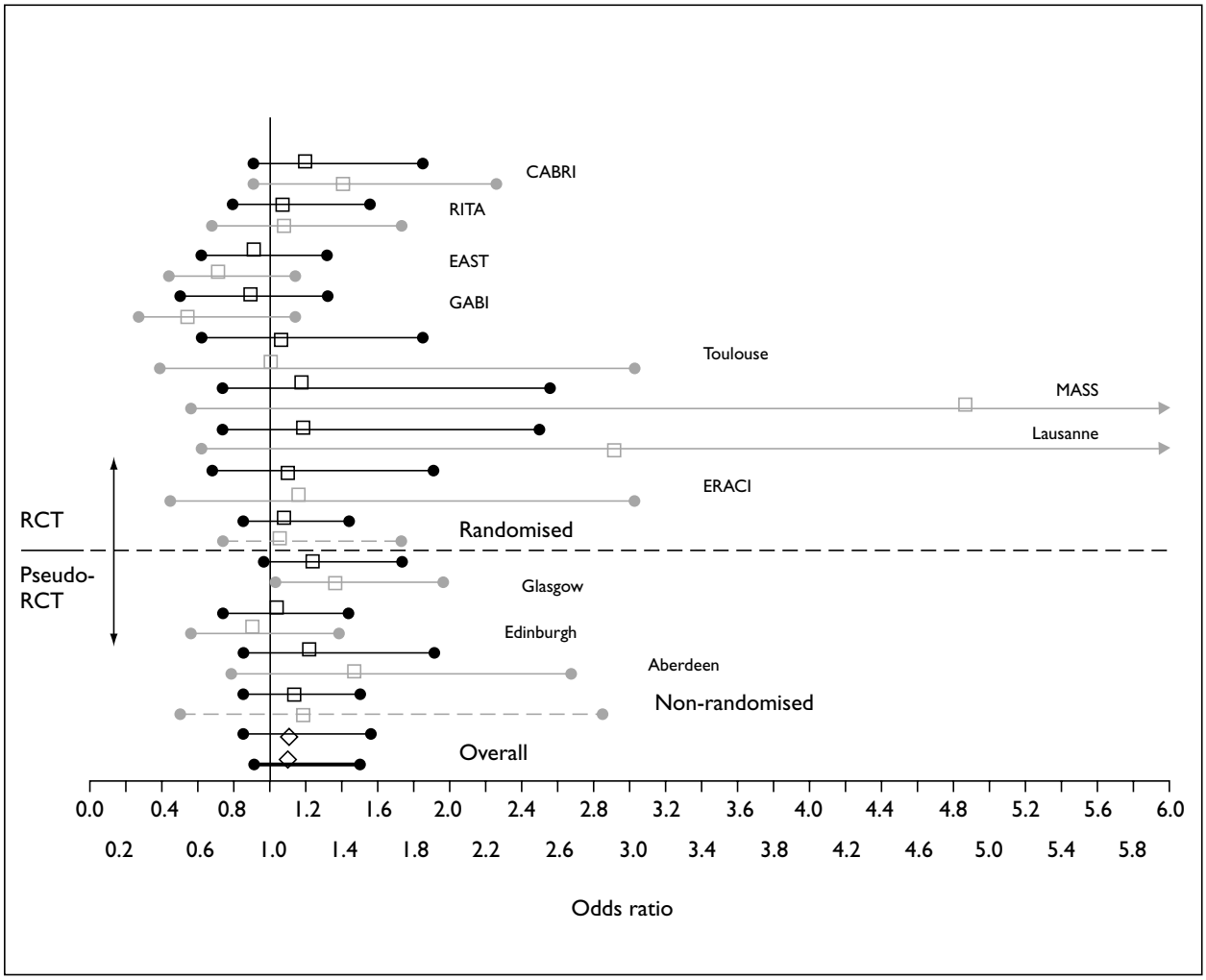


FIGURE 10 *Estimated odds ratio with 95% CI of cardiac death or acute MI at 1 year obtained from the random-effects model accounting for study type [2]. Study-specific treatment effect estimates (——); random-effects meta-analysis on each type (——)*

TABLE 15 *Posterior estimates of μ, θ_1, and θ_2, given different prior beliefs of σ_2^2 and σ_1^2 (95% CI in round brackets, width of 95% CI in square brackets)*

a	*μ*	θ_1	θ_2
2	1.09 (0.77 to 1.56) [0.79]	1.06 (0.82 to 1.37) [0.55]	1.12 (0.81 to 1.55) [0.74]
3	1.08 (0.77 to 1.52) [0.75]	1.06 (0.82 to 1.34) [0.52]	1.11 (0.79 to 1.55) [0.76]
4	1.08 (0.77 to 1.52) [0.75]	1.05 (0.83 to 1.33) [0.50]	1.10 (0.78 to 1.54) [0.76]
5	1.08 (0.77 to 1.53) [0.76]	1.05 (0.83 to 1.32) [0.49]	1.10 (0.78 to 1.55) [0.77]
6	1.07 (0.77 to 1.51) [0.74]	1.05 (0.83 to 1.32) [0.49]	1.10 (0.78 to 1.55) [0.77]
7	1.08 (0.77 to 1.52) [0.75]	1.05 (0.83 to 1.31) [0.48]	1.10 (0.78 to 1.55) [0.77]
8	1.07 (0.76 to 1.50) [0.74]	1.05 (0.83 to 1.31) [0.48]	1.10 (0.78 to 1.54) [0.76]
9	1.08 (0.77 to 1.52) [0.75]	1.05 (0.83 to 1.31) [0.48]	1.10 (0.78 to 1.55) [0.77]
10	1.07 (0.77 to 1.50) [0.73]	1.05 (0.83 to 1.31) [0.48]	1.10 (0.78 to 1.56) [0.78]

It is evident that changing beliefs about variability between the RCTs compared with variability between the centres contributing to the routine data has not greatly affected the posterior results. This insensitivity is pleasing because it suggests robust results: although marginal, the effect of assuming greater variability between the three centres is that the overall treatment effect moves towards the pooled randomised effect.

It might also be argued that randomised studies have less bias than the routine data – in which case the random-effects ε_2 will be greater than ε_1. This seems reasonable because the earlier logistic regression demonstrated selection bias in the pseudo-RCT population. *Table 16* reports the estimated overall, randomised and non-randomised treatment effects (μ, θ_1, and θ_2, respectively) obtained by assuming different values for *b*, the ratio of ε_2 and ε_1.

Changing the prior beliefs about biasing had an effect on outcome estimates. Although the point estimates do not alter greatly, the widths of the CIs are sensitive to the ratio of ε_2 and ε_1. The CI for the overall treatment effect narrows as the suggestion that there is less bias amongst the RCT results increases, thus providing greater assurance that the point estimate of the relative risk is correctly estimated. Not surprisingly, the width of the CI associated with the pooled non-randomised estimate increases as it is proposed that more bias is associated with the routine data centres.

The uses of Bayesian methods

The immediate value of these illustrations is to provide support for our earlier conclusion that the results of the pseudo-RCT analysis were essentially the same as those derived from the Pocock meta-analysis.[49] A rather more funda-

TABLE 16 *Posterior estimates of μ, θ_1, and θ_2, given different prior beliefs of ε_2 and ε_1 (95% CI in round brackets, width of 95% CI in square brackets)*

b	*μ*	θ_1	θ_2
2	1.05 (0.66 to 1.59) [0.93]	1.09 (0.84 to 1.42) [0.58]	1.12 (0.85 to 1.48) [0.63]
3	1.00 (0.69 to 1.52) [0.83]	1.07 (0.81 to 1.43) [0.62]	1.14 (0.84 to 1.56) [0.72]
4	1.00 (0.72 to 1.49) [0.77]	1.07 (0.80 to 1.43) [0.63]	1.15 (0.84 to 1.59) [0.75]
5	1.04 (0.72 to 1.48) [0.76]	1.06 (0.79 to 1.43) [0.64]	1.15 (0.83 to 1.62) [0.79]
6	1.04 (0.73 to 1.47) [0.74]	1.06 (0.79 to 1.43) [0.64]	1.16 (0.82 to 1.64) [0.82]
7	1.05 (0.74 to 1.49) [0.75]	1.06 (0.79 to 1.45) [0.66]	1.16 (0.83 to 1.66) [0.83]
8	1.05 (0.74 to 1.48) [0.74]	1.06 (0.78 to 1.45) [0.67]	1.16 (0.83 to 1.67) [0.84]
9	1.04 (0.74 to 1.46) [0.72]	1.06 (0.78 to 1.43) [0.65]	1.16 (0.82 to 1.68) [0.86]
10	1.05 (0.75 to 1.46) [0.71]	1.06 (0.78 to 1.44) [0.66]	1.17 (0.82 to 1.67) [0.85]

mental value is that they provide a basis for linking two kinds of information: by design, that from RCTs is unbiased, but – also because of their design – may not reflect experience from the real world. Information from routine data is more likely to reflect the real world, but is also likely to be biased – sometimes in unknown ways. The problem of evaluating the true benefits of a new technology, therefore, is that of finding a balance between information from the two sources. Bayesian approaches involve judgements about expected outcomes so that their use in this and similar contexts provides a means of augmenting the evidence from RCTs and supporting (or otherwise) the utility of their findings in the practice of the real world. A discussion of classical and Bayesian approaches to meta-analysis can be found elsewhere.[59]

Conclusions

Using routine data for the study of coronary revascularisation

Of the three studies we report, the assessment of coronary artery vascularisation is the most satisfactory from the perspective of parallels with the methods of RCTs. This was mainly because the data included in the SMR1 records were fairly close to those employed in the studies contributing to the Pocock meta-analysis,[49] making it possible to make fairly direct comparisons between the relative risks estimated by the two approaches. This result is important from the standpoint of comparing the results achieved in the context of specially designed studies with those observed when these procedures are more generally adopted. In this sense, the pseudo-RCT confirms (or supports) the conclusions of the meta-analysis when it can be regarded as an 'audit' of outcomes from these interventions at the stage following RCT evidence of their effectiveness.

This said, however, our earlier comments about the completeness of routinely assembled data – and their use as a way of answering questions in circumstances where RCTs may be difficult or expensive – are also illustrated by this study. Again, the problem turns on the lack of data that might add precision to information about the complexity or severity of ischaemic heart disease and thus the diagnostic mix of different patient subgroups. We note above that there was evidence of interactions with other covariates in this analysis and that, as expected, outcomes for the excluded group of patients were less favourable than those for patients who satisfied the RCT criteria. As the applications of new technologies spread beyond the patient criteria on which original RCTs are based, it is desirable to evaluate outcomes in these other categories of patient. It would, of course, be possible to undertake further RCTs with these objectives, but their design might well be complex and a fairly large number of such studies might be necessary. There is now an increasing body of literature on the dangers of extrapolation from the results of RCTs to wider populations.[60] Given the size of the population on which the pseudo-RCT sample that we report was based, a more comprehensive approach to the assembly of routine data has the potential for alternative approaches.

Chapter 6

Other possible studies

A number of other case studies were proposed in the original grant application:

(i) a comparison of laparoscopic with open surgery for cholecystectomy and for hernia repair
(ii) a comparison of modes of treatment for female breast cancer
(iii) a comparison of outcomes following surgery for colorectal cancer performed by general and specialist surgeons.

Our inability to undertake these investigations illustrates two general features of the inadequacies of routine data as a complement to RCTs. We discuss these in greater detail below; in brief they are, first, the need to ensure that routine data provide a more precise account of the clinical characteristics of patients and the objectives of their care (for example, by including information about the staging of cancers and the distinction between curative interventions and palliative care) and, second, the need to maintain up-to-date coding systems that properly reflect therapeutic procedures in a timely way.

These additional case studies were not taken further, principally because early explorations of the data demonstrated that their limitations precluded useful analysis of them. In terms of the broader purposes of the research, these constraints are exemplified by the three case studies described above. More specific reasons for our inability to pursue them further are listed below.

(i) For the cholecystectomy and hernia studies, there is no coding within OPCS4 to identify laparoscopic procedures.
(ii) For breast cancer, the choice of management (especially mastectomy versus local excision) depends on clinical characteristics (such as the size of the tumour and nodal status) that are not recorded. Any analysis of outcomes would have been seriously confounded by unknown case mix variables. Unrecorded adjuvant therapy would also have a major impact on outcomes.
(iii) For colorectal cancer, issues of confidentiality mean that the identity (and specialty) of the operating surgeon is not included in the data set provided routinely by the Information and Statistics Division of the NHS. There are mechanisms to allow access to this additional information, which were followed, for example in the SAH case study in order to allow comparisons between the SMR1 with the operating theatre log of one neurosurgical centre. Acquiring these data for the very large number of cases involved would have been a substantial task. As with the breast cancer example, necessary case mix variables such as Dukes' grade and operative intent (curative versus palliative) did not form part of the data available to us. Colorectal cancer specialists are likely to treat a high proportion of patients with advanced disease and so our inability to perform appropriate case mix adjustment would have seriously compromised the treatment comparisons we might have made.

Chapter 7
Conclusions

The uses of routine data

In the introduction of this report, the argument was put forward that routinely assembled NHS data might have value as a complement or alternative to RCTs in certain circumstances. It was, also, suggested that these data might have potential value for determining whether the benefits expected from the evidence provided by RCTs were achieved when new technologies or procedures were adopted in the wider framework of clinical practice. The three studies described were chosen as a way of exploring different aspects of these propositions while also testing the utility of presently available routine data for these purposes. The use of the Scottish linked SMR1 data set had value in this context because these data have two important advantages: first, that they relate to all hospital admissions over a long period with the benefit that (at least for hospital admissions) they describe complete populations of cases and can thus possibly clarify issues relating to case selection or sampling. Second, by linking episodes of care to each other and to deaths, it is possible to gain information about prior medical histories and subsequent outcomes. Although the content of individual SMR1 records is broadly similar to those used elsewhere in the UK, these added features are not usually a part of the routine data sets created in other systems. Both were a necessary part of the analyses described and are probably a prerequisite for comparable work of this kind.

Support for this conclusion is found in all three illustrative studies. In that concerned with SAH, the sequence of samples describing the epidemiology of hospital care and the patterns of its provision is an illustration of the potential for locating the care provided in specialised centres and thus a context for more focused (RCT) studies of particular interventions. The initial objective in this example – that of devising an alternative in circumstances where a formal RCT was not feasible – failed for two main reasons: first, the lack of necessary detail in the data and, second, our quite serious reservations about the quality of the data. On the other hand, the **potential** (we stress the word) was considerable; it was possible to assemble information about a large series of cases of a fairly uncommon condition and – had the data permitted – to undertake multicentre comparisons of the uptake of a new technique and its outcomes. We return to questions about the scope and quality of routine data later in this discussion; it is important, however, that questions about the potential of routine data should not be dismissed because of the inadequacies of present circumstances.

Problems with the scope and content of the data were also evident in the study of prostatectomy although – even with these limitations – this analysis illustrated other ways in which routine data can complement information from RCTs. The RCT results favouring TURP rather than OPEN are now fairly old – TURP has become the procedure of choice and new RCTs comparing the two procedures (perhaps for selected groups of patients) are unlikely for this reason. The uses of routine data in these contexts rest mainly on a further need to take account of co-morbidities or other case mix variables and to employ reasonably long timeframes when evaluating outcomes. Both are difficult in most RCT designs, partly because of the size of the populations that would be necessary and partly due to the need to assess outcomes over quite long periods. These activities are not a substitute for RCTs; instead, they provide a possible means of refining or improving conclusions reached by RCTs and answering the question of whether the benefits they conclude apply to a broader range of patients. The prostatectomy study also highlights a further deficiency of present routine data, which is that the interventions they report are almost always surgical; data of this kind rarely report drug regimes. However, the general category of applications sketched above is likely equally pertinent for the evaluation of long-term drug therapies.

The content of the data is particularly important when considering the issues surrounding **risk adjustment** or **severity adjustment**. Each of the case studies relied on the use of patient characteristics as a means of adjusting for differences between hospitals in terms of their case mix,

but such adjustments are obviously restricted to such information as is recorded in the data. A substantial body of literature exists on the use of routine data to adjust for differences between patients both generally[61–66] and with specific reference to CABG.[67,68]

We were fortunate in the study of coronary revascularisation, partly because the Pocock meta-analysis[49] provided a more direct basis for comparing results from two sources and partly because the routine data were sufficiently complete to devise a pseudo-RCT that permitted this comparison. In our view, this is an important illustration because it suggests that a future use of (augmented) routine data could be at this second stage of RCT methodology rather than as a substitute for initial RCTs. This is not to argue against meta-analysis as a means of pooling data from a number of single RCTs; instead, it is to repeat the suggestion that routine data can have value for estimating benefits from the more general application of RCT findings. In practical terms – and as the analysis in this study indicates – the two activities are fairly close together. Our account of the Bayesian synthesis of evidence from both approaches is an example of this possible complementarity and of how further methodological exploration along these lines might have value.

The link between RCTs as an investigative method and the movement towards evidence-based medical practice rests on the argument that evidence from RCTs should inform clinical judgements about appropriate interventions in the care of individual patients. Extensions of this argument suggest that such evidence is also relevant to larger-scale assessments – specifically, in the audit of clinical practice and the wider question of whether new technologies achieve expected benefits for the health of populations. These latter issues are important for both contemporary policy and practice because they underpin the present need to achieve accountability and yet it is difficult to see how they can be satisfied except through the development of routine data systems that permit a closer rapprochement with the methodological standards of RCT practice. The case studies reported provide imperfect examples of how this potential might be addressed while illustrating the shortcomings of present data systems for these purposes.

There are a number of ways of characterising these deficiencies. One is simply that, historically, routine data have been assembled for largely administrative purposes and have been employed at fairly high levels of abstraction to monitor broad trends in hospital activity. This being so, there has, perhaps, been little incentive to maintain them in ways that might encourage or facilitate applications that require greater precision or improved insights into the content of clinical care.[69] In terms of their uses for (say) the evaluation of clinical outcomes, this leads to the further problem that the collection of data generally lacks specific purposes: a 'generic' data set is assembled and is then employed (within its limits) for post hoc purposes that may be devised.[70] This is a conservative approach that limits opportunities for expanding and developing the uses that might be made of the data for either clinical or administrative purposes. Given the growing need to describe NHS activities in terms of outcomes and benefits, this is an important shortfall of data collection systems.[42]

There are two possible explanations for the present generic forms of data collection. One is that it is difficult to manage the collection and validation of data when many hospitals are involved. This argument reflects perceptions of the possible uses of the data – 'fit for function' is a phrase that has been used in this context – and raises doubts about relationships between the management of hospitals and the content and quality of the services they provide. The discrepancies encountered in the case study of SAH illustrate this distinction and emphasise the possible need to strengthen the link between providers of care and those who generate and use routine records. It will be evident that feedback to clinicians is a central feature of this requirement for quality control.

An anonymous referee commented on the choice of the Institute of Neurological Sciences in Glasgow as the centre in which to study the validity of the SMR1 data, stating that the data quality within this unit was known to be poorer than for other specialties in the same hospital (referencing a personal communication from the Information and Statistics Division of the NHS). The centre was chosen because of existing collaborative links and two interesting features of routine data are raised with respect to the claim that the quality of neurosurgical data from this unit is worse than from other such units. The first is the point made in the previous paragraph concerning feedback to clinicians; if they are not aware that their data are deemed to be of poor

quality then there is little that they can do to improve it. Second, any national system of routine data can only be as strong as its weakest link; in this particular case, the unit in question provides neurological services to approximately half the country.

The second possible reason for maintaining generic forms of data collection is that diagnostic and therapeutic variation and complexity may be too great to permit the inclusion of more clinical data in the records. This is an important argument from the standpoint of national data systems but also problematic when considered in terms of trying to ensure that evidence – including that from RCTs – is translated into practice and patient benefit. Some improvements could be fairly simple, including expanding and updating coding systems so that they provide a more up-to-date account of new procedures and treatments. The adoption of ICD-10 codes is likely to make diagnostic codes more sensitive when fully employed, but there are parallel arguments for continuing revisions of the OPCS procedural codes to ensure that both new therapies and new combinations of therapies are adequately described. Feedback to clinicians, including publication of audits of the data's quality, is one way of monitoring the extent to which the data reflect current practices.

One way of circumventing the limitations of generic data sets is to initiate high-quality clinical databases.[69] Another way of overcoming their limitations might be to develop a two-level approach in which basic data continue to be collected as at present but are complemented by more focused investigations of topics of current interest in ways that have analogies with multicentre clinical trials. Essentially, this approach would involve agreement between clinicians about the additional data to be collected for particular diagnostic categories or therapies, which could occur for an appropriate period across the whole system and as a way of answering predetermined questions. There are certainly arguments for more extensive information in the study of coronary artery revascularisation, particularly with regard to covariates not included in the available data. The collection of agreed additional data is even more strongly suggested by our inability to undertake the studies of patient outcomes for breast and bowel cancers because the data were insufficient for these purposes. A more focused approach might also mean that therapies that are not now included in the data (such as most drugs) could also be evaluated. The case studies undertaken were restricted to OPCS definitions of procedures and thus necessarily surgical.

This proposal for a 'rolling programme' of studies selected for their contemporary significance appears to have a number of different merits. It might provide better answers to questions about the relationship between RCTs and routine data with which this study began: expanding the content of routinely collected data targeted towards answering particular questions could bring RCT methods (and their advantages) closer to evaluations of wider hospital practice and provide an insight into the ways in which new technologies are adopted. In essence, this could help in the monitoring of the implementation of evidence, or in answering questions for which more usual RCT methods are inappropriate. More generally, however, this approach might offer ways of engaging clinicians in the larger process of routine data assembly, with advantages for their quality and subsequent use in such activities as clinical governance. As methods of outcome assessment evolve it is important to ensure that the data on which they depend are also developed in ways that permit more secure uses of outcome measures. One part of this requirement is the need to link the increasing sophistication of the statistical methods used in these activities and the scope of the data they employ.

In summary, the main conclusions are as follows.

- Routine data have potential applications for either complementing or supplementing RCTs as a source of evidence in the assessment of new healthcare technologies; information from routine data could improve both the design of RCTs and the interpretation of their findings.
- There are great limitations in using routine data for these purposes, partly due to deficiencies in their content and partly because of the quality of the information they report. These two issues are linked in terms of the purposes for which the data are collected – short of formal RCT practice, there are strong arguments for a closer rapprochement between the administrative uses of routine data and their use as a way of providing an account of patient care and its outcomes; doing so is important for evaluating the applications of evidence from RCTs.
- Improved accounts of the content and implications of patient care are desirable even for administrative purposes. Developments in

information technology mean that it is increasingly possible to assemble data for both longer-term (administrative) and shorter-term (clinical) purposes, and attention should, perhaps, be given to the reform of data systems that can serve both purposes, therefore, enhancing the contextual interpretation of each.
- One problem with RCTs is extrapolation of their findings to larger (or less precisely defined) populations; a difficulty with routine data is that of achieving sufficiently well-defined populations for comparisons of outcomes. Part of the complementarity of the two is the potential for resolving such issues (for example, the relevance of co-morbidities). Similarly, routine data could expand the scope of RCT practice by enabling multicentre studies, either as RCTs or as analogies of them. As demonstrated in the illustration of Bayesian methods, there is a need to develop the applications of appropriate statistical techniques with regard to these alternative ways of evaluating the outcomes of healthcare interventions.

The purposes of evaluative research are to determine whether or how interventions and medical technologies are of benefit to patients. Although regarded as a 'gold standard' for some kinds of evidence, RCTs may fall short of a comprehensive answer to such questions either because of possible limitations in their design (which includes restrictions on the scope of the questions they are able to ask) or because there is the subsequent question of extrapolating their findings to the more general world of clinical practice. In the past, routine data systems were devised to answer different kinds of questions – mainly those to do with the deployment of health service resources. Contemporary preoccupations, both clinical and administrative, are bringing these worlds closer together with benefit to the patient as their common objective. The need to bring them together – to the advantage of both – is the principal reason for improving the utility of routine NHS data for clinical purposes.

Implications for policy

- There has been a shift in the purposes for which routinely assembled NHS data might be used towards their applications for clinical audit and governance, for assessments of the outcomes of care, and the extent to which clinical evidence is deployed in practice. Routine data systems do not appear to have kept pace with these developments, and it, therefore, may be necessary to examine the processes of data collection, specifically with regard to the quality and content of the data, the extent to which coding systems reflect actual practice, and the investment in local arrangements for doing so. This might be done by publishing audits of the quality and validity of the data, with follow-up of identified deficiencies and the reasons for them. Doing so may go some of the way towards dispelling (or confirming) present perceptions that data of this kind have only limited use for local clinical evaluations of clinical practice.
- We have commented on the apparent distance between the uses that are made of the data (usually by central agencies) and feedback of the findings of these applications to those responsible for collecting the data on which they are based. If the quality and validity of the data are to be improved, there are strong arguments for improving this relationship and encouraging local uses of the data. This view has implications for local support of information departments (including a clearer definition of responsibility for the data) together with a central response in terms of returning data quickly in ways that – for example – make it possible to compare clinical services in one setting with those of another. Moreover, there is a need to ensure that quality improvement of routine data systems is implemented as a result of quality control.
- We have argued that one way forward in making more effective use of routine data might be to explore ways of expanding their content to facilitate a closer relationship between RCT evidence (and the methods that RCTs employ) and the uses that can be made of routine data. Central support for the development of these activities may improve the uses of routine data for studies of outcomes and, more generally, may encourage good practice in the analysis of the data and its interpretation. Common ownership of the data and its uses could also be encouraged.

Further research

This report describes three case studies that explore the application of routine data as alternatives to formal RCT designs. Our broad

conclusion is that such data can be valuable for such uses but are limited in four important ways:

(i) the scope and content of the data may not include the information needed for the assessment of clinical practice
(ii) uncertainties about the quality of the data may lead to questions about the validity of findings from studies that employ them
(iii) their responsiveness to changes in practice as illustrated by delays in adaptation of their coding systems
(iv) their inability to distinguish different purposes for hospital care – for example, that between curative and palliative care.

In each of these ways, studies based on routine data fall short of the rigour that one expects in RCT designs when comparability between the different arms of a trial is fundamental to the assessments that are made.

On the other hand, routine data can have substantial advantages. The very large size of the data sets means that it is at least potentially possible to take better account of covariate interactions. Their continuity can provide a longer timescale for assessment than is often possible with RCT designs, and the fact that they can describe practice in a wide variety of different settings means that studies employing them can reflect the diversity of everyday clinical practice in more realistic ways than the constraints implicit in most RCTs. In addition, we have noted that there are also other ways in which routine data can be employed to address questions that – for various reasons – are not susceptible to RCT investigations.

One of the suggestions in this discussion has been the possible need for a closer rapprochement between evidence about the effectiveness of a technology derived from RCTs and a demonstration that the expected benefits are achieved in practice. The potential for investigating this link between issues is illustrated by the case study of coronary revascularisation. The broad research question emerging from the case studies, therefore, concerns this link and the ways in which it might be promoted. Overcoming our criticisms of the present system of routine data collection suggests one area in which change might be necessary, and there is scope for specific research projects in this area: the feasibility and implications of a more public audit of the system and appropriate ways of feeding information back to those who provide is one example. Although only sketched briefly above, expanding records for particular purposes along the lines of multicentre trials is likely worthy of more detailed consideration and has the potential for demonstration projects that might identify the strengths and weaknesses of the proposal. Finally, the question of technical ('good practice') support for uses of the data in areas such as clinical governance requires exploration – again, perhaps, in the form of demonstration projects that can explore the needs of local users and feedback to the data system itself as a way of improving its ability to respond to changing patterns of practice.

Acknowledgements

We are grateful to the Information and Statistics Division of the NHS in Scotland for the provision of the linked hospital discharge and mortality records. We are also grateful to our neurosurgical colleagues at the Institute of Neurological Sciences, Southern General Hospital, Glasgow, especially Mr Philip Barlow, for their help with the SAH study; and to Dr Keith Abrams, University of Leicester, for his advice on the Bayesian analysis. The project was funded by the UK NHS as part of its Health Technology Assessment programme, and the Public Health Research Unit was supported financially by the Chief Scientist Office of the Scottish Office Department of Health. Opinions and conclusions contained in this report are those of the authors, who are responsible for any errors, and do not necessarily reflect those of the Scottish Office Department of Health. We also thank the referees for their perseverance in reading the report and the quality of their comments.

References

1. Joffe M, Chapple J, Paterson C, Beard RW. What is the optimal caesarean section rate? An outcome based study of existing variation. *J Epidemiol Community Health* 1994;**48**:406–11.
2. Leyland AH, Pritchard CW, McLoone P, Boddy FA. Measures of performance in Scottish maternity hospitals. *BMJ* 1991;**303**:389–93.
3. Berry DA, Eick SG. Adaptive assignment versus balanced randomization in clinical trials: a decision analysis. *Stat Med* 1995;**14**:231–46.
4. Ball SG. ACE inhibition in acute myocardial infarction. In: Ball SG, editor. Myocardial infarction: from trials to practice. Petersfield: Wrightson Biomedical Publishing; 1995.
5. Hall AS, Murray GD, Ball SG. ACE inhibitors in and after myocardial infarction. In: Pitt B, Julian D, Pocock S, editors. Clinical trials in cardiology. London: WB Saunders; 1997.
6. The ASPIRE Steering Group. A British Cardiac Society survey of the potential for the secondary prevention of coronary disease: ASPIRE (Action on Secondary Prevention through Intervention to Reduce Events). *Heart* 1996;**75**:334–42.
7. Sapsford RJ, Hall AS, Robinson MB. Evaluation of angiotensin-converting enzyme inhibitor dosage and prescription rates in post-myocardial infarction patients. *Eur Heart J* 1996;**17**:62 (Abstract).
8. Cochrane AL. Effectiveness and efficiency: random reflections on health services. London: Nuffield Provincial Hospitals Trust; 1972.
9. Byar DP. Why data bases should not replace randomised clinical trials. *Biometrics* 1980;**36**:337–42.
10. Spiegelhalter DJ, Freedman LS, Parmar MKB. Bayesian approaches to randomized trials (with Discussion). *J Roy Statistical Soc, Series A* 1994;**157**:357–416.
11. Berry DA. A case for Bayesianism in clinical trials. *Stat Med* 1993;**12**:1377–93.
12. Fendrick AM, Escarce JJ, McLane C, Shea JA, Schwarz JS. Hospital adoption of laparoscopic cholecystectomy. *Med Care* 1994;**32**:1058–63.
13. Kendrick S, Clarke J. The Scottish record linkage system. *Health Bull (Edinb)* 1993;**51**:72–9.
14. World Health Organisation. Manual of the international statistical classification of diseases, injuries and causes of death. Ninth revision. Geneva: WHO; 1977.
15. Office of Population Censuses and Surveys. Classification of surgical operations and procedures. Fourth revision. London: OPCS; 1987.
16. Hop JW, Rinkel GJE, Algra A, van Gijn J. Case-fatality rates and functional outcome after subarachnoid haemorrhage. A systematic review. *Stroke* 1997;**28**:660–4.
17. Jennett B, Lindsay KW. An introduction to neurosurgery, 5th Edition. Oxford: Butterworth Heinemann; 1994.
18. Linn FHH, Rinkel GJE, Algra A, van Gijn J. Incidence of subarachnoid haemorrhage. Role of region, year, and rate of computed tomography: a meta-analysis. *Stroke* 1996;**27**:625–9.
19. Mayberg MR, Batjer H, Dacey R, Diringer M, Haley EC, Heros RC, *et al.* Guidelines for the management of aneurysmal subarachnoid hemorrhage. A statement for healthcare professionals from a special writing group of the Stroke Council, American Heart Association. *Circulation* 1994;**90**:2592–605.
20. Drake CG. Report of World Federation of Neurological Surgeons Committee on a Universal Subarachnoid Haemorrhage Grading Scale. *J Neurosurg* 1988;**68**:985–6.
21. Teasdale GM, Jennett B. Assessment of coma and impaired consciouness. *Lancet* 1974;**ii**:81–4.
22. Craig JJ, Patterson VH, Cooke RS, Rocke LG, McKinstry CS. Diagnosis of subarachnoid haemorrhage (letter). *Lancet* 1997;**350**:216–7.
23. van Gijn J. Slip-ups in diagnosis of subarachnoid haemorrhage. *Lancet* 1997;**349**:1492.
24. Mayer PL, Awad IA, Todor R, Harbaugh K, Varnavas G, Lansen TA, *et al.* Misdiagnosis of symptomatic cerebral aneurysm. Prevalance and correlation with outcome at four institutions. *Stroke* 1996;**27**:1558–63.
25. Vermeulen M, Lindsay KW, Murray GD, Cheah F, Hijdra A, Muizelaar JP, *et al.* Antifibrinolytic treatment in subarachnoid haemorrhage. *N Engl J Med* 1984;**311**:432–7.
26. Ohman J, Heiskanen O. Timing of operation for ruptured supratentorial aneurysms: a prospective randomized study. *J Neurosurg* 1989;**70**:55–60.
27. Hunt WE, Hess RM. Surgical risk as related to time of intervention in the repair of intracranial aneurysms. *J Neurosurg* 1968;**28**:14–20.

28. ISAT. International Subarachnoid Haemorrhage Trial. http://users.ox.ac.uk/~isat/

29. Pickard JD, Murray GD, Illingworth R, Shaw MD, Teasdale GM, Foy PM, *et al.* Effect of oral nimodipine on cerebral infarction and outcome after subarachnoid haemorrhage: British aneurysm nimodipine trial. *BMJ* 1989;**298**:636–42.

30. Kassell NF, Torner JC, Haley EC Jr, Jane JA, Adams HP, Kongable GL, *et al.* The international co-operative study on the timing of aneurysm surgery. Part 1: overall management results. *J Neurosurg* 1990;**73**:18–36.

31. Melville A, Donovan J, Sheldon T, Peters T. Benign prostatic hyperplasia. *Qual Health Care* 1996;**5**:111–19.

32. Cockett ATK, Barry MJ, Holtgrewe HL, Sihelnick S, Williams R, McConnell J. Indications for treatment of benign prostatic hyperplasia: the American Urological Association Study. *Cancer* 1992;**70**:280–3.

33. Hunter DJW, McKee CM, Sanderson CFB, Black NA. Appropriate indications for prostatectomy in the UK – results of a consensus panel. *J Epidemiol Community Health* 1994;**48**:58–64.

34. Jenkins BJ, Sharma P, Badenoch DF, Fowler CG, Blandy JP. Ethics, logistics and a trial of transurethral versus open prostatectomy. *Br J Urol* 1992;**69**:372–4.

35. Roos NP, Wennberg JE, Malenka DJ, Fisher ES, McPherson K, Andersen TF, *et al.* Mortality and reoperation after open and transurethral resection of the prostate for benign prostatic hyperpplasia. *N Engl J Med* 1989;**320**:1120–4.

36. Holtgrewe HL. Current trends in management of men with lower urinary tract symptoms and benign prostatic hyperplasia. *Urology* 1998;**51**:1–7.

37. Hargreave TB, Heynes CF, Kendrick SW, Whyte B, Clarke JA. Mortality after transurethral and open prostatectomy in Scotland. *Br J Urol* 1996;**77**:547–53.

38. Greenfield S. The state of outcome research: are we on target? *N Engl J Med* 1989;**320**:1142–3.

39. Andersen TF, Bronnumhansen H, Sejr T, Roepstorff C. Elevated mortality following transurethral resection of the prostate for benign hypertrophy – but why? *Med Care* 1990;**28**:870–81.

40. Crowley AR, Horowitz M, Chan E, Macchia RJ. Transurethral resection of the prostate versus open prostatectomy: long-term mortality comparison. *J Urol* 1995;**153**:695–7.

41. McKee M, Coles J, James P. 'Failure to rescue' as a measure of quality of hospital care: the limitations of secondary diagnosis coding in English hospital data. *J Public Health Med* 1999;**21**:453–8.

42. Leyland AH, Boddy FA. League tables and acute myocardial infarction. *Lancet* 1998;**351**:555–8.

43. Favaloro RG. Saphenous vein autograft replacement of severe segmental coronary artery occlusion. Operative technique. *Ann Thorac Surg* 1968;**5**:334–9.

44. Killip T, Ravi K. Coronary bypass surgery. In: Pitt B, Julian D, Pocock S, editors. Clinical trials in cardiology. London: WB Saunders; 1997.

45. Black N, Langham S, Coshall C, Parker J. Impact of the 1991 NHS reforms on the availability and use of coronary revascularisation in the UK (1987–1995). *Heart* 1996;**76**(Suppl 4):1–30.

46. Grüntzig AR, Senning A, Siegenhaler WE. Nonoperative dilatation of coronary-artery stenosis. *N Engl J Med* 1979;**301**:61–8.

47. EPISTENT Investigators. Randomised placebo-controlled and balloon angioplasty-controlled trial to assess safety of coronary stenting with use of platelet glycoprotein-IIb/IIIa blockade. *Lancet* 1998;**352**:87–92.

48. Knudtson ML. Commentary. *Evidence-Based Med* 1996;**1**:83–4.

49. Pocock SJ, Henderson RA, Rickards AF, Hampton JR, King SB III, Hamm CW, *et al.* Meta-analysis of randomised trials comparing coronary angioplasty with bypass surgery. *Lancet* 1995;**346**:1184–9.

50. BARI Investigators. Comparison of coronary bypass surgery with angioplasty in patients with multivessel disease. *N Engl J Med* 1996;**335**:217–25.

51. White HD. Angioplasty versus bypass surgery. *Lancet* 1995;**346**:1174–5.

52. CABRI Trial Participants. First year results of CABRI (Coronary Angioplasty versus Bypass Revascularisation Investigation). *Lancet* 1995;**346**:1179–83.

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

54. King SB III, Lembo NJ, Weintraub WS, Kosinski AS, Barnhart HX, Kutner MH, *et al.* A randomised trial comparison of coronary angioplasty with coronary bypass surgery. *N Engl J Med* 1994;**331**:1044–50.

55. Spiegelhalter DJ, Myles JM, Jones DR, Abrams KR. An introduction to Bayesian methods in health technology assessment. *BMJ* 1999;**319**:508–12.

56. Spiegelhalter DJ, Freedman LS, Parmar MKB. Bayesian approaches to randomized trials. *J Roy Statistical Soc, Series A* 1994;**157**:357–87.

57. Smith TC, Abrams KR, Jones DR. Hierarchical models in generalised synthesis of evidence: an example based on studies of breast cancer screening. In: Sixteenth International Society for Clinical Biostatistics Conference: Technical Report. University of Leicester; 1995.

58. Thomas A, Spiegelhalter DJ, Gilks WR. BUGS: a program to perform Bayesian inference using Gibbs sampling. In Bernardo J, Berger J, Dawid A, Smith A, editors. Bayesian Statistics 4. Oxford: Oxford University Press; 1992.

59. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. Systematic reviews of trials and other studies. *Health Technol Assess* 1998;**2**(19).

60. Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Choosing between randomised and non-randomised studies: a systematic review. *Health Technol Assess* 1998;**2**(13).

61. Blumberg, MS. Risk adjusting health care outcomes: a methodologic review. *Med Care Rev* 1986;**26**:1129–48.

62. DesHarnais SI, McMahon LF Jr, Wroblewski RT, Hogan AJ. Measuring hospital performance. The development and validation of risk-adjusted indexes of mortality, readmissions, and complications. *Med Care* 1990;**28**:1127–41.

63. Rosen AK, Ash AS, McNiff KJ, Moskowitz MA. The importance of severity of illness adjustment in predicting adverse outcomes in the Medicare population. *J Clin Epidemiol* 1995;**48**:631–43.

64. Petryshen P, Pallas LL, Shamian J. Outcomes monitoring: adjusting for risk factors, severity of illness, and complexity of care. *J Am Med Inform Assoc* 1995;**2**:243–9.

65. Iezzoni LI, Ash AS, Shwartz M, Daley J, Hughes JS, Mackiernan YD. Judging hospitals by severity-adjusted mortality rates: the influence of the severity-adjustment method. *Am J Public Health* 1996;**86**:1379-87.

66. Silber JH, Rosenbaum PR. A spurious correlation between hospital mortality and complication rates: the importance of severity adjustment. *Med Care* 1997;**35**(Suppl):OS77–92.

67. Orr RK, Maini BS, Sottile FD, Dumas EM, OMara P. A comparison of four severity-adjusted models to predict mortality after coronary artery bypass graft surgery. *Arch Surg* 1995;**130**:301–6.

68. Landon B, Iezzoni LI, Ash AS, Shwartz M, Daley J, Hughes JS, Mackiernan YD. Judging hospitals by severity-adjusted mortality rates: the case of CABG surgery. *Inquiry* 1996;**33**:155–66.

69. Black N. High quality clinical databases: breaking down the barriers. *Lancet* 1999;**353**:1205–6.

70. Scottish Office. Clinical Outcome Indicators 1994. Edinburgh: Clinical Resource and Audit Group; 1995.

Health Technology Assessment panel membership

This report was identified as a priority by the Methodology Group.

Acute Sector Panel

Current members

Chair:
Professor Francis H Creed
University of Manchester

Professor Clifford Bailey
University of Leeds

Ms Tracy Bury
Chartered Society of Physiotherapy

Professor Collette Clifford
University of Birmingham

Dr Katherine Darton
M.I.N.D.

Mr John Dunning
Papworth Hospital, Cambridge

Mr Jonathan Earnshaw
Gloucester Royal Hospital

Mr Leonard Fenwick
Freeman Group of Hospitals, Newcastle-upon-Tyne

Professor David Field
Leicester Royal Infirmary

Ms Grace Gibbs
West Middlesex University Hospital NHS Trust

Dr Neville Goodman
Southmead Hospital Services Trust, Bristol

Professor Mark Haggard
MRC Institute of Hearing Research, University of Nottingham

Professor Robert Hawkins
University of Manchester

Dr Duncan Keeley
General Practitioner, Thame

Dr Rajan Madhok
East Riding Health Authority

Dr John Pounsford
Frenchay Hospital, Bristol

Dr Mark Sculpher
University of York

Dr Iqbal Sram
NHS Executive, North West Region

Mrs Joan Webster
Consumer member

Past members

Professor John Farndon*
University of Bristol

Professor Senga Bond
University of Newcastle-upon-Tyne

Professor Ian Cameron
Southeast Thames Regional Health Authority

Ms Lynne Clemence
Mid-Kent Health Care Trust

Professor Cam Donaldson
University of Aberdeen

Professor Richard Ellis
St James's University Hospital, Leeds

Mr Ian Hammond
Bedford & Shires Health & Care NHS Trust

Professor Adrian Harris
Churchill Hospital, Oxford

Dr Gwyneth Lewis
Department of Health

Mrs Wilma MacPherson
St Thomas's & Guy's Hospitals, London

Dr Chris McCall
General Practitioner, Dorset

Professor Alan McGregor
St Thomas's Hospital, London

Professor Jon Nicholl
University of Sheffield

Professor John Norman
University of Southampton

Professor Michael Sheppard
Queen Elizabeth Hospital, Birmingham

Professor Gordon Stirrat
St Michael's Hospital, Bristol

Dr William Tarnow-Mordi
University of Dundee

Professor Kenneth Taylor
Hammersmith Hospital, London

* Previous Chair

continued

continued

Diagnostics and Imaging Panel

Current members

Chair:
Professor Mike Smith
University of Leeds

Dr Philip J Ayres
Leeds Teaching Hospitals NHS Trust

Dr Paul Collinson
St George's Hospital, London

Dr Barry Cookson
Public Health Laboratory Service, Colindale

Professor David C Cumberland
University of Sheffield

Professor Adrian Dixon
University of Cambridge

Mr Steve Ebdon-Jackson
Department of Health

Mrs Maggie Fitchett
Association of Cytogeneticists, Oxford

Dr Peter Howlett
Portsmouth Hospitals NHS Trust

Professor Alistair McGuire
City University, London

Dr Andrew Moore
Editor, *Bandolier*

Dr Peter Moore
Science Writer, Ashtead

Professor Chris Price
London Hospital Medical School

Dr William Rosenberg
University of Southampton

Mr Tony Tester
South Bedfordshire Community Health Council

Dr Gillian Vivian
Royal Cornwall Hospitals Trust

Dr Greg Warner
General Practitioner, Hampshire

Past members

Professor Michael Maisey*
Guy's & St Thomas's Hospitals, London

Professor Andrew Adam
Guy's, King's & St Thomas's School of Medicine & Dentistry, London

Dr Pat Cooke
RDRD, Trent Regional Health Authority

Ms Julia Davison
St Bartholomew's Hospital, London

Professor MA Ferguson-Smith
University of Cambridge

Dr Mansel Haeney
University of Manchester

Professor Sean Hilton
St George's Hospital Medical School, London

Mr John Hutton
MEDTAP International Inc., London

Professor Donald Jeffries
St Bartholomew's Hospital, London

Dr Ian Reynolds
Nottingham Health Authority

Professor Colin Roberts
University of Wales College of Medicine

Miss Annette Sergeant
Chase Farm Hospital, Enfield

Professor John Stuart
University of Birmingham

Dr Ala Szczepura
University of Warwick

Mr Stephen Thornton
Cambridge & Huntingdon Health Commission

Dr Jo Walsworth-Bell
South Staffordshire Health Authority

Methodology Group

Current members

Chair:
Professor Martin Buxton
Health Economics Research Group, Brunel University

Professor Doug Altman
ICRF/NHS Centre for Statistics in Medicine, University of Oxford

Dr David Armstrong
Guy's, King's & St Thomas's School of Medicine & Dentistry, London

Professor Nicholas Black
London School of Hygiene & Tropical Medicine

Professor Ann Bowling
University College London Medical School

Dr Mike Clarke
UK Cochrane Centre, Oxford

Professor Paul Dieppe
MRC Health Services Research Collaboration, University of Bristol

Professor Mike Drummond
Centre for Health Economics, University of York

Dr Vikki Entwistle
University of Aberdeen

Professor Ewan Ferlie
Imperial College, London

Professor Ray Fitzpatrick
University of Oxford

Mrs Jenny Griffin
Department of Health

Professor Jeremy Grimshaw
University of Aberdeen

Dr Stephen Harrison
University of Leeds

Mr John Henderson
Department of Health

Professor Richard Lilford
R&D, West Midlands

Professor Theresa Marteau
Guy's, King's & St Thomas's School of Medicine & Dentistry, London

Dr Henry McQuay
University of Oxford

Dr Nick Payne
University of Sheffield

Professor Maggie Pearson
NHS Executive North West

Dr David Spiegelhalter
Institute of Public Health, Cambridge

Professor Joy Townsend
University of Hertfordshire

Ms Caroline Woodroffe
Standing Group on Consumers in NHS Research

Past members

Professor Anthony Culyer*
University of York

Professor Michael Baum
Royal Marsden Hospital

Dr Rory Collins
University of Oxford

Professor George Davey Smith
University of Bristol

Professor Stephen Frankel
University of Bristol

Mr Philip Hewitson
Leeds FHSA

Mr Nick Mays
King's Fund, London

Professor Ian Russell
University of York

Professor David Sackett
Centre for Evidence Based Medicine, Oxford

Dr Peter Sandercock
University of Edinburgh

Dr Maurice Slevin
St Bartholomew's Hospital, London

Professor Charles Warlow
Western General Hospital, Edinburgh

* Previous Chair

Pharmaceutical Panel

Current members

Chair:
Professor Tom Walley
University of Liverpool

Dr Felicity Gabbay
Transcrip Ltd

Dr Peter Golightly
Drug Information Services,
NHS Executive Trent

Dr Alastair Gray
Health Economics
Research Centre,
University of Oxford

Professor Rod Griffiths
NHS Executive
West Midlands

Mrs Jeanette Howe
Department of Health

Professor Trevor Jones
ABPI, London

Ms Sally Knight
Lister Hospital, Stevenage

Dr Andrew Mortimore
Southampton & SW Hants
Health Authority

Mr Nigel Offen
NHS Executive Eastern

Dr John Reynolds
The Oxford Radcliffe Hospital

Mrs Marianne Rigge
The College of Health,
London

Mr Simon Robbins
Camden & Islington
Health Authority, London

Dr Frances Rotblat
Medicines Control Agency

Dr Eamonn Sheridan
St James's University Hospital,
Leeds

Mrs Katrina Simister
National Prescribing Centre,
Liverpool

Dr Ross Taylor
University of Aberdeen

Past members

Professor Michael Rawlins*
University of Newcastle-
upon-Tyne

Dr Colin Bradley
University of Birmingham

Professor Alasdair
Breckenridge
RDRD, Northwest Regional
Health Authority

Ms Christine Clark
Hope Hospital, Salford

Mrs Julie Dent
Ealing, Hammersmith &
Hounslow Health Authority,
London

Mr Barrie Dowdeswell
Royal Victoria Infirmary,
Newcastle-upon-Tyne

Dr Tim Elliott
Department of Health

Dr Desmond Fitzgerald
Mere, Bucklow Hill, Cheshire

Professor Keith Gull
University of Manchester

Dr Keith Jones
Medicines Control Agency

Dr John Posnett
University of York

Dr Tim van Zwanenberg
Northern Regional
Health Authority

Dr Kent Woods
RDRD, Trent RO,
Sheffield

Population Screening Panel

Current members

Chair:
Professor Sir John
Grimley Evans
Radcliffe Infirmary, Oxford

Mrs Stella Burnside
Altnagelvin Hospitals Trust,
Londonderry

Mr John Cairns
University of Aberdeen

Professor Howard Cuckle
University of Leeds

Dr Carol Dezateux
Institute of Child Health,
London

Mrs Anne Dixon-Brown
NHS Executive Eastern

Professor Dian Donnai
St Mary's Hospital,
Manchester

Dr Tom Fahey
University of Bristol

Mrs Gillian Fletcher
National Childbirth Trust

Dr JA Muir Gray
National Screening
Committee, NHS Executive
Oxford

Professor Alexander Markham
St James's University Hospital,
Leeds

Dr Ann McPherson
General Practitioner,
Oxford

Dr Susan Moss
Institute of Cancer Research

Mr John Nettleton
Consumer member

Mrs Julietta Patnick
NHS Cervical
Screening Programme,
Sheffield

Dr Sarah Stewart-Brown
Health Service Research Unit,
University of Oxford

Past members

Dr Sheila Adam*
Department of Health

Professor George Freeman
Charing Cross & Westminster
Medical School, London

Dr Mike Gill
Brent & Harrow
Health Authority

Dr Anne Ludbrook
University of Aberdeen

Professor Theresa Marteau
Guy's, King's &
St Thomas's School of
Medicine & Dentistry,
London

Professor Catherine Peckham
Institute of Child Health,
London

Dr Connie Smith
Parkside NHS Trust,
London

Ms Polly Toynbee
Journalist

Professor Nick Wald
University of London

Professor Ciaran Woodman
Centre for
Cancer Epidemiology,
Manchester

* Previous Chair

continued

continued

Primary and Community Care Panel

Current members

Chair:
Dr John Tripp
Royal Devon & Exeter Healthcare NHS Trust

Mr Kevin Barton
East London & City Health Authority

Professor John Bond
University of Newcastle-upon-Tyne

Dr John Brazier
University of Sheffield

Ms Judith Brodie
Cancer BACUP

Mr Shaun Brogan
Ridgeway Primary Care Group, Aylesbury

Mr Joe Corkill
National Association for Patient Participation

Dr Nicky Cullum
University of York

Professor Pam Enderby
University of Sheffield

Dr Andrew Farmer
Institute of Health Sciences, Oxford

Dr Jim Ford
Department of Health

Professor Richard Hobbs
University of Birmingham

Professor Allen Hutchinson
University of Sheffield

Dr Aidan MacFarlane
Independent Consultant

Professor David Mant
Institute of Health Sciences, Oxford

Dr Chris McCall
General Practitioner, Dorset

Dr Robert Peveler
University of Southampton

Professor Jennie Popay
University of Salford

Dr Ken Stein
North & East Devon Health Authority

Past members

Professor Angela Coulter*
King's Fund, London

Professor Martin Roland*
University of Manchester

Dr Simon Allison
University of Nottingham

Professor Shah Ebrahim
Royal Free Hospital, London

Ms Cathy Gritzner
King's Fund, London

Professor Andrew Haines
RDRD, North Thames Regional Health Authority

Dr Nicholas Hicks
Oxfordshire Health Authority

Mr Edward Jones
Rochdale FHSA

Professor Roger Jones
Guy's, King's & St Thomas's School of Medicine & Dentistry, London

Mr Lionel Joyce
Chief Executive, Newcastle City Health NHS Trust

Professor Martin Knapp
London School of Economics & Political Science

Dr Phillip Leech
Department of Health

Professor Karen Luker
University of Liverpool

Dr Fiona Moss
Thames Postgraduate Medical & Dental Education

Professor Dianne Newham
King's College London

Professor Gillian Parker
University of Leicester

Dr Mary Renfrew
University of Oxford

Ms Hilary Scott
Tower Hamlets Healthcare NHS Trust, London

National Coordinating Centre for Health Technology Assessment, Advisory Group

Current members

Chair:
Professor John Gabbay
Wessex Institute for Health Research & Development

Dr Sheila Adam
Department of Health

Professor Nicholas Black
London School of Hygiene and Tropical Medicine

Professor Martin Buxton
Health Economics Research Group, Brunel University

Mr Harry Cayton
Alzheimer's Disease Society

Professor Angela Coulter
The King's Fund, London

Professor Paul Dieppe
MRC Health Services Research Collaboration, University of Bristol

Professor Mike Drummond
Centre for Health Economics, University of York

Professor Shah Ebrahim
MRC Health Services Research Collaboration, University of Bristol

Ms Lynn Kerridge
Wessex Institute for Health Research & Development

Professor Jos Kleijnen
NHS Centre for Reviews and Dissemination, University of York

Dr Ruairidh Milne
Wessex Institute for Health Research & Development

Ms Kay Pattison
Research & Development Directorate, NHS Executive

Professor James Raftery
Health Economics Unit, University of Birmingham

Professor Ian Russell
Department of Health Sciences & Clinical Evaluation, University of York

Dr Ken Stein
North & East Devon Health Authority

Professor Andrew Stevens
Department of Public Health & Epidemiology, University of Birmingham

Professor Kent Woods
Department of Medicine & Therapeutics, University of Leicester

Past member

Dr Paul Roderick
Wessex Institute for Health Research & Development

HTA Commissioning Board

Current members

Chair:
Professor Shah Ebrahim
Professor of Epidemiology of Ageing, University of Bristol

Professor Doug Altman
Director, ICRF Medical Statistics Group, Centre for Statistics in Medicine, University of Oxford

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

Mr Peter Bower
General Manager and Independent Health Advisor, Thames Valley Primary Care Agency

Ms Christine Clark
Honorary Research Pharmacist, Hope Hospital, Salford

Professor Martin Eccles
Professor of Clinical Effectiveness, University of Newcastle-upon-Tyne

Dr Mike Gill
Regional Director of Public Health, NHS Executive South East

Dr Alastair Gray
Director, Health Economics Research Centre, University of Oxford

Professor Mark Haggard
Director, MRC Institute of Hearing Research, University of Nottingham

Dr Jenny Hewison
Senior Lecturer, Department of Psychology, University of Leeds

Professor Alison Kitson
Director, Royal College of Nursing Institute

Dr Donna Lamping
Senior Lecturer, Department of Public Health, London School of Hygiene & Tropical Medicine

Professor Alan Maynard
Joint Director, York Health Policy Group, University of York

Professor David Neal
Joint Director, York Health Policy Group, University of York

Professor Jon Nicholl
Director, Medical Care Research Unit, University of Sheffield

Professor Gillian Parker
Nuffield Professor of Community Care, University of Leicester

Dr Tim Peters
Reader in Medical Statistics, Department of Social Medicine, University of Bristol

Professor Martin Severs
Professor in Elderly Health Care, University of Portsmouth

Dr Sarah Stewart-Brown
Health Service Research Unit, University of Oxford

Professor Ala Szczepura
Director, Centre for Health Services Studies, University of Warwick

Dr Gillian Vivian
Consultant, Royal Cornwall Hospitals Trust

Professor Graham Watt
Department of General Practice, University of Glasgow

Professor Kent Woods
Professor of Therapeutics, University of Leicester

Dr Jeremy Wyatt
Senior Fellow, Health Knowledge Management Centre, University College London

Past members

Professor Ian Russell*
Department of Health Sciences & Clinical Evaluation, University of York

Professor Charles Florey*
Department of Epidemiology & Public Health, Ninewells Hospital & Medical School, University of Dundee

Professor David Cohen
Professor of Health Economics, University of Glamorgan

Mr Barrie Dowdeswell
Chief Executive, Royal Victoria Infirmary, Newcastle-upon-Tyne

Dr Michael Horlington
Head of Corporate Licensing, Smith & Nephew Group Research Centre

Professor Sir Miles Irving
Professor of Surgery, University of Manchester, Hope Hospital, Salford

Professor Martin Knapp
Director, Personal Social Services Research Unit, London School of Economics & Political Science

Professor Theresa Marteau
Director, Psychology & Genetics Research Group, Guy's, King's & St Thomas's School of Medicine & Dentistry, London

Professor Sally McIntyre
MRC Medical Sociology Unit, Glasgow

Professor David Sackett
Centre for Evidence Based Medicine, Oxford

Dr David Spiegelhalter
MRC Biostatistics Unit, Institute of Public Health, Cambridge

Professor David Williams
Department of Clinical Engineering, University of Liverpool

Dr Mark Williams
Public Health Physician, Bristol

* Previous Chair

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.

Copies of this report can be obtained from:

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