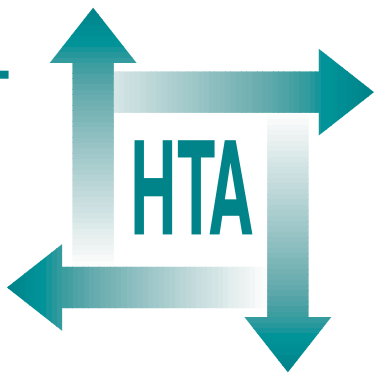
Health Technology Assessment 2000; Vol. 4: No. 8

Methodology

An introduction to statistical methods for health technology assessment

SJ White
D Ashby
PJ Brown

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.

An introduction to statistical methods for health technology assessment

SJ White[1]
D Ashby[2*]
PJ Brown[3]

[1] Department of Mathematical Sciences, University of Liverpool, Liverpool, UK
[2] Wolfson Institute of Preventive Medicine, Queen Mary and Westfield College, London, UK
[3] Institute of Mathematics and Statistics, University of Kent, Canterbury, UK

[*] Corresponding author

Competing interests: none declared

Published May 2000

This report should be referenced as follows:

White SJ, Ashby D, Brown PJ. An introduction to statistical methods for health technology assessment. *Health Technol Assess* 2000;**4**(8).

Health Technology Assessment is indexed in *Index Medicus*/MEDLINE and *Excerpta Medica*/EMBASE. Copies of the Executive Summaries are available from the NCCHTA web site (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 93/50/02.

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, Ruairidh Milne, Ken Stein and John Gabbay
Monograph Editorial Manager: Melanie Corris

The editors have tried to ensure the accuracy of this report but cannot accept responsibility for any errors or omissions. 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

ASRU	Applied Statistics Research Unit
CI	confidence interval
CPMP	Committee for Proprietary Medicinal Products
FDA	Food and Drug Administration
HTA	health technology assessment
ICH	International Conference on Harmonisation
MRC	Medical Research Council
PK–PD	Pharmacokinetic–pharmacodynamic
PSI	Statisticians in the Pharmaceutical Industry
QoL	quality of life
RCT	randomised controlled trial
SOP	standard operating procedure

Executive summary

Objectives

The ability of the Health Technology Assessment (HTA) programme to answer questions about the effectiveness and cost-effectiveness of new technologies relies on the availability of appropriate methodologies including statistics. The aims of this report were to:

- document recommended practice in relevant and related areas
- document current practice critically
- map current methodological research
- identify areas relevant to health technology assessment where statistical methodology is either inadequate or not being employed to full advantage, and
- identify suitable areas for further research.

Methods

To meet these objectives a series of linked reviews were undertaken. These were of:

- textbooks used in the training of medical statisticians on three MSc courses
- guidelines covering statistical aspects of evaluation of technologies
- publications appearing in the statistical literature during 1994–95
- publications on various study designs using MEDLINE for 1993–96
- publications on methods for analysis of follow-up studies using MEDLINE for 1993–96
- the needs of the HTA programme as evidenced by current work.

Findings

Statistical training

The review of textbooks from MSc courses showed that students are being offered courses in statistical theory and methods, design of experiments, linear models and generalised linear models, survival analysis, repeated measures, spatial statistics, multivariate methods, multilevel models, distribution-free statistics, Bayesian inference and methods, measurement errors, computational statistics, clinical trials and epidemiology. This represents a wider range than any one person can learn in a year. Much is relevant to health technology assessment, but the links are not yet very explicit, and there are no directly relevant textbooks recommended.

Statistical guidelines

Statistical guidelines have been developed in areas relevant to health technology assessment, in particular drug regulation, and in systematic reviews of randomised trials, through the Cochrane Collaboration. The linchpin technology in both of these areas is the randomised controlled trial. However, they mostly emphasise principles and ways of working rather than detail, with only meta-analysis covered in depth.

Publications

A review of the papers potentially relevant to health technology assessment, published in statistical journals in 1994–95 yielded 505 papers. These were predominantly about new methodology rather than discussion or review papers, mainly used classical rather than Bayesian approaches, and largely used re-analysed or simulated data, rather than primary analyses. Most related to preclinical or clinical trials rather than other kinds of studies.

Study designs

Much of the statistical literature on study designs that relate to health technology assessment comes from clinical trials; there are relatively few publications that cover the more complex experimental designs, meta-analysis or studies of drug safety. Within the medical literature, of the more complex experimental designs, (bio-)equivalence and crossover studies can be identified but not other designs in large numbers.

Methods for analysis

Of the statistical literature relevant to health technology assessment on analysis of follow-up studies, both survival and longitudinal data feature regularly, and repeated measures (non-longitudinal) occur less often. Within the medical literature, survival analysis is extremely common, particularly proportional hazards and Cox regression. Much of this work is in the context of cancer and heart disease. Identifying longitudinal studies

in the medical literature is straightforward, and they cover a range of conditions, but identifying the use of longitudinal methods of analysis is much harder.

Needs of the HTA programme

In health technology assessment the question 'Does the technology work?' is most easily answered using standard statistical methods. 'For whom?' raises statistical questions of subgroup analysis and interactions, and wider questions of generalisability. 'At what cost?' raises questions of identification and measurement of costs, with appropriate handling of associated uncertainty. 'How does it compare with alternatives?' brings a need for more formal decision analysis, and revisiting work on complex experimental design.

Recommendations

- The NHS R&D programme could consider training strategies for continuing professional development of statisticians, for example by allocating a fund to allow attendance of relevant courses.
- The NHS R&D programme could consider commissioning induction courses for statisticians working in health technology assessment. The purpose of such courses would partly be to give an introduction to health technology assessment and associated disciplines, and partly to focus on reinforcing statistical methods particularly pertinent to health technology assessment.
- Researchers in health technology assessment could avail themselves of existing guidelines, specifically those in drug regulation, and, if involved in meta-analysis, those of the Cochrane Collaboration.
- The NHS R&D programme could sponsor workshops to bring together statisticians and others who have been working in health technology assessment to identify and develop future statistical issues in health technology assessment.
- Case studies are needed on decision making under uncertainty using established Bayesian methodology to integrate health outcomes with wider costs.
- Established statistical methodology on design of experiments is potentially relevant to complex questions in health technology assessment. The development of specimen protocols explicitly using such methodology could be commissioned.

Chapter 1
Background

Health technology assessment

'Health technology' is an internationally recognised term that covers any method used by those working in the health services to promote health, prevent and treat disease and improve rehabilitation and long-term care. It includes the use of devices, equipment, drugs, procedures and care. The Health Technology Assessment (HTA) programme aims to answer questions of purchasers, providers and users of health services on effectiveness and cost-effectiveness of interventions. Questions include:

- Does this treatment work?
- For whom?
- At what cost?
- How does it compare with other treatments that are available?

The ability to answer these questions depends on the availability of appropriate methodologies, including statistics. In recognition of this, the HTA programme commissioned several systematic reviews on statistical methodology. This is the broadest review, and has the remit to look at current practice, and to review current statistical research to see which recent novel developments might be pertinent.

What is statistics?

The term 'statistics' means different things to different people. A helpful taxonomy was provided by David Bartholomew is his Presidential address to the Royal Statistical Society.[1] Type I statistics has as its main thrust "the collection and presentation of numerical data in a manner calculated to reveal their numerical features". This often refers to populations and, although it may seem 'practical' rather than 'theoretical', Bartholomew argues that the form of the presentation and interpretation is linked to our conceptual models. An example from the health field might be cancer registration statistics, which are routinely presented by age and sex, reflecting our understanding of basic determinants of cancer risk.

Type II statistics is concerned with making inferences from samples to large populations, and with surveys and experimental design. In health technology assessment the design and analysis of clinical trials stems straight from Type II statistics.

Type III statistics is concerned with large and complex systems. Modelling is an essential element. Understanding inequalities in health would fall into this category, as would epidemiology of most chronic diseases or assessing the impact of a national screening programme.

Type IV statistics involves a wider stage where policy decisions are taken under uncertainty, and where Bartholomew argues statisticians could play a greater role. Evidence-based medicine is an example of this broader remit.

Statistical research can cover a wide spectrum, from that which is a branch of mathematics, such as the study of probability theory, through to a broad range of very applied work in fields such as agriculture, chemometrics, biology, industrial processes and, of course, health. One of the fascinations of statistics is the way developments in one area lend insight in a totally different area.

The wide, though inter-related meanings of 'statistics', as well as its span from the mathematical to the applied are what give it its power as a discipline, but they also present very serious challenges in undertaking a comprehensive review of statistical methodology in health technology assessment.

In this review we will attempt to describe current practice, illustrate some current areas of development of statistics relevant to health technology assessment, and to highlight some areas that are less well developed. However, the reader would do well to take this review as an introduction to the subject, rather than the final work. Other projects under the HTA programme that develop particular areas further include *Methods for the analysis of quality-of-life and survival data in health technology assessment.*[2]

Perspective

The team of researchers all have their backgrounds in statistics, one in mathematical statistics and two in medical statistics. To some extent the review reflects this: it asks what is happening in the training of and research from statisticians, and then assesses how this is relevant to health technology assessment. A complementary approach would be to start from assessing work in health technology assessment, and to assess the statistical challenges and whether adequate methodology exists.

Research questions

- **To document recommended practice in relevant and related areas**. Textbooks that are recommended for MSc in medical statistics and guidelines covering statistical aspects of the evaluation of technologies were reviewed.
- **To document current practice critically**. This was too broad a subject to cover fully but was partially addressed by the electronic searching.
- **To identify novel methodological developments**. A review of research appearing in statistical journals in 1994 and 1995 was undertaken. Articles were extensively cross-classified, and the resulting database used for two reviews, one of work in study design, and one on work in analysis of studies that follow patients over time.
- **To provide examples of methodological applications**. Electronic searching of MEDLINE was used to document the use of particular statistical techniques within the medical and health literature.
- **To identify areas where statistical methodology is either inadequate or is not being employed to full advantage, and to identify areas for further research**. The needs of the HTA programme have become more explicit since this review was first commissioned. In particular, they are demonstrated by the current range of commissioned work. Current areas of statistical research work with the current activities of the HTA programme have been compared in an attempt to identify those areas that are covered in neither the various guidelines, nor the current statistical research output.

Chapter 2

Current statistical practice: review of MSc courses and textbooks

Introduction

The main purpose of this project is to review new developments in statistical methodology relevant to health technology assessment. Before deciding what is 'new' it is useful to establish a baseline.

We shall review statistical textbooks that are recommended for Masters of Medical Statistics and similar degrees. Books broadly fall into three categories: the standard statistical textbooks, those aimed for general medical audiences, and those aimed at statistical methods in particular areas, for example epidemiology or clinical trials.

Aims

The aim of this review was to review courses and text books used in current MSc courses in medical statistics, in order to establish what might be seen as 'standard' material with which a newly qualified statistician should be familiar.

Methods

Institutions running the four MSc courses in the UK that offer medical statistics as a major option were contacted and course details were requested. Little additional benefit would have been gained from investigating those courses that were more generic, but included a medical statistics option. The recommendations from each course were then synthesised into a common framework, using the topic structure of the MSc courses. Courses or options on applications a long way removed from HTA, such as genetics and biological assay, were excluded from the review, as were generic research skills and computing. The books have essentially been classified on the basis of title. An introduction to each area is given below.

Results

Course guides and reading lists were obtained from the MSc in Medical Statistics at the London School of Hygiene and Tropical Medicine (1997–98), the MSc in Statistics with Applications in Medicine at the University of Southampton (1997–98) and the MSc in Biometry at Reading (1997–98).

Whole course

Each course was slightly different in emphasis, and some had a relatively wide range of options. Only one course recommended overall course books, advising to choose between Altman,[3] Armitage and Berry,[4] or Fisher and van Belle.[5]

Statistical theory and methods

Many, but not all student on these courses will have covered some statistical material at undergraduate level. All courses contained some topics on general statistical theory and methods, but recommended texts varied,[6–11] or were not recommended at all (*Table 1*).

Design of experiments

Design of experiments is a classic topic for more general statistical training. It covers principles including randomisation, replication and factorial designs. It has its roots in agriculture, and has wide application to industrial processes. The two-group randomised controlled trial (RCT) frequently employed in medical research is actually a very simple experimental design conceptually, though with special challenges, notably sequential accrual of subjects. More complex designs are employed within the pharmaceutical industry for preclinical research. As health technology assessment develops, it is likely to need more complex studies, and hence these methods to answer questions about packages of interventions and interactions between them. There are parallels between constraints in agricultural experiments (individual fields have particular orientations and for some purposes must be treated as a whole unit) and work in organisations, where some aspects of care, but not all, may need to be randomised at, say, primary care team level. Not all courses offer this topic; for those that do, the textbooks mainly reflect the non-medical heritage of the subject.[12–16] None recommend the classic text by Cochran and Cox.[17]

TABLE 1 *Recommended books for study*

Reference	Title
Statistical theory and methods	
Beaumont, 1980[6]	Intermediate mathematical statistics
Cox & Hinkley, 1974[7]	Theoretical statistics
Le, 1992[8]	Fundamentals of biostatistical inference
Mood *et al*., 1974[9]	Introduction to the theory of statistics
Larson, 1982[10]	Introduction to probability and statistical inference
Hettmansperger, 1984[11]	Statistical inference based on ranks
Design of experiments	
Mead, 1988[12]	The design of experiments: statistical principles for practical application
Atkinson & Donev, 1994[13]	Optimum experimental designs
Box *et al*., 1978[14]	Statistics for experiments
Jones & Kenward, 1989[15]	The design and analysis of cross-over trials
John & Williams, 1995[16]	Cyclic and computer generated designs
Linear models and generalised linear models	
Mead *et al*., 1993[18]	Statistical methods in agriculture and experimental biology
Montgomery & Peck, 1982[19]	Introduction to linear regression analysis
Aitken *et al*., 1989[20]	Statistical modelling in GLIM
Dobson, 1990[21]	An introduction to generalised linear models
Collett, 1991[22]	Modelling binary data
Draper & Smith, 1981[23]	Applied regression analysis
Kleinbaum, 1994[24]	Logistic regression: a self-learning text
Kleinbaum *et al*., 1988[25]	Applied regression analysis or other multivariate methods
Survival analysis	
Collett, 1994[26]	Modelling survival data in medical research
Cox & Oakes, 1984[27]	Analysis of survival data
Kalbfleisch & Prentice[28]	The statistical analysis of failure-time data
Repeated measures	
Hand & Crowder,1990[29]	Repeated measures analysis
Spatial statistics	
Cressie, 1991[30]	Statistics for spatial data
Diggle, 1983[31]	Statistical analysis of spatial point patterns
Ripley, 1981[32]	Spatial statistics
Webster & Oliver, 1990[33]	Statistical methods in soil and land resource survey
Multivariate methods	
Chatfield & Collins, 1980[34]	Introduction to multivariate analysis
Johnson & Wichern, 1992[35]	Applied multivariate analysis
Krzanowski, 1988[36]	Principles of multivariate analysis: a user's perspective
McCullagh & Nelder, 1989[37]	Generalized linear models
Manly, 1986[38]	Multivariate statistical methods – a primer
Mardia *et al*., 1979[39]	Multivariate analysis
Multilevel models	
Goldstein, 1995[40]	Multi-level statistical models
Bryk & Raudenbush, 1992[41]	Hierarchical linear models: applications and data analysis methods
Distribution-free statistics	
Siegel, 1988[42]	Non-parametric statistics for the behavioural sciences
Bayesian inference and methods	
Gelman *et al*., 1995[43]	Bayesian data analysis
Measurement error	
Fleiss, 1986[44]	The design and analysis of clinical experiments
Strike, 1991[45]	Statistical methods in laboratory medicine

continued

TABLE 1 contd *Recommended books for study*

Reference	Title
Computational statistics	
Thisted, 1988[46]	Elements of statistical computing
Silverman, 1986[47]	Density of estimation
Hastie & Tibshirani, 1990[48]	Generalised additive models
Efron & Tibshirani, 1993[49]	An introduction to the bootstrap
Clinical trials	
Pocock, 1983[50]	Clinical trials – a practical approach
Sequential trials	
Whitehead, 1997[51]	The design and analysis of sequential clinical trials
General epidemiology	
Hennekens & Buring, 1987[52]	Epidemiology in medicine
Ashton, 1994[53]	The epidemiological imagination
Lilienfeld, 1994[54]	Foundations of epidemiology
Stolley & Lasky, 1995[55]	Investigating disease patterns
Statistical methodologies for epidemiology	
Kahn & Sempos, 1989[56]	Statistical methods of epidemiology
Clayton & Mills, 1993[57]	Statistical models in epidemiology
Schlesselman, 1982[58]	Case-control studies: design, conduct, analysis
Breslow & Day, 1980[59]	Statistical methods of cancer research I: the analysis of case-control studies
Breslow & Day, 1987[60]	Statistical methods in cancer research II: the design and analysis of cohort studies

Linear models and generalised linear models

Linear models include simple and multiple regression models for continuous outcomes (e.g. blood pressure). These have been generalised for other outcomes such as binary endpoints (e.g. dead or alive). Generalisations include logistic regression, which is widely used in the analysis of observational and experimental studies. When a study's main result can be expressed as an odds ratio, logistic regression is the technique necessary to obtain odds ratios adjusted for confounding factors. All courses include modules, citing a range of text books (*Table 1*).[18–25]

Survival analysis

Survival analysis deals with the analysis of data in the form of 'time to an event' which may be death, a non-fatal event such as an epileptic seizure, or in a non-medical context, failure of a machine or component part. All courses have modules in survival analysis, with a small number of recommended texts (*Table 1*).[26–28]

Repeated measures

When there are repeated measures made, for example by different observers grading the same image, the usual assumptions of statistical independence break down, and models need to reflect the more complex structure. All courses mentioned repeated measures, but only one had a full unit on the topic, with one recommended book.[29]

Spatial statistics

When measures have a spatial relationship, this structure should be allowed for. An example is the geographic distribution of cancers. When monitoring changes over time, perhaps following the introduction of a new screening procedure, such spatial dependence may need to be taken into account. Only one course offered a module, with a choice of texts (*Table 1*).[30–33]

Multivariate methods

Multivariate analysis is a traditional statistical area dealing with multiple observations on each individual. Within the health context they are probably most widely applied in psychology, because of the widespread use of rating scales, but are potentially important in health technology assessment in areas such as quality of life (QoL). Two courses offer modules, recommending a variety of books, none specifically targeted towards health (*Table 1*).[34–39]

Multilevel models

Multilevel modelling is a rather newer area of statistical research, used for example when data are structured (e.g. patients within primary care units within geographical/administrative

localities, or patients treated by surgeons within hospitals). It can also deal with repeated measures, for example within the patients. Only one course offered a module, with a choice of two books.[40,41]

Distribution-free statistics

Distribution-free methods are fairly widely used in medicine, but only one course explicitly offers a module, using a classic text (*Table 1*).[42]

Bayesian inference and methods

Bayesian methods formalise the process of quantifying existing evidence and combining these with observed data to update the evidence. The principles have been well known to statisticians for decades, but computational advances are now making them available for a range of applications including health technology. Two courses offer options, but only one textbook is recommended.[43]

Measurement error

Measurement error problems abound in applied research, but only one course offers an option. The 'textbooks' are two individual chapters within books, one of which is specifically for laboratory medicine.[44,45]

Computational statistics

Many advances in statistics are due to increasing computational power to deal with complexity. Two courses offer an option in this explicitly, with mainstream statistical texts.[46–49]

Clinical trials

The RCT is central to the work of many medical statisticians. All courses offer at least one module, with recommended books on clinical trials, and on sequential trials.[50,51] Although courses cover meta-analyses, there are no recommended texts for the area.

Epidemiology

Epidemiology is traditionally about causes of disease, but both the results and the methodologies are relevant to health technology assessment. Books recommended cover both general epidemiological work, and those covering statistical methodologies for epidemiology.[52–60]

Summary

The courses tend to be geared towards traditional epidemiology, and the pharmaceutical industry. The major emphases are on individual studies, rather than designing or assessing packages of evidence, which are relevant in both pharmaceutical contexts for drug regulation, and for health technology assessment. It is hard to comment on areas not covered, especially as some courses had the potential to pick modules from a wide area to supplement the core/recommended ones. There is nothing explicit on pharmacoepidemiology, nor on economic assessments, though such issues may be addressed via examples within courses.

Discussion

Training and qualifications necessary to practice medical statistics

For a medical statistician, there is no unique career path. For a mathematical statistician, the path might follow the traditional academic pattern of a BSc (say, in mathematics and statistics), possibly an MSc, a PhD, a research assistant post, then an established post within a university department. Whilst some medical statisticians follow this path, many more do an MSc in medical statistics, but no further formal study. Some go straight to the pharmaceutical industry at this stage, others go into other forms of medical research, generally on short-term contracts, typically 1–3 years. Some stay on this type of contract for many years. Those who wish to move to permanent posts may do so within universities, government or the pharmaceutical industry.

An MSc course in medical statistics typically covers 9 months' coursework plus a dissertation. For many the dissertation will be the first taste of research. Those who do embark on a PhD will receive further training in aspects of research relevant to their research area. For others going into research posts, training is frequently 'on the job'. The quality of such training may depend heavily on the availability of experienced statisticians and other colleagues.

In addition to the academic qualifications already mentioned, statisticians may apply for the professional qualification of Chartered Statistician (CStat) awarded by the Royal Statistical Society. This reflects both academic qualification, and experience at a responsible level.

The survey above is a concatenation of three courses. No one course, let alone individual student, could cover all of the topics in a year.

However, the variety does illustrate the range of skills required of medical statisticians, and much of it is directly or potentially relevant to health technology assessment. However, the link is not explicit, and there are no directly relevant text books available.

One issue for the HTA programme is how to allow statisticians to expand or update their skills. This comes under the area of continuing professional development. This can be quite generic, but there is also a specific challenge of bridging the gap between current medical statistical training, and a health technology assessment perspective.

Recommendations

- The NHS R&D programme could consider training strategies for continuing professional development of statisticians, for example by allocating a fund to allow attendance of relevant courses.
- The NHS R&D programme could consider commissioning induction courses for statisticians working in health technology assessment. The purpose of such courses would partly be to give an introduction to health technology assessment and associated disciplines, and partly to focus on reinforcing statistical methods particularly pertinent to health technology assessment.

Chapter 3

Current practice: review of guidelines

Introduction

Having reviewed what medical statisticians learn during training, the next step in reviewing current practice is to review advice given to practitioners. As there are various guidelines that have been developed, this seems a good place to start.

Aims

The aims of this review were to establish which guidelines exist for medical researchers, to describe their content, and to make a judgement on the extent to which they are helpful to health technology assessment researchers.

Methods

For identification of guidelines, organisations known to be involved with medical research were contacted. Guidelines fell into natural groupings, and for each the background, context, content, and structure are described before reviewing the underlying statistical philosophies.

Regulatory guidelines

Although there is far more to health technology than pharmaceutical products, they nevertheless have a very important part to play. For historical reasons their use has been more heavily regulated than any other part of healthcare delivery, with formal procedures to be carried out in order to gain a product licence for a new medicinal product. As a result there has been more explicit discussion of acceptable practice, and it is here that most of the written guidelines can be found.

Until fairly recently the regulation of medicines has been done on a country by country basis, with moves towards harmonisation at the European level being a fairly recent development, which has been followed by moves toward global harmonisation. The development of guidelines on various issues, including statistics, reflect these changes.

American Food and Drug Administration Guidelines

The Food and Drug Administration (FDA) is a body within the US Department of Health and Human Services responsible, among other things, for the regulation of new medicinal products. These guidelines[61] describe in detail how the clinical and statistical sections of a new drug application should be written, what information should be included and how information and data should be presented and arranged. They do not give much detailed statistical advice but imply the importance of statistical quality and integrity by stating the need for including specific information.

The aim of these guidelines are to produce reports of an appropriate standard, which contain all the information needed by the regulatory authorities to make a judgement on whether a drug licence should be granted. These guidelines are very detailed and refer to other documents that are available from the FDA. The guidelines relate to the regulatory procedure in the USA, but because of the size of the US market, they have a large influence internationally.

The European Guidelines

The Committee for Proprietary Medicinal Products (CPMP) European Guidelines[62] were drawn up over a period of 3 years by statisticians from the regulatory environment as well as those from industry and academia from all over Europe.

The European guidelines are concerned mainly with the planning and methodology of the clinical trials that are to be carried out to collect scientific evidence about the efficacy and safety of a new drug product. The guidelines provides information on the general development of a product, clinical trial designs and issues of design, primary and secondary variables, the case report form, pre-specified data analysis, the conduct and monitoring phase, data and computer issues, and summaries of the clinical database.

As stated in its introduction, the purpose of the guidelines is not intended to be a textbook or to specify methodology. The researcher is still responsible for making reasoned choices between statistical approaches and procedures to solve

specific problems. It is recognised that innovation may be necessary in the design and analysis of a clinical trial. The guidelines are written very much from a classical statistical perspective.

The guidelines give warnings of potential problems, such as a carry-over effect in a crossover trial. They then offer advice on how to avoid such a problem, recommending, for example for carry-over, that sufficient knowledge of the disease area under study and the new medication are necessary to guide the decision on the length of a washout period needed.

Throughout the guidelines there is emphasis on the completeness of the protocol and other documentation. Any decisions made as to the planning of the trial or the methodology to be used must be documented comprehensively. There must be a contingency plan set out prospectively for every possible situation that may arise, for example the breaking of blindness or the treatment of missing values and outliers.

International Conference on Harmonisation Guidelines

The International Conference on Harmonisation (ICH) is a tri-partite initiative to standardise requirements for marketing applications submitted in Europe, Japan and the USA. There is a series of guidelines that have been developed or are currently under development. Many touch on statistical issues, including:

E1: The extent of population exposure to assess clinical safety
E2a: Clinical safety data management: definitions and standards for expediting reporting
E2b: Clinical safety data management: data elements for transmission of individual case safety reports
E2c: Clinical safety data management: periodic safety update reports for marketed drugs
E3: Structure and content of clinical study reports
E4: Dose response information to support drug registration
E5: Ethnic factors in the acceptability of foreign clinical data
E6: Good clinical practice: consolidated guidelines
E7: Studies in support of special populations: geriatrics
E8: General considerations for clinical trials
E9: Statistical principles for clinical trials
E10: Choice of control groups in clinical trials
M1: Standardisation of medical terminology for regulatory purposes
M3: Non-clinical safety studies for the conduct of human clinical trials for pharmaceuticals.

ICH-E3: Structure and content of clinical study reports

This document of guidelines[63] is similar to the FDA guidelines document, as it is not directly concerned with the biostatistical methodology that was used in a study, but instead in how the clinical study report, to be submitted to the regulatory authorities, is written in terms of its structure and content. The clinical study report referred to is the report that is submitted to a regulatory body with the objective of a medicinal products licence being granted.

Following these guidelines should result in a clinical study report that is acceptable to all the regulatory authorities of the ICH regions, namely Europe, the USA and Japan. In addition to the clinical study report, individual regulatory authorities can issue modules that specify extra information to be included in the appendices.

What these guidelines aim to do is to produce clinical study reports that integrate "the clinical and statistical description, presentation and analysis" and not simply join the clinical and statistical reports together. Although guidelines are not given on what statistical methodology should be used, as this is not the purpose of this document, there is detailed information on what statistical information should be documented and presented. It also explains why it is important for specific information should be included. For example, in the 'Multiple Comparisons/Multiplicity' section, it is mentioned that the number of false-positive findings increases as the number of hypothesis tests carried out on a dataset increases. So if multiple hypothesis tests are necessary, perhaps because of subgroup analysis, then any adjustment made to the p-value should be stated, or if no adjustment was made then it should be stated why this was felt appropriate action.

These guidelines are structured in the form of a report, and for each section and subsection details are given as to what information should be included. It also describes what should be included in the appendices and how they should be presented. Guidance is also given on ways to lay out tables and listings, and there are examples of how study design and assessment procedures could be represented.

ICH-E9: Statistical principles for clinical trials

These guidelines were under development at the time of this review, but are a development of the CPMP guidelines described above.

The Safety Assessment of Marketed Medicines guidelines

The guidelines described so far refer to clinical trials in Phases I–III, that is before a drug is granted a licence and is marketed. The Safety Assessment of Marketed Medicines guidelines have been drawn up to assist those researchers who are assessing the safety of marketed medicines under company sponsorship.

These guidelines were drawn up in response to a paper published in the *British Medical Journal* in June 1992[64] which reported the general weaknesses of company-sponsored studies of safety of marketed medicines. Original guidelines published in 1988[65] were thought ineffectual and insufficient so they were reviewed and their scope was expanded. The current guidelines[66] are relevant to all studies that are sponsored by a pharmaceutical company and that assess the safety of newly marketed medicines. They were drawn up by a Working Party made up of representatives from the Medicines Control Agency, Committee on Safety of Medicines, the Royal College of General Practitioners, the British Medical Association and the Association of The British Pharmaceutical Industry.

These guidelines are not detailed but outline the type of study that should follow the recommendations, possible study designs that can be used, the conduct of studies, the required liaison with regulatory authorities, ethical considerations, complaints procedures, the participation of doctors and company promotion issues. No direct statistical advice is given. The possible study designs are described and advice given as to circumstances in which each should be used. As the study is assessing the safety of a new drug in the general population, this should be reflected in the selection of patients. It is essential that a patient is prescribed the drug under study solely on clinical grounds, rather than just to increase the enrolment of patients to the study. This is emphasised for all of the prospective study designs.

When describing the liaison with regulatory authorities that a study must maintain, it is stated that a study plan must be submitted to the Medicines Control Agency a month before the commencement of a trial. This study plan must include the aims and objectives of the study, methods to be used, the record keeping to be maintained and details of the statistical analysis to be carried out on the collected data.

Statistical review

A review of the above guidelines reveals a growing recognition of statistical principles and inferential framework. The FDA guidelines, published in 1988, make no explicit statement of inferential framework. From mentions of the null hypothesis, alternative hypotheses, power, significance level, *t*-test, and 95% confidence intervals (CIs), one can deduce that a classical hypothesis testing approach has been assumed, but the emphasis on technique rather than principles is highlighted by the reference to 'SAS Type III sums-of squares'. In contrast, the CPMP guidelines[62] explicitly acknowledge that "Although this *Note for Guidance* is written largely from the classical (frequentist) viewpoint, the use of Bayesian or other well-argued approaches is quite acceptable." Two years later, the ICH-E3 guidelines still made no explicit statement on inferential mode, but a detailed Annex on statistical considerations follows an implicitly frequentist line again, mentioning statistical model, null hypothesis, alternative hypotheses, CIs, prespecification of analysis plans, significance levels, *t*-tests, and analysis of variance. The ICH-E9 guidelines currently under development emphasise principles as being more important than technique, with minimisation of bias, maximisation of precision and robustness being the most important. Again, inferential mode is explicitly acknowledged with the statement: "This guideline largely refers to the use of frequentist methods when discussing hypothesis testing and/or confidence intervals. However, the use of Bayesian or other approaches may be considered when the reasons for their use are clear and when the resulting conclusions are sufficiently robust compared to the alternative assumption". Both Bayesian approaches and frequentist methods are explicitly defined in an accompanying glossary to that report.

Although most of the principles in these guidelines could apply to many study types, much of the detail applies to Phase III studies. Issues particularly pertinent to these trials, for example subgroup analysis, multicentre trials, and interim analysis, are discussed. There is very little on modelling strategies for more complex data, such as survival or longitudinal data, though the need for explicit analysis plans, to avoid accusations of data-dredging, is emphasised.

Pharmaceutical industry guidelines

Whether regulatory guidelines focus on detail or principle, those in the pharmaceutical industry have the responsibility of making sure their work will satisfy such guidelines. This has led to the development of standard operating procedures (SOPs). Statisticians in the Pharmaceutical

Industry (PSI) have developed generic guidelines aimed at satisfying statistical guidelines, and individual pharmaceutical companies have also laid out operating procedures that their employees must adhere to when involved in a clinical trial. Many companies have developed their own guidelines. In this report, we have not attempted to be comprehensive, but have reviewed one set from a pharmaceutical company and one from a contract research organisation.

The PSI guideline SOPs

The PSI guideline SOPs[67] were drawn up in order for individual institutions to adapt them into working SOPs for their specific needs. It was recognised that it would not be possible to draw up SOPs that were appropriate for all PSI members as they come from various areas of clinical research, for example pharmaceutical companies, contract research organisations and academic institutions.

These SOPs were issued in order to promote and publicise good statistical practice in all aspects of clinical trials. Working SOPs prepared using these guidelines should ensure compliance with the requirements of international Good Clinical Practice.

The Guideline SOPs are divided into 11 sections and cover all stages of a clinical trial and issues that arise during the research. Each section, for example clinical trial protocols, has a statement of objective and then procedures that need to be followed in order to reach the objective. Each procedure outlines the role of the statistician and his/hers responsibility in each specific task.

For the personnel involved in the clinical trial it is of paramount importance that the integrity of the trial, and therefore the results, are maintained in order for the regulatory bodies to grant a product licence if the drug under study reaches this ultimate stage. The SOPs therefore stress heavily the importance of documentation. The European guidelines, described above, talk about extensive documentation but not to the degree as is required by the pharmaceutical SOPs. In addition to a clinical development plan and the clinical trial protocol PSI describe all that should be included in a statistical analysis plan and a data management plan.

AstraZeneca statistical guidelines

AstraZeneca Pharmaceuticals has developed statistical guidelines for its employees to be directed by in the phases of drug development. They address statistical methodology in the areas of clinical trial design, analysis and the presentation of data and analytical results.

The guidelines were developed within AstraZeneca by senior statisticians and other senior staff to whom they apply (data management and computer programming staff). Each guideline has an 'owner', a senior member of the statistical staff in the company, and in the development of a guideline all practising statisticians are invited to contribute in the review and consultation process. Where appropriate other skill groups are also consulted.

These guidelines were drawn up prior to the development of PSIs SOPs and so do not follow their structure. However, AstraZeneca are familiar with their content and have checked that their own guidelines cover all the areas outlined by PSI and continue to do this as updates of the PSI SOPs are published. The AstraZeneca guidelines also meet the requirements of drug regulatory bodies. By doing this they gain maximum consistency and efficiency and it also aids them in the global aspect of their business. Biometric units across the world are able to produce standardised presentation of results by following these guidelines.

The guidelines are split into appropriate sections and cover all areas where statistical input is needed and so guidance is required. Some of the areas covered are statistical input to clinical trial protocols, sample size estimation and power calculations, randomisation, statistical analysis plans, the design of bioequivalence studies, the analysis of all types of data structures and computer programming standards and conventions.

Applied Statistics Research Unit SOPs

The Applied Statistics Research Unit (ASRU) is a contract research organisation, with the majority of its clients in the pharmaceutical industry. It is involved in all stages of drug development in a statistical capacity. Some of its employees work full time at a client's site of work and these employees would follow the guidelines set down by the client. However the remainder of its employees working on pharmaceutical contracts follow SOPs that ASRU have drawn up.

These SOPs are split into four main sections, general, data management, statistical analysis and computing. A senior member of staff involved in each area was responsible for developing the SOPs and a degree of coordination was required between the sections to ensure minimal overlap and maximum coverage. In the development of

SOPs, attention was paid to the FDA guidelines and the document on good clinical practice. The PSI guidelines were not used as these were published after ASRU drew up its SOPs. However, the SOPs have been reviewed with respect to the PSI document.

Statistical review

The PSI SOPs are comprehensive in one sense, in that they cover the process from trial planning to final report. However, they function more as a checklist of points to cover (e.g. methods for handling missing data) rather than giving details of how to go about this. By contrast, individual companies must grapple with detail, and we are grateful to AstraZeneca and to ASRU for discussions about this. AstraZeneca's guidelines are detailed and represent considerable investment by the company in a very competitive market. For this reason, they are not in the public domain, and cannot be reviewed in detail. For similar reasons, we will not give details of ASRU's procedures. Ironically, reviewing company material would probably give us the best overview of 'standard practice', but in the absence of this, we note the priority given to statistical procedures in speedy evaluation of new medicines.

Research community

The academic research community is much less structured than the pharmaceutical industry, and so, unsurprisingly, we have unearthed relatively little formal statistical material. However, the main gateways to successful research are funding bodies, ethics committees and publication in refereed journals, and it is these that we have reviewed.

As a means of ensuring good quality statistics in journal publications, various journals issue statistical checklists that are used by referees reviewing prospective publications. We did not attempt to be comprehensive, but selected the *British Medical Journal* as a leading journal, and included its guidelines and checklists.

Guidelines from the Medical Research Council

The Medical Research Council (MRC) is one of the major funders of medical research. It issue six booklets known as the MRC Ethics Series[68] that lay out guidelines for various issues in medical research, for example conduct of research on children or the mentally incapacitated and use of animals in medical research.

Only one of the six (*Principles in the Assessment and Conduct of Medical Research and Publicising Results*) mentions the importance of statistical aspects. It states: "In planning research, careful design, including statistical considerations where appropriate, is essential". However, there are no guidelines specifically concerned with statistical issues.

Ethics committees

The new briefing pack for Research Ethics Committee members has no explicit statement on statistics, but in the case studies, there are repeated references to the need for statistical review, for example on study size, and appropriateness of outcome measures. Although there are currently very few ethics committees with statisticians as members, the new Multicentre Research Ethics Committees must have a statistician as a member by constitution.[69]

Guidelines for journal publication

Guidelines published in the *British Medical Journal*[70] were drawn up by four statisticians and are concerned with assisting researchers who are carrying out research intended for publication. These guidelines have been approved by *Statistics in Medicine*,[71] and authors are referred to them by the *British Medical Journal* and other journals to use for their research projects.

In the introduction to the guidelines, the authors discuss how the need for such guidelines was recognised, because many studies were being published in medical journals that had had no input from a statistician in the design and analysis stage. This can lead to the incorrect or weak use of statistics by clinical researchers who have, at best, knowledge of just basic statistics. In an attempt to remedy this, the guidelines were drawn up. The guidelines can be used as a source for medical researchers to use to help them become aware of important statistical principles and also to aid them in writing their papers by outlining the type of statistical information that should be included in the paper.

The authors have not tried to lay out a set of rules of how and how not to carry out/analyse a study but instead give general information and guidance about what they consider to be the important aspects of statistical design, analysis and presentation. It is stressed in the introduction that however closely a researcher is able to follow these guidelines that there is no substitute for consulting with a statistician.

Statistical checklists for referees

To ensure good quality statistics in journal publications, various journals issue statistical checklists, which are used by referees reviewing prospective publications. Statistical checklists[70] have been drawn up by the *British Medical Journal* for various user groups. They can be used in different ways by editorial staff, statistical referees and also by authors, both in the design and analysis of a study and in its writing up. They aid systematic assessment of a paper on its statistical content, and a satisfactory completion of the checklist would mean that a paper was of a statistical quality satisfactory for publication.

There are two checklists, a general one and one for clinical trials, which also encompasses the general checklist.

Statistical review

Both the MRC and ethics committee documents have little detailed statistical content, referring to the role of statisticians instead. The journal guidelines are more detailed, but work as checklists for design and analysis rather than detailed prescriptions.

Guidelines for Cochrane reviewers

The Cochrane Collaboration aims to synthesise knowledge from RCTs, organised by specialty and subspecialty. It is an international collaboration. Standardised ways of working are evolving, and there is a *Handbook for Reviewers* which is continually being updated.[72]

Briefly for a particular review, a protocol is drawn up by a team of reviewers, detailing the objectives, methods for identifying and selecting studies, outcomes of interest and analysis methods. If suitable studies exist, meta-analysis is the techniques for combining estimates from studies.

Handbook for Cochrane reviewers

Statistical considerations on design aspects, such as formulating objectives are covered throughout the handbook, but in addition, there is a chapter on analysing and presenting results. This covers, in detail, the rationale for using meta-analysis, how to handle binary and continuous data, fixed versus random effects, how to display results, investigation of heterogeneity between studies, and subgroup analysis. Software, freely available to reviewers, in a package called REVMAN (Update Software), carries out these procedures. In addition, meta-regression, Bayesian meta-analysis, exact methods and meta-analysis of survival data are mentioned in the handbook, though currently the REVMAN software cannot handle these.

Statistical review

Of all the guidelines reviewed, this is by far the most detailed. The guidelines have evolved from the experiences of statisticians working within the collaboration, who form the Statistical Methods Working Group of the collaboration. They are designed to give concrete advice to a range of reviewers, most of whom will not be statisticians.

Initially, guidelines concentrate on standard methods, but future plans include issues driven by experience within the Cochrane context, such as meta-analysis of cluster-randomised trials, meta-analysis of crossover trials and publication bias.

Guidelines for health technology assessment

There are currently no guidelines aimed specifically at health technology assessment, though a recent *Health Technology Assessment* report[2] is an excellent guide to dealing with QoL and survival data. However, experiences in two areas may be pertinent – that of drug regulation and the Cochrane Collaboration. In each of these, statisticians and other researchers have reached a shared understanding of the context and objectives of their worth, and from this standards of good practice have evolved and are continuing to be developed.

Health technology assessment builds on both of these processes; much of the work done in the context of drug regulation and the Cochrane Collaboration is directly relevant, both in terms of substantive results or the efficacy and safety of products, and in terms of methodologies such as Phase III trials and meta-analysis. However, both of these initiatives have RCTs as the lynch pin, whereas health technology assessment needs to be able to answer questions beyond this, for example about use in practice, comparative performance and economics. The statistical methods used will therefore need to be more complex than either drug regulation or Cochrane work.

Perhaps the most important lesson for health technology assessment is to reflect on the

process by which consensus was reached in these two areas. Statisticians working in different institutions meet together regularly, through, for example, PSI meetings and Drug Information Association meetings for the pharmaceutical industry, and at a series of general and statistical meetings for the Cochrane Collaboration. From presentations and discussions, good practice evolves. To some extent, this process is beginning to happen within health technology assessment, but a focused series of workshops could act as a stimulus.

Recommendations

- Researchers in health technology assessment could avail themselves of existing guidelines, specifically those in drug regulation, and, if involved in meta-analysis, those of the Cochrane Collaboration.
- The NHS R&D programme could sponsor workshops to bring together statisticians and others who have been working in health technology assessment to identify and develop future statistical issues in health technology assessment.

Chapter 4

A database of statistical methodology used in health technology assessment

Introduction

Having reviewed statistical techniques taught in MSc courses, and recommendations to practitioners, the next stage is to review current statistical research relevant to health technology assessment. As other health technology assessment reviews have been commissioned to look at particular statistical areas, this project took a deliberately broad view.

Aims

The aims of this stage were to:

- obtain a broad classification of work likely to be relevant to health technology assessment
- provide a basis for further reviews of the application of these techniques in the medical literature.

Methods

Major statistical journals have been handsearched in order to identify those papers describing statistical methodology that can be applied to health technology assessment. These papers are then classified according to the type of paper they are, the statistical framework being used, the data source, type of study, data structure of the outcome variable, type of analysis being undertaken and any other discriminating features. The references for all of these papers and their classification have then been entered onto a bibliographic database called ProCite.[73] The classification scheme has been structured to allow accurate retrieval of papers investigating similar methodological areas.

Sources of references in database

The following scientific journals were handsearched for 1994 and 1995:

- *Statistics in Medicine*
- *Journal of the Royal Statistical Society* (Series A–D)
- *Statistical Methods in Medical Research*
- *Biometrics*
- *Biometrika*
- *Journal of the American Statistical Association*
- *Controlled Clinical Trials.*

Inclusion/exclusion criteria

A paper was included in the database if it is concerned with biostatistical methodology used in health technology assessment. For the purposes of this project, primarily because of time limitations, epidemiological work was excluded, as were those looking at genetics. An exception to the genetics papers are those investigating genetic susceptibility as these can lead to treatments or early detection of some diseases.

The classification of papers

Once a paper was identified as being relevant to health technology assessment then it was classified. A paper can be classified by a combination of the following terms, and can be described by more than one term per section. For example, the methodology being presented may be appropriate for both continuous and ordinal discrete outcome variables and so this can be recorded. This is often the case in review papers. Also, many survival analysis studies are of a longitudinal nature and when this has been recognised in the paper this is recorded by including both the terms 'survival data' and 'longitudinal data' in the keyword field.

The following is a reference from the database followed by its classification terms.

> Kenward MG, Lesaffre E, Molenberghs G. An application of maximum likelihood and generalised estimating equations to the analysis of ordinal data from a longitudinal study with cases missing at random. *Biometrics* 1994;**50**:945–53.

Methodology/Classical/Re-analysis/Clinical trial/ Observational Study/Estimation/Discrete-ordinal/ Drop-outs/Longitudinal data/Multi-centre/ Missing values.

This paper is developing **classical methodology**, using data that has been primarily presented elsewhere, which can be applied to both **multicentre clinical trials** and **observational studies**. The paper

will be concerned with **estimation using discrete-ordinal longitudinal data**, which is subject to **missing values**, some of which is due to **drop-outs**.

Developing the classification scheme

Due to the wide scope of the papers to be classified it was impossible to develop a classification system prior to handsearching the journals. It was anticipated that *Statistics in Medicine* would contain the widest range of relevant papers and so this journal was handsearched first in order to establish a classification scheme that would encompass the subject area. During the handsearching of this journal therefore each paper contributed to the development and complexity of the scheme. When the remaining journals were searched the classification scheme was found to accommodate all the papers found to be relevant to health technology assessment, each paper could be classified by the classification scheme as it stood after searching the 2 years of *Statistics in Medicine*. While the scheme was under development it became more complex and complete, so that the first papers to be classified would have been less well defined by their classification than those classified by the final version of the scheme. Before entry onto the computer database the papers in *Statistics in Medicine* were reclassified using the final version.

Entry of a paper onto the ProCite database

A paper was classified using the terms described above and then entered on the database. The bibliographic reference was first entered using the fields:

- author
- title
- journal
- volume
- pages.

In the keyword field a combination of the classification terms was then entered to classify the paper.

Results

The classification scheme

The final classification scheme is presented in a flow chart type representation in *Table 2*.

Definition of the major terms used in the classification scheme

Type of paper

- **Discussion**: a paper discussing an issue or a method, perhaps motivated by a new application.
- **Review**: a paper reviewing, comparing and contrasting various methods that can be used for the same situation.
- **Methodology**: a paper that introduces new methodology, most often in response to a new situation or set of clinical circumstances. This can also be an adaptation of established methodology or extension to it.

Statistical framework

- **Classical**: classical frequentist statistical theory used.
- **Bayesian**: Bayesian or empirical Bayes theory used.
- **Both**: some papers, particularly the review papers, use both classical and Bayesian theory.

Data source

- **Original**: data being presented for the first time, that is, it has not been reported in any other paper previously.
- **Simulated**: data simulated using distribution assumptions, random number generators or hypothetical data.
- **Re-analysis**: data that have been reported elsewhere previously.
- **Database**: data taken as a subset of a database, national register, for example, a cancer registry.

Study design

- **Preclinical**: trials carried out on animals, in petri dishes or on healthy volunteers.
- **Clinical trial**: trials carried out on patients to assess efficacy of treatments; patients are allocated a treatment by a non-clinical mechanism.
- **Screening**: studies that assess the benefit of screening programmes.
- **Observational study**: the study of groups of the general population with a specific diagnosis with no experimental control over their treatment.
- **Safety**: post marketing studies looking at safety of treatments when in general use.
- **Meta-analysis**: studies synthesising results from multiple independent trials.
- **Hospital audit**: all studies concerned with the performance and/or conduct of hospitals and other healthcare departments.

Further classifications, such as **data structures**, and the **free-text terms** are used to further specify a paper in more detail, perhaps in terms of the type of model being used (hierarchical, random effects), or the data involved (binary response, non-parametric). The possible terms for the field are kept to a minimum in order that similar papers, for example, all papers using

TABLE 2 *The classification scheme*

Type of paper	**Statistics**	**Issues**	
Discussion	Classical	Randomisation	Measurement error
Review	Bayesian	Missing values	Outliers
Methodology	Both	Clustered data	Measures of agreement
		Outcome measures	Data sources
Type of data	**Study type**	Transformations	Data monitoring
Original	Preclinical	Sensitivity and specificity	Trial development
Database	Clinical trial	Sample size determination	Standardisation charts
Simulated	Observational study	Covariates	Power calculations
Reanalysis	Screening	Multiple testing	
	Safety		
Preclinical study	Meta-analysis	**Free-text terms**	
Bio-assay	Hospital audit	Binary response	Bivariate response
PK–PD		Bootstrap	Calibration
Dose determination	**Study design**	Categorical covariates	Censored data
Bio-equivalence	Crossover	Clinical subgroups	Composite measurement scales
Toxicity	Equivalence	Computer software	Continual reassessment method
	Experimental design	Correspondence analysis	Cox regression
Data structure	Multicentre	Diagnostic markers	Drop-outs
Count data		Finite mixture models	Fixed effects
Discrete ordinal	**Type of analysis**	Frailty models	Goodness of fit
Discrete nominal	Estimation	Hierarchical	Lead-time
Continuous	Modelling	Markov models	Mean residual life
Survival data	Hypothesis testing	Mixed effects	Neural networks
Time series	Discrimination methods	Non-linear models	Non-parametric data
Repeated measures	Graphical modelling	Optimal design	Paired data
Repeated events	Graphical presentation	Physiological flow data	Pooling blood samples
Multivariate	Confidence intervals	Prognostic factors	Quality of life
Longitudinal		Random effects	Regression splines
Multi-state		Residual plots	Risk–benefit
Scoring scales		Robustness	ROC curves
Rates/proportions		Selection bias	Stopping rules
Contingency tables		Transitional models	Treatment allocation
		Tree regression	Trends
		Vaccine efficacy	

ROC, receiver operator characteristic

Cox regression, are classified similarly and so can be retrieved from the database in a systematic manner.

Issues

Those papers concerned with an aspect of statistical methodology rather than a specific type of analysis (e.g. randomisation, sample size calculations, outcome measures, missing values), are additionally classified as 'issue' papers.

Statistical research relevant to health technology assessment published in 1994–95

The completed database has 505 references, 271 from 1994 and 234 from 1995. A total of 855 different authors are included, two authors were involved in 11 papers each, four with six papers each.

Table 3 gives the major classification terms, with the number of papers that have been classified with this term.

Table 4 shows some of the major issues that were classified.

Discussion

Issues arising during review

This review was started in 1995 using fairly loose definitions of health technology assessment. Now that the HTA programme has been running for some years, there is a better shared understanding of health technology assessment as well as concrete examples, for example projects commissioned under the HTA programme. In chapter 7 we revisit the work with the benefits of this hindsight.

TABLE 3 *The demographics of the database*

Classification term	No. of papers
Type of paper	
Discussion	21
Review	54
Methodology	430
Statistics	
Classical	420
Bayesian	60
Both	20
Data source	
Original	12
Database	10
Re-analysis	224
Simulated	117
Study type	
Preclinical trials	45
Clinical trials	275
Observational study	44
Screening trials	25
Meta-analysis studies	15
Safety trials	6
Hospital audit	5
Type of analysis	
Modelling	133
Hypothesis testing	75
CIs	18
Estimation	81
Discrimination techniques	3
Graphical presentation	11
Graphical modelling	2
Data structure	
Discrete ordinal	22
Discrete nominal	28
Continuous	57
Survival data	102
Longitudinal data	47
Repeated measures	13
Repeated events	6
Multi-state data	8
Count/score data	8
Rates/proportions	14

TABLE 4 *Papers studying specific statistical issues*

Classification term	No. of papers
Measurement error	17
Missing values	37
Trial monitoring	47
Sample size determination	29
Power calculations	7
Multiple testing	19

However, many of the issues encountered would not change substantially.

Perhaps the most difficult issue for a review like this is the transferable nature of much statistical methodology. For example, work on missing values may be presented in the context of a clinical trial application, but be equally relevant to an aetiological epidemiological investigation or to a study of uptake of a screening programme. Initially in the classification scheme, such methods were coded as a default to parallel group trials, but it became apparent that this classification was meaningless, and so it was abandoned.

As a general rule, the more theoretical the paper, the harder it was to attribute a particular classification relevant to health technology assessment. A pragmatic rule was adopted: if a paper talked about 'estimating treatment' effects, it was included, but if it was phrased as investigating 'the properties of a location parameter' it was excluded. In a less general review, with more focused questions, these problems would largely resolve.

Generalisability

This review was fairly comprehensive in the sense it looked at most major statistical journals, likely to be publishing relevant material. It was limited to just 2 years as the original intention to expand it was too ambitious within a 1-year project. What is reasonably generalisable is the broad span and balance of this kind of statistical research. This will change relatively slowly, though for example, the recent explosion of interest in meta-analysis is likely to result in an increase in publications on the topic in the statistical literature. However, the field of statistics evolves, and so many of the papers will lessen in importance as that knowledge feeds into further research. The actual database is probably therefore of less relevance than the lessons learned from it. Partly for this reason, but mainly because of the resources that would be needed, the database is not being made generally available.

Further work

Two more in-depth reviews following on from this work are presented in chapters 4 and 5 of this report. Study design is fundamental to good statistical practice, so formed the basis of one review. In the initial trawl of papers, one of the most difficult areas to disentangle were the papers relating to methods of analysis looking at patients followed over time, so this formed the second review.

Chapter 5

Systematic review of study designs used in health technology assessment

Introduction

Good design underpins good-quality work relevant to the health technology assessment. Although not the only consideration, statistical aspects are integral to good study design, and so this area was chosen for a more in-depth review.

Aims

The aims of this review were to:

- describe the work being published in the statistical literature relevant to study designs
- describe the use of study designs appearing in the medical literature.

Methods

Data collection

Papers to be discussed in this review are those that have been classified as being concerned with a specific study design when entered onto the database as detailed in chapter 3. Of the 505 references on the database, 378 are relevant for inclusion into this review. The papers being reviewed are developing statistical methodology relevant to a specific study design or number of different study designs. For example, a paper on the interpretation of regression plots will be relevant to all study designs when the analysis requires regression to be used. *Table 5* indicates how many papers were found to be concerned with each study design.

Results

The results section will be presented in the following format. Each study design, or group of designs, will be described at the beginning of the subsection. A table will give an overall picture of how each design is represented in the database. For those designs represented by fewer papers will be reviewed in some detail. This may give an insight on why less research has been done on them and what needs doing in the future.

We have then used electronic searching of MEDLINE for the years 1993–96 and instances of each type of statistical methodology being used in practice. The major keywords used to describe the study designs will be used for text searching on as well as other terms describing statistical methodology that could be used in that study design.

Preclinical trials

Preclinical trials in the context of this review are those trials that are concerned with toxicity testing, pharmacokinetics/pharmacodynamics work (PK–PD), bioassays, determination of a dose that is both effective and safe, and bioequivalence studies. All these types of designs occur prior to a drug being administered to patients in a routine manner. *Table 6* displays the number of papers concerned with each preclinical study design.

These papers, which use preclinical study designs, are not directly relevant to health technology assessment but are building blocks towards it.

TABLE 5 *Frequency of study design papers*

Study design	No. of papers
Preclinical trials	45
Clinical trials	275
Screening trials	25
Meta-analytic studies	15
Hospital audit studies	5
Safety trials	6
Observational studies	44

TABLE 6 *Frequencies of preclinical trial designs*

Study design	No. of papers
Bioassay	8
Bio-equivalence	9
PK–PD	8
Dose determination	14
Toxicity	9

The drugs and treatments that successfully pass this stage in these studies then become the health technologies to be assessed in routine use. We do not review these papers in detail but have instead concentrated on those study designs that relate more directly to health technology assessment.

Clinical trials

By far the majority of the papers in this review use a clinical trials type study design. These trials may be a combination of the following, randomised or not, placebo controlled or not, crossover or parallel design, perhaps be multicentre trials or be of an experimental design layout. Frequencies of these papers are in *Table 7*.

TABLE 7 *Frequency of clinical trial types*

Study design	No. of papers
Crossover	20
Experimental design	9
Multicentre trials	11
Equivalence trials	7

From *Table 7* we see that the total number of papers referenced as one of these type of clinical trial designs is a lot less than the total number of papers referenced as concerned with clinical trials. The remaining clinical trials papers are not specific to any design. Instead they are concerned with methodology that would be applicable to many different situations or they are exploring an issue or technique that is non-study design-specific such as measurement error or sample size calculations.

Crossover trials

Crossover trials (*Table 8*) are often used to assess the effectiveness of a new drug when the disease being studied is chronic and its symptoms can be adequately controlled by medication but worsen when medication is withdrawn. One of the advantages of this design is that a smaller sample size is required as the within-subject variability is minimised by the subject acting as their own control.

Handsearch database (statistical)

Looking first at those papers studying the straightforward analysis of a two-period crossover trial, three of them are review papers.[74–76] Senn[74] gives a historical perspective of crossover trials and discusses many issues that are important in the design and analysis of these trials. The author asserts that the following issues would benefit from further research: Bayesian analysis of crossover trials, robust estimation, the role of time and alternatives to the AB/BA design. Kenward and Jones[75] review the methods of analysis of crossover trial when the data are discrete. Recent methodology that has been developed for the analysis of correlated categorical data is shown to be usefully applied to the analysis of crossover trials. Tudor and Koch[76] reviews the non-parametric methods that can be used to analyse crossover trials. Non-parametric methods can be used when the response variable is ordinal or has a censored time-to-event nature. These review papers include many references to other important papers and are a useful tool as a point of first contact for people involved in the analysis of two-period crossover trials. In Senn[77] the author includes his personal view on some issues in the use of crossover trials. Myers and co-workers[78] describes methodology for analysing crossover trials when the response is measured repeatedly over time and also the case of bivariate response. The restricted estimate maximum likelihood is shown to be applicable to crossover trials in Brown and Kempton.[79] The method estimates variance components in multi-classified data and so increases the precision of treatment estimates as information is efficiently combined across strata.

Bayesian methodology has been developed for the analysis of crossover trials and Grieve[80] gives a review of Bayesian analyses of crossover trials studying both cases of continuous and binary data. The author describes various crossover designs and investigates some more recent Bayesian work that can be regarded as being more robust in nature. Grieve[81] describes the use of a Bayesian analysis of a crossover trial when there are missing data, and its relative efficiency to analysing the data with those patients with missing data being excluded is evaluated. A graphical method has been developed in Grieve,[82] which aids in the Bayesian analysis of a two-period crossover trial with baseline measurements. Adjusting for baseline measurements can be important if there is thought to be a seasonal effect. In Forster[83] the Bayesian analysis of binary crossover data is investigated. The paper explains that the data considered are a 2^3 contingency table with the three binary factors being order of treatments. Waclawiw[84] is developing the empirical Bayes estimation and inference for a binary response random-effects model. This methodology is applied to the analysis of binary crossover data and is based on a fully parametric bootstrapping method.

TABLE 8 *Papers for crossover trials*

Reference	Title
Senn, 1994[74]	The AB/BA crossover: past present and future?
Kenward & Jones,1994[75]	The analysis of binary and categorical data from crossover trials
Tudor & Koch, 1994[76]	Review of nonparametric methods for the analysis of crossover studies
Senn, 1995[77]	A personal view of some controversies in allocating treatment to patients in clinical trials
Myers *et al*., 1994[78]	Fitting multivariate polynomial growth curves in two-period crossover designs
Brown & Kempton, 1994[79]	The application of REML in clinical trials
Grieve, 1994[80]	Bayesian analyses of two-treatment crossover studies
Grieve, 1995[81]	Extending a Bayesian analysis of the two-period crossover to accommodate missing data
Grieve, 1994[82]	Extending a Bayesian analysis of the two-period crossover to allow for baseline measurements
Forster, 1994[83]	A Bayesian approach to the analysis of binary crossover data
Waclawiw & Liang, 1994[84]	Empirical Bayes estimation and inference for the random effects model with binary response
Matthews, 1994[85]	Multi-period crossover trials
Carriere, 1994[86]	Crossover designs for clinical trials
Balagtas *et al*., 1995[87]	Marginal modelling of categorical data from crossover experiments
Bellavance & Tardif, 1995[88]	A nonparametric approach to the analysis of three-treatment three-period crossover designs
Liu, 1995[89]	Use of the repeated cross-over designs in assessing bioequivalence
Vuorinen & Tuominen, 1994[90]	Fieller's confidence intervals for the ratio of two means in the assessment of average bioequivalence from crossover data
Liu & Weng, 1995[91]	Bias of two one-sided tests procedures in assessment of bioequivalence

Multiperiod crossover trials are used when there are three or more treatment periods. They are of particular use when a carryover effect is present as this effect can be estimated in this model but not the two-period design when the inclusion of a carryover effect would produce a biased estimate of the direct treatment effect. Matthews[85] reviews multiperiod crossover trials and uses examples when the response variable is continuous. Carriere[86] shows how the three-period crossover design is more efficient than the two-period case particularly when faced with the practical limitations that can arise in clinical trials. This practical problem may be termination of a trial after the second treatment period or a large proportion of missing values. Balagtas and co-workers[87] shows how marginal models can be used to analyse a categorical response variable in a three-period crossover trial. A non-parametric approach to the analysis of the three treatment three-period crossover design in Bellavance and Tardif.[88]

Three papers use the crossover design in bioequivalence testing.[89–91] These papers will not be described further because, as explained in the previous section, bioequivalence testing is not of direct interest to health technology assessment.

TABLE 9 *Methodology used in crossover clinical trials*

Search number (#)	Keyword	No. of hits	No. method-ological
1	Crossover Crossover design Crossover designed Crossover studies	3029	
2	#1 and Bayesian	4	1
3	#1 and multi-period	3	2
4	#1 and bio-equivalence	68	11
5	#1 and carry-over	20	5

Electronic searching

For the period January 1993 to November 1996 an electronic search was done using MEDLINE (*Table 9*). Crossover studies are used regularly, and only a small proportion of the papers are methodological.

Experimental design

Handsearch database

The seven papers on the database classified as using an experimental design layout are more

complex designs other than crossover trials to analyse a variety of different situations (*Table 10*). Factorial, randomised block designs are seen, and balanced and unbalanced designs and both parametric and non-parametric data are dealt with.

Toman[92] and Slud[93] both look at the factorial design. Toman[92] uses a Bayesian approach "to solve the experimental design problems of selecting a fraction of the complete factorial experiment". Slud[93] looks at the use of a two-way factorial design for survival experiments. The framework of a proportional hazards model is used.

Methods to analyse non-parametric data, by using the ranked values instead, are put forward by Marden[94] and Alvo and Cabilio.[95] In Bellavance and Tardiff[88] non-parametric analysis of a treatment three-period crossover trial is transformed into a randomised block design. Giani and Strasburg[96] uses a randomised block design to test and select equivalent treatments. Finally, Ibrahim and Laud[97] views the analysis of designed experiments as a model selection problem and introduces the use of a predictive Bayesian criterion to facilitate model choice.

Multicentre trials

Handsearch database

Ten papers concerned with multicentre trials were entered onto the database (*Table 11*). Four of these papers use Bayesian analysis and of these, three are concerned with a survival outcome.[98–100] Gray[100] and Stangl[99] use a hierarchical model whereas Gustafson[98] uses a random effects model. The remaining Bayesian paper, Raghunathan[101] combines estimates of treatment effect across centres when the outcome is a binary response. Papers using classical methodology in the same situation are Hirji and co-workers[102] and O'Gorman and co-workers.[103]

Brunner and co-workers[104] and Kenward and co-workers[105] both present non-parametric

TABLE 10 *Papers for experimental design*

Reference	Title
Bellavance & Tardif, 1995[88]	A nonparametric approach to the analysis of three-treatment three-period crossover designs
Toman, 1994[92]	Bayes optimal designs for two- and three-level factorial experiments
Slud, 1994[93]	Analysis of factorial survival experiments
Marden & Muyot, 1995[94]	Rank tests for main and interaction effects in analysis of variance
Alvo & Cabilio, 1995[95]	Testing ordered alternatives in the presence of incomplete data
Giani & Strasburger, 1994[96]	Testing and selecting for equivalence with respect to a control
Ibrahim & Laud, 1994[97]	A predictive approach to the analysis of designed experiments

TABLE 11 *Papers for multicentre trials*

Reference	Title
Brown & Kempton, 1994[79]	The application of REML in clinical trials
Gustafson, 1995[98]	A Bayesian analysis of bivariate survival data from a multi-centre cancer clinical trial
Stangl, 1995[99]	Prediction and decision making using Bayesian hierarchical models
Gray, 1994[100]	A Bayesian analysis of institutional effects in a multicenter cancer clinical trial
Raghunathan, 1994[101]	Monte Carlo methods for exploring sensitivity to distributional assumptions in a Bayesian analysis of a series of 2 x 2 tables
Hirji *et al.*, 1994[102]	Efficient power computation for exact and mid-P tests for the common odds ratio in several 2 x 2 tables
O'Gorman *et al.*, 1994[103]	A comparison of two methods of estimating a common risk difference in a stratified analysis of a multi-center trial
Brunner *et al.*, 1995[104]	Nonparametric methods for stratified two-sample designs with application to multiclinic trials
Kenward *et al.*, 1994[105]	An application of maximum likelihood and generalised estimating equations to the analysis of ordinal data from a longitudinal study with cases missing at random
Davis & Chung, 1995[106]	Randomization model methods for evaluating treatment efficacy in multicenter clinical trials

methods to analyse multicentre trials when the outcome variables are scores or more general ordinal variables respectively. A randomisation model is introduced by David and Chung,[106] and Brown and Kempton[79] presents methodology that allows estimation of the variance components of a model. Therefore the variance attributable to different centres could be estimated.

Equivalence trials

Handsearch database

This type of hypothesis test has one major difference from the standard test of the efficacy of a treatment. In equivalence trials the null hypothesis states that the treatments are not equivalent, the alternative hypothesis, that the treatments are equivalent. Seven papers were classified as being concerned with equivalence trials, trials that are testing whether a new treatment has equivalent efficacy to an established standard treatment (*Table 12*). Nam[107] and Roebuck and Kuhn[108] both look at methods to calculate sample sizes for these types of studies. The five papers[96,108–111] are all concerned with the hypothesis testing element of these trials, that is, testing to see if the treatment are equivalent.

The final paper with this classification is Spiegelhalter and co-workers,[112] a general paper of the use of Bayesian analysis in clinical trials that includes a section about the Bayesian analysis of equivalence trials.

Electronic searching

Table 13 shows the MEDLINE searches for the experimental design, multicentre and equivalence type trials. It shows that multicentre trials are well used, though it is not obvious what type of methodology has been employed.

In search number 6 all three papers were found to be methodological, three of the four in number 7 were methodological, as were all three in number 8.

Screening trials

Screening programmes are used mainly to detect a disease in its early stages when no symptoms are apparent, for example for breast or cervical cancer. A large number of people are screened and those showing the early signs of the disease being screened for are then referred for further tests and contact with a specialist. By diagnosing a patient early, treatment of the disease is more effective. A screening programme is only worthwhile if treatment at the stage at which the cancer is detected by screening means that survival after diagnosis is longer than it would have been without a screening programme. RCTs are used to evaluate the potential benefits of a screening programme (i.e. screening trials). It is statistical methodology

TABLE 13 *Results of MEDLINE searches for clinical trial designs*

Search number (#)	Search term	No. of hits
1	Clinical trial(s)	16,833
2	#1 and experimental design	29
3	#1 and latin square	4
4	#1 and randomized block	1
5	#1 and multi-centre	1179
6	#5 and hierarchical	3
7	#5 and random effect(s)	4
8	#5 and Bayesian	3
9	#1 and equivalence	54

TABLE 12 *Papers for equivalence trials*

Reference	Title
Giani & Strasburger, 1994[96]	Testing and selecting for equivalence with respect to a control
Nam, 1995[107]	Sample size determination in stratified trials to establish the equivalence of two treatments
Roebruck & Kuhn, 1995[108]	Comparison of tests and sample size formulae for proving therapeutic equivalence based on the difference of binomial probabilities
Bofinger & Bofinger, 1995[109]	Equivalence with respect to a control: stepwise tests
Yamagousa *et al.*, 1994[110]	Mantel–Haenszel type tests for testing equivalence or more than equivalence in comparative clinical trials
Ng, 1995[111]	Conventional null hypothesis testing in active control equivalence trials
Spiegelhalter *et al.*, 1994[112]	Bayesian approaches to randomized trials

used in evaluating screening trials that is of interest to health technology assessment.

Handsearch database

Of the 23 papers concerned with screening programmes on the database two major groups of papers can be identified. Thirteen papers are concerned with the analysis of screening trials, the other ten with other forms of screening. The former are mostly cancer screening, and are represented in *Table 14*.

The review papers[113–116] do not necessarily introduce any new or innovative methodology but provide a valuable reference point for researchers involved in screening trials. They include many references to key papers and give an overview of the area.

The importance of RCTs in evaluation of screening is discussed in Etzioni and co-workers[114] and Stevenson.[113] The biases that can arise in screening studies, length and lead time bias can be minimised in RCTs as well as the biases inherent in all studies (i.e. selection bias). O'Neill and co-workers[115] illustrates the technical features of the design and analysis of a screening trial with reference to the South Australian Cancer Registry.

Many factors are important in the context of a screening trial. Some of these factors may be: the clinical benefit of a screening programme in terms of reducing mortality; identification of variables that influence the effectiveness of screening (perhaps by increasing compliance); the advantages and disadvantages of screening in a targeted population; and also the feasibility and cost-effectiveness of mass screening. In order to be able to evaluate the above factors mathematical models can be constructed and used as a tool to interpret the data and provide answers to all the questions that a screening trial promotes. The objectives of cancer screening models are to formally describe and analyse the processes involved, parameter estimation and hypothesis testing, predicting the results of screening policies in a realistic setting, analysis of the cost-effectiveness of a screening programme and the optimisation of screening policies.[116] Stevenson[113] and Van Ootmarssen and co-workers[116] review screening models and Duffy and co-workers[117] uses a specific type of a model, a Markov chain model, to model the screening policy.

TABLE 14 *Papers for screening trial methodology*

Type of analysis	Reference	Title
Trial design	Stevenson, 1995[113]	Statistical models for cancer screening
	Etzioni *et al.*, 1995[114]	Design and analysis of cancer screening trials
	O'Neill *et al.*, 1995[115]	A review of the technical features of breast cancer screening illustrated by a specific model using South Australian cancer registry data
Modelling	Stevenson, 1995[113]	Statistical models for cancer screening
	Van Oortmarssen *et al.*, 1995[116]	Modelling issues in cancer screening
	Duffy *et al.*, 1995[117]	Estimation of mean sojourn time in breast cancer screening using a Markov chain model of both entry to and exit from the preclinical detectable phase
Estimation and hypothesis testing	Connor & Prorok, 1994[118]	Issues in the mortality analysis of randomised controlled trials of cancer screening
	Duffy *et al.*, 1995[117]	Estimation of mean sojourn time in breast cancer screening using a Markov chain model of both entry to and exit from the preclinical detectable phase
	Etzioni & Self, 1995[119]	On the catch-up time method for analyzing cancer screening trials
	Kafadar & Prorok, 1994[120]	A data-analytic approach for estimating lead time and screening benefit based on survival curves in randomized cancer screening trials
	Stukel *et al.*, 1994[121]	Standardized rates of recurrent outcomes
	Xu & Prorok, 1995[122]	Non-parametric estimation of the post-lead-time survival distribution of screen-detected cancer cases
	Self & Etzioni, 1995[123]	A likelihood ratio test for cancer screening trials
	Thomson, 1995[124]	A hybrid paired and unpaired analysis for the comparison of proportions

Hughes and co-workers[125] specifically deals with case of a screening trial when risk factors are measured with error and repeated measures are instrumental in aiding the analysis of such data.

The following two papers discuss issues that do not relate just to screening trials but could be relevant to a researcher engaging in analysing a screening trial. Green and Freedman[126] discusses the situation where multiple interventions (multiple screening tools) are being tested. It may be possible to stop using specific screening tools and change to others before the end of the trial as the questions being asked about a specific tool could be answered in a shorter time than the length of the whole trial. The paper discusses the stopping of such trials. In many cancer trials, including cancer screening trials, surrogate endpoints are used as this can lead to a trial being of a shorter length. The use of surrogate endpoints in cancer and AIDS research is discussed in Fleming and co-workers.[127]

Wade and co-workers[128] describes the development of a screening tool to assess the visual acuity of children aged between 2 and 9 years. Age-related reference ranges are constructed against which children between these ages are compared.

Other papers were classified as being concerned with screening but not the type of screening described above. Eight papers[125,129–135] were concerned with the screening of blood and urine samples for a rare disease, the HIV virus in most of these papers, where samples are pooled and the prevalence of the disease is estimated. The method of pooling samples is favoured as it is cost-effective. This type of screening will not be discussed in this review. There is also a form of screening where patients are screened for eligibility for entrance to a clinical trial or, in the case of, a vaccine trial.[136]

Electronic searching

The MEDLINE database was searched for the years 1993 to the present using the keyword **Screening trial(s)** (*Table 15*).

Once the 32 screening trials were identified, the area they were covering was noted by scanning their abstracts in MEDLINE, not using keyword searching as might be implied (*Table 15*).

TABLE 15 *Clinical fields in which screening trials are carried out*

Search number (#)	Keyword	No. of hits
1	Screening trials	32
2	#1 and breast cancer	13
3	#1 and prostate cancer	5
4	#1 and ovarian cancer	2
5	#1 and (other) cancers	3
6	#1 and eardrum pathology	1
7	#1 and cystic fibrosis	1
8	#1 and (methodology)	7

Safety trials

Safety trials are conducted on a large scale on drugs pre- and postmarketing. Those that are premarketing trials make up part of the evidence submitted to the regulatory authorities as to the efficacy and safety of a new drug. Postmarketing studies are important in order to pick up adverse events that are rare or occur in long-term use of a drug or in a specific patient population.

Handsearched database

The six papers referenced as being concerned with drug safety (*Table 16*) can be split into two groups. Four are primarily interested in

TABLE 16 *Papers for safety trials*

Reference	Title
Enas & Goldstein, 1995[137]	Defining monitoring and combining safety information in clinical trials
O'Neill, 1995[138]	Statistical concepts in the planning and evaluation of drug safety from clinical trials in drug development: Issues of international harmonization
Lancar *et al.*, 1995[139]	Non-parametric methods for analysing recurrent complications of varying severity
Farrington, 1995[140]	Relative incidence estimation from case series for vaccine safety evaluation
Cook & Farewell, 1994[141]	Guidelines for monitoring efficacy and toxicity responses in clinical trials
Chueng-Stein, 1994[142]	A new proposal for benefit-loss risk analysis in clinical trials

the safety assessment of a new drug and the other two consider the risk–benefit trade-off, which is relevant in many drug therapies (e.g. chemotherapy).

Of this first group of papers Enas and Goldstein[137] and O'Neill[138] are concerned with design issues, which are of paramount importance when building up the safety profile of a new drug. It is emphasised that an awareness of subgroup differences in adverse event rates, the relationship between dose exposure and adverse event rates, and definitions of safety measurements are all issues that need to be taken into account in designing a package of trials that come together to provide a safety profile for a new drug. O'Neill[138] puts forward ideas for "statistical treatment of the three areas of dose response, subgroup differences and population exposure" stating that these are three areas of safety assessment that would benefit from further work. Methodology put forward includes adapting survival analysis techniques, in particular Cox's proportional hazard modelling with time-dependent covariates, to estimate the dose/toxicity–response relationship. Methods are also suggested to test for subgroup differences and the sample sizes required for such testing, and the sample size determination and length of follow-up issues in identifying rare adverse events in patients with chronic conditions. Enas and Goldstein[137] does not concentrate so much on the statistical methodology but gives guidelines for the design of both pre- and postmarketing surveillance studies and discusses 'stopping rules' for safety trials when a prespecified number of adverse events have been reported.

Lancar and co-workers[139] presents a non-parametric method of analysing a specific type of adverse event, that is, those complications which recur, for example headaches and dizziness. They also take into account the varying severity of adverse events. Farrington[140] looks at methodology involved in assessing the safety of a vaccine preparation.

Of the two risk–benefit papers, Cook and Farewell[141] describes guidelines that can be used when efficacy and toxicity responses are being evaluated simultaneously in a clinical trial, whereas Cheung-Stein[142] puts forward methodology that conducts a risk analysis of benefit–loss in clinical trials.

Electronic searching

The MEDLINE database was searched for the time period January 1993 to November 1996. The following search expression 'safety assessment, not Phase 1, and human' identified 47 papers. Further searches were not carried out because of the wide diversity in this area as described below.

There are many papers assessing safety of drugs being used in standard clinical practice are many. However, through browsing through the abstracts of these papers it appears that a simplistic approach to the presentation of safety assessment is taken. By far the majority of the papers, if not all, simply present data on safety by simple frequencies of number of adverse events recorded in the follow-up time. Some of the papers, particularly the large studies, graded the severity of the adverse events.

Through the electronic searching of MEDLINE other types of safety assessment were identified. A lot of work is published on the safety of drugs at the preclinical laboratory stage when the drugs are tested on animals. Safety assessment of drugs at this stage is not directly relevant to health technology assessment as it is likely that the majority of these drugs will never go into standard clinical use, though for those that do, observations at this stage may help guide appropriate assessment of safety subsequently.

Another area of safety assessment not concerned with health technology assessment but is more of a public health issue is the safety testing of food packaging and other possible sources of food contaminants. A number of papers were found in this area.

An area that is related to health technology assessment, if somewhat indirectly, is the safety of equipment in hospital, for example magnetic resonance imaging machines or substances injected that show up on imaging machinery. Many of these substances are radioactive and so their safety has to be assured.

Many clinical trials carried out to assess the efficacy of a new drug also made some assessment of the safety of the drug. This assessment was mostly simply in terms of frequencies of complications reported. The safety assessment of the drug was not the primary objective of the trial in most cases.

Meta-analyses

Meta-analysis studies are studies that combine (synthesise) the results of a set of trials investigating the same or similar hypothesis. The trials

being studied are most likely to be RCTs but methodology is now being developed to combine the results of different study designs.

This area has not been further investigated as a full systematic review of meta-analysis has been funded.[143]

Observational studies

Observational study designs are frequently used in epidemiology but have other applications as well. Some drug treatments may have to be assessed over a long period of time and a clinical trial is not suitable. Therefore, the long-term performance of a drug can be assessed in this way. Also, safety assessment of a drug is often performed in an observational study design framework.

Some papers have been classified as referring to the observational study design because the methodology they are developing would be equally applicable to that design even if in the paper another study design is referred to. This is particularly true of many of the generalised linear modelling papers. In these cases a paper was classified as to being concerned with more than one study design. However, having taken a decision to exclude pure epidemiological papers, this section is somewhat incomplete, and we do not pursue it further at this stage.

Discussion

Issues arising from review

It is striking how much of the statistical work relating to study design comes from preclinical and simpler clinical trials. Relatively little comes from safety studies, more complex types of experimental design or meta-analysis. For reasons of time constraints, traditional epidemiological studies were excluded.

From the perspective of health technology assessment, the preclinical and clinical trials are important, but work in synthesing studies, such as meta-analysis, and safety studies, looking especially at the use of technologies in populations are particularly important. It is hard to believe there are no statistical challenges left in this work, so the parity of current work is disappointing.

The MEDLINE searches were interesting, but only offer a glimpse. Again, due to time constraints, these could not be pursued more thoroughly.

Further work

Further work on statistical methodology of study designs would be useful, but should be rooted in the more mature understanding of health technology assessment that is now possible (see chapter 7).

Chapter 6

Systematic review of methods of analysis used in studies that follow patients over time

Introduction

Most data in health technology assessment are longitudinal in nature. In the simplest case, baseline measurements can be recorded and then another measurement is taken at the end of the study. In this case the change in the response variable may be the value to be analysed and is therefore reduced to a single variable. In this review this simple kind of longitudinal data is not considered, and only those cases when two or more measurements on each subject have been recorded are treated as longitudinal.

The nature of a disease will determine the structure of the outcome measure. For example, if a disease is terminal, as with some cancers, then the data to be analysed will be survival data or failure time data, the length of time from diagnosis to death. However, if a disease is chronic, such as asthma, then it would be more suitable to study patients over a long time recording specific outcomes at various times over the total length of study (i.e. longitudinal data).

The response variable in longitudinal data can be discrete or continuous and will be measured at prespecified intervals. These data are very prone to missing values. Also patients may be not measured at the same times leading to unbalanced messy datasets. These problems complicate analysis.

In survival data the outcome in its simplest form is the time to death or a prespecified definable event. However, the data may be in the form of repeated events, where a symptom recurs at various intervals. Alternatively there could be multistate data, which would occur when there is a separate length of time to each stage of the disease. Survival data are also complicated by the presence of censored observations.

The term 'repeated measures' sometimes refers to longitudinal data as was described above but it is also used for replicated measures. This can occur when measurements are made a number of times immediately after each other, for example a peak flow measurement of lung capacity. These replicated measurements are made in order to reduce random error in the estimate. However these types of measures are often termed repeated measures and so terminology can become confusing.

Time series also follow patients over time by taking repeated measurements over a prespecified length of time. However, time series differ from longitudinal data and repeated measures data as they typically arise following a small number of patients with short intervals between measurements and many measurements being made.

All these types of data where multiple measurements are made on each patient can be considered as multivariate data. This introduces the complication of correlated data as each measure on a specific patient will be measuring the same feature. Multivariate longitudinal data are also possible where more than one response variable is being recorded at each timepoint, and the correlation structure will be even more complex.

In this review survival data, longitudinal data and repeated measures data will be examined with respect to the methods of analysis that are being developed to deal with them. Time series data will not be examined in detail.

Examples of types of data used for studying patients over time[144]

The occurrence of a specified event, such as death or the detection of metastases in patients with cancer of the colon,[145] is the focus of many studies and, particularly, RCTs of cancer treatments. **Survival data** measure the time between entry to study and exit from it for each patient recruited. A patient may leave a study for a variety of reasons other than experiencing the event of interest, for example withdrawal or migration. These latter reasons are collectively called censoring, the patient ceases to contribute survival time to the study, but the time to the most recent follow-up is included in the analysis.

Longitudinal data occur when an outcome of interest is measured on more than one occasion. An example is the measurement of CD4 counts at successive clinic visits in patients with HIV.[146] Longitudinal measures of QoL are becoming increasingly important, and can be combined with survival data to evaluate the trade-off between quantity and QoL in fatal diseases.

More than one measurement is sometimes obtained for a patient for reasons other than the follow-up time. In this review we have called these **repeated measures data**. (The term 'repeated measures' is sometimes used interchangeably with the term 'longitudinal data' but we have made a distinction between these types of data for the review.) For example, independent opinions may be sought from orthodontists on the need for extraction of children's teeth.[147]

A **time series** is a sequence of measurements over time, such as the number of new HIV notifications on a monthly basis.[148] When the primary interest is the analysis of an individual (long) series, time series are the most appropriate method. Analysis centres on identifying underlying trends and periodicity in the data, for example seasonal effects, often with a view to prediction for the future. When there are (often quite short series) on many individuals, methods developed for longitudinal data (see above) are likely to be more appropriate. In both cases, the statistically important feature is the correlation between observations within a series, which must be appropriately modelled if correct conclusions are to be drawn.

Aims

The aims of this review were to:

- describe the work being published in the statistical literature related to longitudinal data, repeated measures and survival data
- describe the use of these three sets of techniques in the medical literature.

Methods

The database was searched for all papers that were classified as being concerned with the following data types; longitudinal data, survival data, repeated measures and time series.

Various features of the papers identified were tabulated including the source of the data, structure of response variable and type of statistics used. Following this, each type of data was treated separately, and the methods being developed in the literature with which they are analysed was described.

Electronic searching of MEDLINE was done for the period January 1993 to October 1996. These electronic searches were to identify how these types of studies are analysed in practice.

Basic classifications

Table 17 shows the number of papers classified as being concerned with each data structure under study in this review.

TABLE 17 *Frequency of classification term*

Data structure	No. of papers
Longitudinal data	47
Repeated measures	13
Survival data	102
Time series	6

Table 18 shows by far the majority of papers are using real data rather than simulated. It seems the new methodology has been developed in response to real practical problems uncovered by researchers who have carried out longitudinal type studies.

TABLE 18 *Source of data used in the papers in this review*

Data structure	Data source	No. of papers
Longitudinal data	No data	1
	Original	1
	Re-analysis	33
	Simulated	7
	Re-analysis and simulated	6
Repeated measures	No data	2
	Re-analysis	8
	Simulated	3
Survival data	No data	23
	Original	3
	Re-analysis	46
	Simulated	18
	Re-analysis and simulated	13
	Database	5
Time series	Original	2
	Re-analysis	4

The numbers for the time series are biased downwards as papers that were not directly relevant to health technology assessment were not included. There are many theoretical papers written on time series but to a non-specialist in time series it was not possible to ascertain how the theory was likely to be applied in practice if at all.

The few papers that are classified as not using any data were discussion or review papers, which did not use data to illustrate their point.

As can be seen in the *Table 19* very few papers use Bayesian methodology. Those papers classified as 'both' are likely to be review papers that demonstrate both classical and Bayesian methodology in a particular research area.

The majority of papers are developing methodology for continuous response variables (*Table 20*). Continuous response variables are much more straightforward to analyse than discrete variables but the problems of analysing discrete variables in a longitudinal setting has been addressed by various authors.

TABLE 19 *Statistical framework of papers in this review*

Data structure	Statistical framework	No. of papers
Survival data	Classical	104
	Bayesian	3
	Both	1
Longitudinal data	Classical	43
	Bayesian	3
	Both	2
Repeated measures	Classical	12
	Bayesian	0
	Both	1
Time series	Classical	6

TABLE 20 *Data structure of papers used in this review*

Data structure	Data type	No. of papers
Longitudinal data	Not specified	2
	Continuous	26
	Discrete ordinal	7
	Discrete nominal	5
	Binary response	8
Repeated measures	Not specified	2
	Continuous	9
	Discrete ordinal	2
	Discrete nominal	1
	Binary response	1

Longitudinal data

Handsearched database

Of the 48 papers classified as developing methods for longitudinal six are reviews (*Table 21*). Everitt[149] offers a very practical review of methods used to analyse data where the same response variable is recorded on each observational unit on several different occasions. In Fitzmaurice and co-workers[150] the authors review the different models that can be used when dealing with the specific case of binary responses with the drop-out mechanism being informative. The analysis of binary responses are once again being examined in Ten Have and co-workers[150] but the issue of clustered observations is of prime importance here specifically in the case of dental studies where observations are made on multiple tooth sites. Wang-Chow and co-workers[152] reviews different estimators of treatment effect when the missing data arise from one of four processes. The estimators are assessed by means of a simulation study. Both parametric and non-parametric methods for analysing longitudinal data with incomplete observations are reviewed in Wu and co-workers.[153] The methods are assessed both under the assumption of random censoring and informative censoring. The analysis of recurrent (or repeated) events can be analysed using survival methods but in Lawless[154] longitudinal methods are presented to analyse this type of data.

TABLE 21 *Papers for longitudinal data*

Reference	Title
Everitt, 1995[149]	The analysis of repeated measures: a practical review with examples
Fitzmaurice *et al.*, 1995[150]	Regression models for longitudinal binary responses with informative drop-outs
Ten Have *et al.*, 1995[151]	Association models for peridontal disease progression: a comparison of methods for clustered binary data
Wang-Clow *et al.*, 1995[152]	A simulation study of estimators for rates of change in longitudinal studies with attrition
Wu *et al.*, 1994[153]	Testing for differences in changes in the presence of censoring: parametric and non-parametric methods
Lawless, 1995[154]	The analysis of recurrent events for multiple subjects

The two main ways of analysing longitudinal data are marginal modelling (incorporating generalised estimating equations) and random- or mixed-effect modelling. The references developing these approaches are listed in *Table 22*.

Transitional models can also be used and Follman[162] deals with these type of models.

The main aim of longitudinal analysis is to reduce the multiple observations on an individual to a summary statistic and to then use the summary statistics of the sample of individuals in further analysis. These summary statistics are often in terms of growth curves, when continuous data are involved. Six papers refer to growth curves[78,174–178] and other summary statistics (*Table 23*).[149,179]

Lindsey[180] and Tsiatis[164] both look at the relationship between survival and longitudinal analysis. In Diggle and Kenward[181] and Little[182] the drop-out process or attrition is modelled simultaneously with the data so that the often inappropriate assumption of random drop-out need not be made.

TABLE 22 *Papers for longitudinal data*

Modelling approach	Reference	Title
Marginal models, generalised estimating equations	Lipsitz *et al.*, 1994[155]	Performance of generalized estimating equations in practical situations
	Robins *et al.*, 1995[156]	Analysis of semiparametric regression models for repeated outcome in the presence of missing data
	Williamson *et al.*, 1995[157]	Analyzing bivariate ordinal data using a global odds ratio
	Lipsitz *et al.*, 1994[158]	Analysis of repeated categorical data using generalized estimating equations
	Robins & Rotnitzky, 1995[159]	Semiparametric efficiency in multivariate regression models with missing data
	Miller, 1995[160]	Analysing categorical responses obtained from large clusters
	Fitzmaurice & Lipsitz, 1995[161]	A model for binary time series data with serial odds ratio patterns
	Follmann, 1994[162]	Modelling transitional and joint marginal distributions in repeated categorical data
	Mark & Gail, 1994[163]	A comparison of likelihood-based and marginal estimating equation methods for analysing repeated ordered categorical responses with missing data: application to an intervention trial of vitamin prophylaxis for oesophageal dysplasia
Random/mixed effect models	Tsiatis *et al.*, 1995[164]	Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS
	Goldstein *et al.*, 1994[165]	Multilevel time series models with applications to repeated measures data
	Pearson *et al.*, 1994[166]	Mixed-effects regression models for studying the natural history of prostate disease
	Follmann & Wu, 1995[167]	An approximate generalized linear model with random effects for informative missing data
	Morrell *et al.*, 1995[168]	Estimating unknown transition times using a piecewise nonlinear mixed-effects model in men with prostate cancer
	Palta *et al.*, 1994[169]	Testing for omitted variables and non-linearity in regression models for longitudinal data
	Zerbe *et al.*, 1994[170]	Studying the relationship between change and initial value in longitudinal studies
	Anderson & Jones, 1995[171]	Smoothing splines for longitudinal data
	Gilula & Haberman, 1994[172]	Conditional log-liner models for analyzing categorical panel data
	Qu *et al.*, 1995[173]	Latent variable models for clustered ordinal data

TABLE 23 *Papers for longitudinal study*

Reference	Title
Myers *et al.*, 1994[78]	Fitting multivariate polynomial growth curves in two-period crossover designs
Lindstrom, 1995[174]	Self-modelling with random shift and scale parameters and a free-knot spline shape function
Mori *et al.*, 1994[175]	Slope estimation in the presence of informative right censoring: modeling the number of observations as a geometric random variable
Donnelly *et al.*, 1995[176]	Prediction and creation of smooth curves for temporally correlated longitudinal data
Cole, 1994[177]	Growth charts for both cross-sectional and longitudinal data
Wang & Taylor, 1995[178]	Inference for smooth curves in longitudinal data with application to an AIDS clinical trial
Dawson, 1994[179]	Stratification of summary statistic tests according to missing data patterns
Everitt, 1995[149]	The analysis of repeated measures: a practical review with examples
Lindsey, 1995[180]	Fitting parametric counting processes by using log-linear models
Tsiatis *et al.*, 1995[164]	Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS
Diggle & Kenward, 1994[181]	Informative drop-out in longitudinal data analysis.
Little, 1995[182]	Modeling the drop-out mechanism in repeated-measures studies
Wu *et al.*, 1994[183]	Use of surrogate information time for monitoring the effect of treatment on the change in a response variable in clinical trials
Lavori *et al.*, 1994[184]	Causal estimation of time-varying treatment effects in observational studies: application to depressive disorder
Lavori *et al.*, 1995[185]	A multiple imputation strategy for clinical trials with truncation of patient data
Cole *et al.,* 1995[186]	An empirical Bayes model for Markov-dependent binary sequences with randomly missing observations
Kim & Lagakos, 1994[187]	Assessing drug compliance using longitudinal marker data with application to AIDS

Data monitoring of a longitudinal study is examined in Wu and co-workers,[183] in Lavori and co-workers[184] the specific issue of time-varying covariates is studied, and a Bayesian approach to the analysis of longitudinal studies is put forward in Lacori and co-workers[185] and in Cole and co-workers.[186] Finally, Kim and Lagakos[187] uses a longitudinal structure to assess drug compliance in persons infected with HIV.

Electronic searching

The search in *Table 24* was undertaken in MEDLINE.

TABLE 24 *Methods of analysis used in practice for longitudinal studies*

Search number (#)	Keyword	No. of hits
1	Longitudinal studies	5319
2	#1 and random effects	20
3	#1 and mixed effects	4
4	#1 and mixed models	1
5	#1 and marginal models	3
6	#1 and generalised estimating equations	1

In *Table 24* search number 2, 13 of these 20 papers were methodological and already on the database. Search number 3 found two applications of this type of analysis, the other two were methodological. The one paper found in search number 4 is on the database and also two of the three found in search number 5. Considering the huge number of longitudinal studies very few can be identified as being analysed using the type of methodology being developed in the statistical literature.

In the introduction it was suggested that longitudinal studies should be used when chronic diseases are being investigated. In an attempt to discover what type of diseases are being analysed using longitudinal studies various diseases were used as keywords (*Table 25*). Terminal diseases were also included so that a comparison could be made with the type of diseases analysed using survival methods.

TABLE 25 *Clinical areas using a longitudinal design*

Search number (#)	Keyword	No. of hits	% of #1
1	Longitudinal studies	5319	–
2	#1 and MI, myocardial infarction, cardiac, or heart disease	267	5.0
3	#1 and psychiat*, depress* or dementia	759	14.3
4	#1 and cancer, leukemia, or lymphoma	259	4.9
5	#1 and AIDS or HIV	298	5.6
6	#1 and stroke	99	1.9
7	#1 and respiratory disease, asthma, or bronchitis	92	1.7
8	#1 and arthritis	127	2.4
9	#1 and epiliepsy	19	0.4
10	#1 and diabetes	202	3.8

Repeated measures

Database

Thirteen papers deal with repeated measures data (*Table 26*), Ten Have and Chinchilli[188] uses a Bayesian hierarchical model to analyse the data but the remaining 12 all use a classical statistical framework. Three papers[189–191] look at sample size estimation in repeated measures designs under varying situations. The estimation and minimisation of within-unit variance was the primary objective for six of the studies[79,192–196] using the non-parametric case for study.

In Hughes and co-workers[125] a repeated measures design is used to better estimate blood pressure, a measurement prone to measurement error, in a screening trial. Numgung and Yang[197] reports a repeated measurement design that was utilised in a dental study in order to reduce outliers. Contingency tables in a repeated measures design are analysed in McCloud and Darroch.[198]

Electronic searching

Using the keyword **repeated measures**, 947 papers were found. However, many of these hits occur when the abstracts report that repeated measures

TABLE 26 *Papers for repeated measures*

Reference	Title
Ten Have & Chinchilli, 1994[188]	Bayesian hierarchical analysis of within-units variances in repeated measures experiments
Lipsitz & Fitzmaurice, 1994[189]	Sample size for repeated measures studies with binary responses
Overall & Doyle, 1994[190]	Evaluating sample size for repeated measures design
Kirby et al., 1994[191]	Sample size estimation using repeated measurements on biomarkers as outcomes
Silverberg, 1994[192]	A simulation study comparing two approximations for a quasi t-quantile used in repeated measures ANOVA
Brown & Kempton, 1994[79]	The application of REML in clinical trials
Dinse, 1994[193]	A comparison of tumour incidence analyses applicable in single-sacrifice animal experiments
Chinchilli *et al.*, 1995[194]	Partial likelihood analysis of within-unit variances in repeated measurement experiments
Lin & Hughes, 1995[195]	Use of historical marker data for assessing treatment effects in Phase I/II trials when subject selection is determined by baseline marker level
Hayter & Hsu, 1994[196]	On the relationship between stepwise decision procedures and confidence sets
Hughes et al., 1995[125]	Optimal sequential screening guidelines for quantitative risk factors measured with error
Namgung & Yang, 1994[197]	Outlier reduction by an option-3 measurement scheme
McCloud & Darroch, 1995[198]	An analysis of categorical repeated measurements in the presence of an external factor

analysis of variance (or multivariate analysis of variance) was used in the analysis of a study. When these terms are removed 325 papers remain.

Survival data

Database

Due to the large number of papers found dealing with survival data and to time constraints, a detailed study could not be undertaken as in the above two sections. However, *Table 27* shows the areas where research is being done in survival analysis.

TABLE 27 *Areas of survival analysis being studied*

Area of study	No. of papers
Non-parametric estimation/modelling	13
Non-parametric tests	7
Proportional hazards modelling	28
Proportional hazards tests	4
Parametric models	9
Semi-parametric models	6
Discrete time models	5
Mixture models	6
Frailty models	6
Trees and splines	5
Residuals/goodness of fit	3
Multi-state data	4
Repeated events	3
Sample size/power calculations	5
Trial monitoring/design	8
QoL	2
Neural nets	1

Most of this work is being done in the field of proportional hazards modelling, much of this Cox regression, a technique widely used in practice (*Table 28*). However, it is also interesting that other approaches to modelling survival are being developed like frailty models, mixture models, semi-parametric approaches and a single paper on neural nets, a type of methodology currently being applied to many different areas of medical research.

Electronic searching

Search numbers 5, 6 and 10 produced three, two and three methodological papers, respectively. *Table 29* displays the type of diseases that are analysed using survival methods.

It can be seen that nearly a third of all the survival data papers are in cancer studies. The other major terminal disease, heart disease (and associated terms) accounts for a further 10%.

TABLE 28 *Types of survival analysis being applied*

Search number (#)	Keyword	No. of hits
1	Survival or survival analysis	51,873
2	#1 and proportional hazards	1746
3	#1 and Cox regression	358
4	#1 and non-parametric	32
5	#1 and semi-parametric	6
6	#1 and parametric	84
7	#1 and quality of life	1407
8	#1 and mixture models	3
9	#1 and repeated events	2
10	#1 and multi-state	5
11	#1 and accelerated life	1

TABLE 29 *Clinical areas where survival analysis is being used*

Search number (#)	Keyword	No. of hits	% of #1
1	Survival	47,281	–
2	#1 and MI, myocardial infarction, cardiac, or heart disease	4506	9.5
3	#1 and psychiat*, depress* or dementia	156	0.3
4	#1 and cancer, leukemia, or lymphoma	15,214	32.2
5	#1 and AIDS or HIV	1251	2.6
6	#1 and stroke	810	1.7
7	#1 and respiratory disease, asthma, or bronchitis	131	0.3
8	#1 and arthritis	155	0.3
9	#1 and epilepsy	52	0.1
10	#1 and diabetes	996	2.1

Discussion

Issues arising during the review

One of the major problems in doing this review was the terminology used by the authors. The terms longitudinal and repeated measures are often used interchangeably and, although in an everyday sense this is not important, it does cause considerable problems when an attempt is made to classify these areas. The description of the terms in the introduction show how the papers were classified for this review. The problem also occurs in electronic searching but not in the same way. When 'repeated measures' is used as a keyword all papers using repeated measures analysis of variance are picked up. In this case the search is picking up the right type of papers. However, 'longitudinal studies' as a keyword picks up numerous papers, many of which, statistically speaking, are survival studies.

We postulate that longitudinal methods are underused in the analysis of potentially longitudinal data. The word potentially is used because in some studies where data have been collected longitudinally, problems such as missing values and an unbalanced design (data not being collected at the same time intervals for all subjects) can deter researchers from analysing longitudinally. They may instead analyse the study cross-sectionally. This not only results in many data not being used but also it can result in incomplete or weak conclusions. Using more powerful methods may, for example, mean that subgroup differences can be identified and trends analysed.

Searching MEDLINE to find where statistical methodology is being used is a very inexact science. When searching on statistical text words many of the papers unearthed will be methodological papers. However, it is difficult to ascertain whether certain statistical techniques or modelling approaches are more widely used than is apparent in the tables above, as authors do not always include brief details of the statistical methodology that has been used in their abstracts. Greater use of structured abstracts should improve the situation, as greater use of statistical terms as search headings is probably not realistic.

Further work

Since carrying out this work, an in-depth health technology assessment review of QoL and survival data has been reported.[2] Although primarily focused on studies that report both, it in fact contains a lot of valuable work on both longitudinal and survival data, and the interested reader is strongly advised to read it.

Chapter 7

Review of the statistical needs of the Health Technology Assessment programme

The development of the HTA programme

This programme was one of the first set of HTA projects commissioned. In retrospect, it would have benefited from a clearer view of health technology assessment. In turn this is now possible as thinking on health technology assessment has matured, and there are plenty of concrete examples through other commissioned projects. It is not possible to repeat the exercise at this stage, but in this chapter we reflect with the benefit of hindsight.

The role of health technology assessment

Health technology assessment is defined by the HTA programme as "an internationally recognised term that covers any method used by those working in health services to promote health, prevent and treat disease and improve rehabilitation and long-term care" (http://www. ncchta.org/abouthta.htm). Technologies, in this context, are not confined to new drugs or pieces of sophisticated equipment.

Health technology assessment considers the effectiveness, appropriateness and cost of technologies. It does this by asking four fundamental questions:

- Does the technology work?
- For whom?
- At what cost?
- How does it compare with alternatives?

We look at each of these four questions in turn, identifying the statistical issues raised, and reflecting to what extent they are addressed through the textbooks, guidelines and statistical research reviewed.

Does the technology work?

What are the issues?

For a particular technology, probably the most important question to establish is what is meant by 'working'. For an anaesthetic, the effect may be easy to define, but for say an asthma treatment, the definition is likely to be in terms of an average improvement in lung function as measured by, say forced expiratory volume in 1 second.

Considering asthma also highlights many of the issues common to chronic diseases. "Does it work?" needs a time frame. Typically for drug regulation purposes short-term placebo-controlled trials (often crossover) establish an effect, and then trials over a few months, often active controlled, establish some medium-term efficacy. A further important question relates to long-term efficiency (asthma treatments are typically taken for years, not months). Also, forced expiratory volume in 1 second is not really of interest in its own right, but rather it is used as a marker for current QoL, and as a predictor of longer-term morbidity and mortality.

Further questions are raised when considering whether technologies are still effective in routine use, outside the context of special studies.

How does existing knowledge help?

Some parts of the question 'Does it work?' are addressed through the guidelines developed for drug regulation. Where initial demonstrations of efficacy can be through randomised trials, methodology is well established. In principle, the best way of establishing long-term efficacy is through appropriate length trials, with appropriate endpoints. When these have not taken place, other methods may need to be considered, whether as an alternative, or as part of the planning of longer-term studies.

This essentially involves modelling. It may be the naïve assumption that effects observed in the medium term continue, or it may involve quite complex modelling. The concept of a 'surrogate' is relevant here. A 'surrogate' is a short-term measure that is predictive of longer-term outcomes. Establishing this for a particular marker in particular circumstances is not trivial. Guidelines for drug regulation touch on this.[62]

For technologies with less clear parallels with drug regulation, other issues are also pertinent. For organisational interventions, appropriate

design and analysis are needed. Within RCTs there is a growing body of knowledge on cluster randomised trials.[199] More generally hierarchical modelling (see chapter 2) is the appropriate technique.

For whom?

What are the issues?

From a narrow perspective, where trials are possible, they should be designed for the target group. Where there are special groups, for example children, the elderly, the renally impaired (for drugs), these need explicitly addressing. However, there are challenges within this framework.

Trials are generally powered to look at the treatment effect overall. Just because, say, both men and women have been entered does not imply the treatment works equally well in both, but there is typically not enough power to look at interactions between sets and treatment effects.

Conversely there may have been exclusion for pragmatic or ethical reasons, such as inability to speak English, or being of child-bearing potential. In this case we need to extrapolate beyond the trial participants.

If a trial took place some time ago, it may be that concomitant treatments have changed, and so the question here is of extrapolation across time.

How does existing knowledge help?

The technology for formal testing of interactions is standard (see chapter 2 on generalised linear models); the usual problem is lack of power. When concerns are sufficiently serious, a new study in the relevant group can be commissioned, again using standard methodology.

Beyond this, we are again in the realms of modelling, either implicitly through simple extrapolation, or more carefully, taking cognisance of underlying biology. Judgements about similarity or the potential range of heterogeneity between groups of patients can be modelled using Bayesian techniques. HTA reviews on subgroup analysis have recently been commissioned.

At what cost?

What are the issues?

First, costs must be defined. This will depend on who is asking the question. Costs to a patient will include side-effects, time (e.g. the time taken by regular physiotherapy for patients with cystic fibrosis) and financial. Costs to a provider can be couched financially or as opportunity costs. For some conditions, parents, relatives or other carers may incur costs of time, money or loss of quality of relationships.

How does existing knowledge help?

The definitions of costs and the collecting of appropriate data fall into several sorts: medical costs are in the province of clinical researchers, social costs are best understood by social scientists and costs generally are the province of economists. Statistically they are all measured with uncertainty, which is not always handled as well as efficacy endpoints. When the costs come form different sources, the integration may require quite complex modelling. Currently commissioned HTA programme reviews on economic issues are relevant.

How does it compare with alternatives?

What are the issues?

Health technologies do not exist in isolation. In practice, choices have to made. For example, a patient may have to choose whether to undergo chemotherapy prior to surgery for a cancer. A primary care team may need to decide which strategy to use for early detection of diabetes. This leads us towards a decision-making perspective, where the decision has to integrate the previous three questions.

Choices made between alternatives may be made by the patient, their doctor, a practice or primary care group, a trust, or at national level. It is necessary to be explicit about the decision-making context, as the considerations may be different to different decision makers.

For a given choice between alternatives for a particular decision maker information is needed on comparative benefits and costs. For an individual these ideally need to be fine-tuned to their own situation, for decisions relating to groups of patients this balance needs to be by appropriate subgroups.

How does existing knowledge help?

Current standards do not go very far in answering these questions. Where it is possible to run comparative trials, standard methodology as taught in MSc courses is applicable. The guidelines for drug regulation are not very pertinent. This is because drug regulation (at least in the UK) is based on demonstration of efficiency and safety in an absolute sense. Comparative efficacy is only a reason for refusing a licence if there is concern about risk–benefit compared with an existing product. This usually means a considerably worse

safety profile with no relative efficacy benefit for any identifiable group of patients.

The whole science of decision making as applied to health needs further work, in terms of studying what people base decisions on, and how to measure and present such information. Some of this is being covered by other HTA programme reviews. From a statistical perspective, work on decision making does not feature in MSc courses, or research in 1994–95, but a developed theory of decision makers under uncertainty has been proposed by various Bayesians (Lindley,[200] Smith,[201] and Bernardo and Smith[202]), which warrants further investigation. In particular what would be useful are fully worked through case studies.

Some of the questions about choices between alternatives may be complex, involving different doses or frequencies of treatment or different combinations of possibilities. The design of trials taught in MSc medical statistics or referred to in drug regulatory guidelines are fairly simple, typically two-group parallel or crossover trials. These cannot answer more complex questions that health technology assessment raises.

However, within more mainstream statistics there is a long tradition of design of experiments stemming from agricultural trials, and more recently industrial settings. Work here includes different factors varying at different levels (perhaps practice level and patient level) and choice of design when resources will not allow for the testing of all possible permutations. This would seem a fruitful area for further work. As a first step, the working-up of specimen protocols would be useful to see where particular issues arise.

Recommendations

- Case studies are needed on decision making under uncertainty using established Bayesian methodology to integrate health outcomes with wider costs.
- Established statistical methodology on design of experiments is potentially relevant to complex questions in health technology assessment. The development of specimen protocols explicitly using such methodology could be commissioned.

Acknowledgements

We would like to thank the HTA programme for financial support. We thank referees for their careful criticisms of the first version of this report. We are grateful to colleagues in Liverpool, especially Simon Fear, Paula Williamson and Carrol Preston for constructive discussion. We thank AstraZeneca and the ASRU for helpful discussions about their statistical guidelines.

References

1. Bartholomew DJ. What is statistics? *J R Stat Soc* 1995;**158**(part A):1–20.
2. Billingham LJ, Abrams KR, Jones DR. Methods for the analysis of quality-of-life and survival data in health technology assessment. *Health Technol Assess* 1999;**3**(10).
3. Altman DG. Practical statistics for medical research. London: Chapman and Hall, 1991.
4. Armitage P, Berry G. Statistical methods in medical research. Oxford: Blackwell, 1994.
5. Fisher LD, Van Belle G. Biostatistics-methodology for the health sciences. New York: Wiley, 1994.
6. Beaumont GP. Intermediate mathematical statistics. London: Chapman and Hall, 1980.
7. Cox DR, Hinkley DV. Theoretical statistics. London: Chapman and Hall, 1974.
8. Le CT. Fundamentals of biostatical inference. New York: Dekker, 1992.
9. Mood AM, Graybill FA, Boes DC. Introduction to the theory of statistics. Third edition. New York: McGraw-Hill International, 1974.
10. Larson HJ. Introduction to probability and statistical inference. Third edition. New York, Chichester: Wiley, 1982.
11. Hettmansperger TP. Statistical inference based on ranks. New York: Wiley, 1984.
12. Mead R. The design of experiments: statistical principles for practical application. Cambridge: Cambridge University Press, 1988.
13. Atkinson AC. Donev AN. Optimum experimental designs. Oxford: Clarendon Press, 1994.
14. Box GEP, Hunter JS, Hunter WG. Statistics for experiments. London: Chapman and Hall, 1978.
15. Jones B, Kenward MG. The design and analysis of cross-over trials. London: Chapman and Hall, 1989.
16. John JA, Williams E. Cyclic and computer generated designs. Second edition. London: Chapman and Hall, 1995.
17. Cochran WG. Cox GM. Experimental design. Second edition. New York, Chichester: Wiley, 1957.
18. Mead R, Curnow RN, Hasted AM. Statistical methods in agriculture and experimental biology. Second edition. London: Chapman and Hall, 1993.
19. Montgomery DC, Peck EA. Introduction to linear regression analysis. New York, Chichester: Wiley, 1982.
20. Aitken M, Anderson D, Francis B, Hinde N. Statistical modelling in GLIM. London: Oxford University Press, 1989.
21. Dobson AJ. An introduction to generalised linear models. London: Chapman and Hall, 1990.
22. Collett D. Modelling binary data. London: Chapman and Hall, 1991.
23. Draper NR, Smith H. Applied regression analysis. Second edition. New York, Chichester: Wiley, 1981.
24. Kleinbaum DG. Logistic regression: a self-learning text. New York, London: Springer-Verlag, 1994.
25. Kleinbaum D, Kupper LL, Muller RE. Applied regression analysis on other multivariate methods. Second edition. Boston MA: PWS-Kent Publications, 1988.
26. Collett D. Modelling survival data in medical research. London: Chapman and Hall, 1994.
27. Cox DR, Oakes D. Analysis of survival data. London: Chapman and Hall, 1984.
28. Kalbfleisch J, Prentice R. The statistical analysis of failure-time data. New York: Wiley, 1980.
29. Hand DF, Crowder M. Repeated measures analysis. London: Chapman and Hill, 1990.
30. Cressie NAC. Statistics for spatial data. New York: Wiley, 1991.
31. Diggle PJ. Statistical analysis of spatial point patterns. London: Academic Press, 1983.
32. Ripley B. Spatial statistics. New York, Chichester: Wiley, 1981.
33. Webster R, Oliver MA. Statistical methods in soil and land resource survey. Oxford: Oxford University Press, 1990.
34. Chatfield C, Collins AJ. Introduction to multivariate statistical analysis. London: Chapman and Hall, 1980.
35. Johnson RA, Wichern DW. Applied multivariate statistical analysis. London: Prentice Hall International, Inc., 1992
36. Krzanowski WJ. Principles of multivariate analysis: a user's perspective. Oxford: Clarendon Press, 1988.

37. McCullagh P, Nelder JA. generalized linear models. Second edition. London: Chapman and Hall, 1989.

38. Manly BFJ. Multivariate statistical methods – a primer. London: Chapman and Hall, 1986.

39. Mardia K, Kent J, Bibby J. Multivariate analysis. London: Academic Press, 1979.

40. Goldstein H. Multi-level statistical models. London: Edward Arnold, 1995.

41. Bryk AS, Raudenbush SW. Hierarchical linear models: applications and data analysis methods. Second edition. Newbury Park, CA: Sage, 1992.

42. Siegel S. Non-parametric statistics for the behavioural sciences. London: McGraw-Hill International, 1988.

43. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis. London: Chapman and Hill, 1995.

44. Fleiss JL. The design and analysis of clinical experiments. New York, Chichester, Wiley, 1985.

45. Strike PW. Statistical methods in laboratory medicine. Oxford: Butterworth Heinemann, 1991.

46. Thisted RA. Elements of statistical computing. London: Chapman and Hall, 1988.

47. Silverman BW. Density of estimation. London: Chapman and Hall, 1986.

48. Hastie TJ, Tibshirani RJ. Generalised additive models. London: Chapman and Hall, 1990.

49. Efron B, Tibshirani RJ. An introduction to the bootstrap. London: Chapman and Hall, 1993.

50. Pocock SJ. Clinical trials – a practical approach. Chichester: Wiley, 1983.

51. Whitehead J. The design and analysis of sequential clinical trials. Second edition. New York: Wiley, 1997.

52. Hennekens CH, Buring JE. Epidemiology in medicine. Boston, MA: Little and Brown, 1987.

53. Ashton J, editor. The epidemiological imagination. Buckingham: Open University Press, 1994.

54. Lilienfeld DE. Foundations of epidemiology. Oxford: Oxford University Press, 1994.

55. Stolley PD, Lasky T. Investigating disease patterns: the science and epidemiology. New York: Scientific American Library, 1995.

56. Kahn HA, Sempos CT. Statistical methods of epidemiology. Oxford: Oxford University Press, 1989.

57. Clayton D, Mills M. Statistical models in epidemiology. Oxford: Oxford University Press, 1993.

58. Schlesselman JJ. Case-control studies: design, conduct, analysis. Oxford: University Press, 1982.

59. Breslow NE, Day NE. Statistical methods of cancer research I: the analysis of case-control studies. Lyon: International Agency for Research on Cancer, 1980. Scientific Publication No 32.

60. Breslow NE, Day NE. Statistical methods in cancer research II: the design and analysis of cohort studies. Lyon, International Agency for Research on Cancer, 1987. Scientific Publication No 82.

61. American Food and Drug Administration. Guideline for the format and content of the clinical and statistical sections of new drug applications. Rockville, MD: FDA, 1988.

62. CPMP Working Party on Efficacy in Medical Products. Biostatistical methodology in clinical trials in applications for market for medicinal products. *Stat Med* 1995;**14**:1659–82.

63. International Conference on Harmonisation. Structure and content of clinical study reports. Rockville, MD: FDA, 1995.

64. Waller PC, Wood SM, Langman MJS, Breckenridge AM, Rawlins MD. Review of company postmarketing surveillance studies. *BMJ* 1992;**304**:1470–2.

65. Joint Committee of ABPI, BMA CSM and RCGP. Guidelines on postmarketing surveillance. *BMJ* 1988;**296**:399–400.

66. Joint Committee of MCA, ABPI BMA CSM and RCGP. Guidelines for company-sponsored safety assessment of marketed medicines. London: Medicines Control Agency, 1993.

67. PSI Professional Standards Working Party. Guideline standard operating procedures for good statistical practice in clinical research. 1991.

68. Medical Research Council. MRC Ethics Series. Macclesfield, London: Statisticians in the Pharmaceutical Industry, 1995.

69. Williamson PR, Hutton JL, Bliss J, Campbell M, Nicholson R. Statistical review by research ethics committee. *J R Stat Soc* 1999;**162**:482–90.

70. Gardner MJ, Altman DG. Statistics with confidence: confidence intervals and statistical guidelines. London: British Medical Journal, 1989.

71. Johnson T. Editorial: statistical guidelines for medical journals. *Stat Med* 1984;**3**:97–9.

72. Cochrane Handbook for Reviewers. <www.cochrane.org>.

73. Personal Bibliographic Software, Inc. ProCite for Windows, Version 3.0. 1995.

74. Senn S. The AB/BA crossover: past present and future? *Stat Method Med Res* 1994;**3**:303–24.

75. Kenward MG, Jones B. The analysis of binary and categorical data from crossover trials. *Stat Method Med Res* 1994;**3**:325–44.

76. Tudor G, Koch GG. Review of nonparametric methods for the analysis of crossover studies. *Stat Method Med Res* 1994;**3**:345–81.

77. Senn S. A personal view of some controversies in allocating treatment to patients in clinical trials. *Stat Med* 1995;**14**:2661–74.

78. Grender JM, Johnson WD. Fitting multivariate polynomial growth curves in two-period crossover designs. *Stat Med* 1994;**13**:931–43.

79. Brown HK, Kempton RA. The application of REML in clinical trials. *Stat Med* 1994;**13**:1601–17.

80. Grieve AP. Bayesian analyses of two-treatment crossover studies. *Stat Method Med Res* 1994;**3**:407–29.

81. Grieve AP. Extending a Bayesian analysis of the two-period crossover to accommodate missing data. *Biometrika* 1995;**82**:277–86.

82. Grieve AP. Extending a Bayesian analysis of the two-period crossover to allow for baseline measurements. *Stat Med* 1994;**13**:905–29.

83. Forster JJ. A Bayesian approach to the analysis of binary crossover data. *J R Stat Soc* 1994; **43**(series D):61–8.

84. Waclawiw MA, Liang K-Y. Empirical Bayes estimation and inference for the random effects model with binary response. *Stat Med* 1994;**13**:541–51.

85. Matthews JNS. Multi-period crossover trials. *Stat Method Med Res* 1994;**3**:383–405.

86. Carriere KC. Crossover designs for clinical trials. *Stat Med* 1994;**13**:1063–9.

87. Balagtas CC, Becker MP, Lang JB. Marginal modelling of categorical data from crossover experiments. *J R Stat Soc* 1995;**44**(series C):63–77.

88. Bellavance F, Tardif S. A nonparametric approach to the analysis of three-treatment three-period crossover designs. *Biometrika* 1995;**82**:865–75.

89. Liu J-P. Use of the repeated cross-over designs in assessing bioequivalence. *Stat Med* 1995; **14**:1067–78.

90. Vuorinen J, Tuominen J. Fieller's confidence intervals for the ratio of two means in the assessment of average bioequivalence from crossover data. *Stat Med* 1994;**13**:2531–45.

91. Liu J-P, Weng C-S. Bias of two one-sided tests procedures in assessment of bioequivalence. *Stat Med* 1995;**14**:853–61.

92. Toman B. Bayes optimal designs for two-and three-level factorial experiments. *J Am Stat Assoc* 1994; **89**:937–46.

93. Slud EV. Analysis of factorial survival experiments. *Biometrics* 1994;**50**:25–38.

94. Marden JI Muyot MET. Rank tests for main and interaction effects in analysis of variance. *J Am Stat Assoc* 1995;**90**:1388–98.

95. Alvo M, Cabilio P. Testing ordered alternatives in the presence of incomplete data. *J Am Stat Assoc* 1995;**90**:1015–24.

96. Giani G, Strasburger K. Testing and selecting for equivalence with respect to a control. *J Am Stat Assoc* 1994;**89**:320–9.

97. Ibrahim JG, Laud PW. A predictive approach to the analysis of designed experiments. *J Am Stat Assoc* 1994;**89**:309–19.

98. Gustafson P. A Bayesian analysis of bivariate survival data from a multicentre cancer clinical trial. *Stat Med* 1995;**14**:2523–35.

99. Stangl DK. Prediction and decision making using Bayesian hierarchical models. *Stat Med* 1995; **14**:2173–90.

100. Gray RJ. A Bayesian analysis of institutional effects in a multicenter cancer clinical trial. *Biometrics* 1994;**50**:244–53.

101. Raghunathan TE. Monte Carlo methods for exploring sensitivity to distributional assumptions in a Bayesian analysis of a series of 2 x 2 tables. *Stat Med* 1994;**13**:1525–38.

102. Hirji KF, Tang M-L, Vollset SE, Elashoff RM. Efficient power computation for exact and mid-P tests for the common odds ratio in several 2 x 2 tables. *Stat Med* 1994;**13**:1539–49.

103. O'Gorman TW, Woolson RF, Jones MP. A comparison of two methods of estimating a common risk difference in a stratified analysis of a multicenter trial. *Controlled Clin Trials* 1994;**15**:135–53.

104. Brunner E, Puri ML, Sun S. Nonparametric methods for stratified two-sample designs with application to multi-center clinic trials. *J Am Stat Assoc* 1995;**90**:1004–14.

105. Kenward MG, Lesaffre E, Molenberghs G. An application of maximum likelihood and generalised estimating equations to the analysis of ordinal data from a longitudinal study with cases missing at random. *Biometrics* 1994;**50**:945–53.

106. Davis CS, Chung Y. Randomization model methods for evaluating treatment efficacy in multicenter clinical trials. *Biometrics* 1995;**51**:1163–74.

107. Nam J-M. Sample size determination in stratified trials to establish the equivalence of two treatments. *Stat Med* 1995;**14**:2037–49.

106. Roebruck P, Kuhn A. Comparison of tests and sample size formulae for proving therapeutic equivalence based on the difference of binomial probabilities. *Stat Med* 1995;**14**:1583–94.

109. Bofinger E, Bofinger M. Equivalence with respect to a control: stepwise tests. *J R Stat Soc* 1995; **57**(series B):721–33.

110. Yamagousa T, Tango T, Hiejima Y. Mantel-Haenszel type tests for testing equivalence or more than equivalence in comparative clinical trials. *Biometrics* 1994;**50**:859–64.

111. Ng T-H. Conventional null hypothesis testing in active control equivalence trials. *Controlled Clin Trials* 1995;**16**:356–8.

112. Spiegelhalter DJ, Freedman LS, Parmar MK. Bayesian approaches to randomized trials. *J R Stat Soc* 1994;**157**(series A):357–416.

113. Stevenson CE. Statistical models for cancer screening. *Stat Method Med Res* 1995;**4**:18–32.

114. Etzioni RD, Connor RJ, Prorok PC, Self SG. Design and analysis of cancer screening trials. *Stat Method Med Res* 1995;**4**:3–17.

115. O'Neill TJ, Tallis GM, Leppard P. A review of the technical features of breast cancer screening illustrated by a specific model using South Australian cancer registry data. *Stat Method Med Res* 1995;**4**:55–72.

116. Van Oortmarssen GJ, Boer R, Habbema JDF. Modelling issues in cancer screening. *Stat Method Med Res* 1995;**4**:33–54.

117. Duffy SW, Chen H-H, Tabar L, Day NE. Estimation of mean sojourn time in breast cancer screening using a Markov chain model of both entry to and exit from the preclinical detectable phase. *Stat Med* 1995;**14**:1531–43.

118. Connor RJ, Prorok PC. Issues in the mortality analysis of randomised controlled trials of cancer screening. *Controlled Clin Trials* 1994;**15**:81–99.

119. Etzioni R, Self SG. On the catch-up time method for analyzing cancer screening trials. *Biometrics* 1995;**51**:31–43.

120. Kafadar K, Prorok PC. A data-analytic approach for estimating lead time and screening benefit based on survival curves in randomized cancer screening trials. *Stat Med* 1994;**13**:569–86.

121. Stukel TA, Glynn RJ, Fisher ES, Sharp SM. Standardized rates of recurrent outcomes. *Stat Med* 1994;**13**:1781–91.

122. Xu J-L, Prorok PC. Non-parametric estimation of the post-lead-time survival distribution of screen-detected cancer cases. *Stat Med* 1995;**14**:2715–25.

123. Self SG, Etzioni R. A likelihood ratio test for cancer screening trials. *Biometrics* 1995;**51**:44–50.

124. Thomson PC. A hybrid paired and unpaired analysis for the comparison of proportions. *Stat Med* 1995;**14**:1463–70.

125. Hughes MD, Thompson SG, Pocock SJ. Optimal sequential screening guidelines for quantitative risk factors measured with error. *J Am Stat Assoc* 1995; **90**:19–26.

126. Green SB, Freedman LS. Early stopping of prevention trials when multiple outcomes are of interest: a discussion. *Stat Med* 1994;**13**:1479–84.

127. Fleming TR, Prentice RL, Pepe MS, Glidden D. Surrogate and auxiliary endpoints in clinical trials with potential applications in cancer and AIDS research. *Stat Med* 1994;**13**:955–68.

128. Wade AM, Ades AE, Salt AT, Jayatunga R. Age-related standards for ordinal data: modelling the changes in visual acuity from 2 to 9 years of age. *Stat Med* 1995;**14**:257–66.

129. Gastwirth JL, Johnson WO. Screening with cost-effective quality control: potential applications to HIV and drug testing. *J Am Stat Assoc* 1994;**89**:972–81.

130. Hughes-Oliver JM, Swallow WH. A two-stage adaptive group-testing procedure for estimating small proportions. *J Am Stat Assoc* 1994;**89**:982–93.

131. Litvak E, Tu XM, Pagano M. Screening for the presence of a disease by pooling sera samples. *J Am Stat Assoc* 1994;**89**:424–34.

132. Phatarfod RM, Sudbury A. The use of a square array scheme in blood testing. *Stat Med* 1994;**13**:2337–44.

133. Tu XM, Litvak E, Pagano M. Screening tests: can we get more by doing less? *Stat Med* 1994;**13**:1905–19.

134. Tu XM, Litvak E, Pagano M. On the informativeness and accuracy of pooled testing in estimating prevalence of a rare disease: application to HIV screening. *Biometrika* 1995;**82**:287–97.

135. Waller JL, Addy CL, Jackson KL, Garrison CZ. Confidence intervals for weighted proportions. *Stat Med* 1994;**13**:1071–82.

136. Shannon WD, Bryant J, Logan TF, Day R. An application of decision theory to patient screening for an autologous tumour vaccine trial. *Stat Med* 1995;**14**:2099–110.

137. Enas G, Goldstein DJ. Defining monitoring and combining safety information in clinical trials. *Stat Med* 1995;**14**:1099–111.

138. O'Neill RT. Statistical concepts in the planning and evaluation of drug safety from clinical trials in drug development: issues of international harmonization. *Stat Med* 1995;**14**:1117–27.

139. Lancar R, Kramar A, Haie-Meder C. Non-parametric methods for analysing recurrent complications of varying severity. *Stat Med* 1995;**14**:2701–12.

140. Farrington CP. Relative incidence estimation from case series for vaccine safety evaluation. *Biometrics* 1995;**51**:228–35.

141. Cook RJ, Farewell VT. Guidelines for monitoring efficacy and toxicity responses in clinical trials. *Biometrics* 1994;**50**:146–52.

142. Chueng-Stein C. A new proposal for benefit-loss risk analysis in clinical trials. *Controlled Clin Trials* 1994;**15**:30–43.

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

144. Ashby D, White SJ, Brown PJ. Statistical methods: good practice and identifying opportunities for innovation. In: Black N, Brazier J, Fitzpatrick R, Reeves B, editors. Health services research methods – a guide to best practice. London: BMJ Books, 1998:141–8.

145. Zahl PH. A proportional regression model for 20 year survival of colon cancer in Norway. *Stat Med* 1995;**14**:1249–61.

146. Satten GA, Longini I Jr. Markov chains with measurement error: estimating the 'true' course of a marker of the progression of human immuno-deficiency virus disease. *Appl Stat* 1996;**45**:275–309.

147. Grubard BI, Korn EL. Regression analysis with clustered data. *Stat Med* 1994;**13**:509–22.

148. Raab GM, Gore SM, Goldberg DJ, Donnelly CA. Bayesian forecasting of the human immuno-deficiency virus epidemic in Scotland. *Stat Soc* 1994;**157**:17–30.

149. Everitt BS. The analysis of repeated measures: a practical review with examples. *J R Stat Soc* 1995;**44**(series D):113–35.

150. Fitzmaurice GM, Molenberges G, Lipsitz SR. Regression models for longitudinal binary responses with informative drop-outs. *J R Stat Soc* 1995;**57**(series B):691–704.

151. Ten Have TR, Landis JR, Weaver SL. Association models for peridontal disease progression: a comparison of methods for clustered binary data. *Stat Med* 1995;**14**:413–29.

152. Wang-Clow F, Lange M, Laird NM, Ware JH. A simulation study of estimators for rates of change in longitudinal studies with attrition. *Stat Med* 1995;**14**:283–97.

153. Wu MC, Hunsberger S, Zucker D. Testing for differences in changes in the presence of censoring: parametric and non-parametric methods. *Stat Med* 1994;**13**:635–46.

154. Lawless JF. The analysis of recurrent events for multiple subjects. *J R Stat Soc* 1995; **44**(series C):487–98.

155. Lipsitz SR, Fitzmaurice GM, Orav EJ, Laird NM. Performance of generalized estimating equations in practical situations. *Biometrics* 1994;**50**:270–8.

156. Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcome in the presence of missing data. *J Am Stat Assoc* 1995;**90**:106–21.

157. Williamson JM, Kim K, Lipsitz SR. Analyzing bivariate ordinal data using a global odds ratio. *J Am Stat Assoc* 1995;**90**:1432–7.

158. Lipsitz SR, Kim K, Zhao L. Analysis of repeated categorical data using generalized estimating equations. *Stat Med* 1994;**13**:1149–63.

159. Robins JM, Rotnitzky A. Semiparametric efficiency in multivariate regression models with missing data. *JASA* 1995;**90**:122–9.

160. Miller ME. Analysing categorical responses obtained from large clusters. *J R Stat Soc* 1995;**44**(series C):173–86.

161. Fitzmaurice GM, Lipsitz SR. A model for binary time series data with serial odds ratio patterns. *J R Stat Soc* 1995;**44**(series C):51–61.

162. Follmann D. Modelling transitional and joint marginal distributions in repeated categorical data. *Stat Med* 1994;**13**:467–77.

163. Mark SD, Gail MH. A comparison of likelihood-based and marginal estimating equation methods for analysing repeated ordered categorical responses with missing data: application to an intervention trial of vitamin prophylaxis for oesophageal dysplasia. *Stat Med* 1994;**13**:479–93.

164. Tsiatis AA, DeGruttola V, Wulfsohn MS. Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. *J Am Stat Assoc* 1995;**90**:27–37.

165. Goldstein H, Healy MJR, Rasbash J. Multilevel time series models with applications to repeated measures data. *Stat Med* 1994;**13**:1643–55.

166. Pearson JD, Morrell CH, Landis PK, Carter HB. Mixed-effects regression models for studying the natural history of prostate disease. *Stat Med* 1994;**13**:587–601.

167. Follmann D, Wu M. An approximate generalized linear model with random effects for informative missing data. *Biometrics* 1995;**51**:151–68.

168. Morrell CH, Pearson JD, Carter HB, Brant LJ. Estimating unknown transition times using a piecewise nonlinear mixed-effects model in men with prostate cancer. *J Am Stat Assoc* 1995;**90**:45–53.

169. Palta M, Yao T-J, Velu R. Testing for omitted variables and non-linearity in regression models for longitudinal data. *Stat Med* 1994;**13**:2219–31.

170. Zerbe GO, Wu MC, Zucker DM. Studying the relationship between change and initial value in longitudinal studies. *Stat Med* 1994;**13**:759–68.

171. Anderson SJ, Jones RH. Smoothing splines for longitudinal data. *Stat Med* 1995;**14**:1235–48.

172. Gilula Z, Haberman SJ. Conditional log-liner models for analyzing categorical panel data. *J Am Stat Assoc* 1994;**89**:645–56.

173. Qu Y, Piedmonte MR, Medendorp SV. Latent variable models for clustered ordinal data. *Biometrics* 1995;**51**:268–75.

174. Lindstrom MJ. Self-modelling with random shift and scale parameters and a free-knot spline shape function. *Stat Med* 1995;**14**:2009–21.

175. Mori M, Woolson RF, Woodworth GG. Slope estimation in the presence of informative right censoring: modeling the number of observations as a geometric random variable. *Biometrics* 1994;**50**:39–50.

176. Donnelly CA, Laird NM, Ware JH. Prediction and creation of smooth curves for temporally correlated longitudinal data. *J Am Stat Assoc* 1995;**90**:984–9.

177. Cole TJ. Growth charts for both cross-sectional and longitudinal data. *Stat Med* 1994;**13**:2477–92.

178. Wang Y, Taylor JMG. Inference for smooth curves in longitudinal data with application to an AIDS clinical trial. *Stat Med* 1995;**14**:1205–18.

179. Dawson JD. Stratification of summary statistic tests according to missing data patterns. *Stat Med* 1994;**13**:1853–63.

180. Lindsey JK. Fitting parametric counting processes by using log-linear models. *J R Stat Soc* 1995;**44**(series C):201–12.

181. Diggle P, Kenward MG. Informative drop-out in longitudinal data analysis. *J R Stat Soc* 1994; **43**(series C):49–93.

182. Little RJA. Modeling the drop-out mechanism in repeated-measures studies. *J Am Stat Assoc* 1995;**90**:1112–21.

183. Wu MC, Lan GKK, Connett JE. Use of surrogate information time for monitoring the effect of treatment on the change in a response variable in clinical trials. *Stat Med* 1994;**13**:945–53.

184. Lavori PW, Dawson R, Mueller TB. Causal estimation of time-varying treatment effects in observational studies: application to depressive disorder. *Stat Med* 1994;**13**:1089–100.

185. Lavori PW, Dawson R, Shera D. A multiple imputation strategy for clinical trials with truncation of patient data. *Stat Med* 1995; **14**:1913–25.

186. Cole BF, Lee M-LT, Whitmore A, Zaslavsky AM. An empirical Bayes model for Markov-dependent binary sequences with randomly missing observations. *J Am Stat Assoc* 1995;**90**:1364–72.

187. Kim HM, Lagakos SW. Assessing drug compliance using longitudinal marker data with application to AIDS. *Stat Med* 1994;**13**:2141–53.

188. Ten Have TR, Chinchilli VM. Bayesian hierarchical analysis of within-units variances in repeated measures experiments. *Stat Med* 1994;**13**:1841–52.

189. Lipsitz SR, Fitzmaurice GM. Sample size for repeated measures studies with binary responses. *Stat Med* 1994;**13**:1233–9.

190. Overall JE, Doyle SR. Evaluating sample size for repeated measures design. *Controlled Clin Trials* 1994;**15**:100–23.

191. Kirby AJ, Galou N, Minez A. Sample size estimation using repeated measurements on biomarkers as outcomes. *Controlled Clin Trials* 1994;**15**:165–172.

192. Silverberg AR. A simulation study comparing two approximations for a quasi t-quantile used in repeated measures ANOVA. *Stat Med* 1994;**13**:1657–63.

193. Dinse GE. A comparison of tumour incidence analyses applicable in single-sacrifice animal experiments. *Stat Med* 1994;**13**:689–708.

194. Chinchilli VM, Esinhart JD, Miller WG. Partial likelihood analysis of within-unit variances in repeated measurement experiments. *Biometrics* 1995;**51**:205–16.

195. Lin H-M, Hughes MD. Use of historical marker data for assessing treatment effects in Phase I/II trials when subject selection is determined by baseline marker level. *Biometrics* 1995;**51**:1053–63.

196. Hayter AJ, Hsu JC. On the relationship between stepwise decision procedures and confidence sets. *J Am Stat Assoc* 1994;**89**:128–36.

197. Namgung YY, Yang MCK. Outlier reduction by an option-3 measurement scheme. *Biometrics* 1994;**50**:173–82.

198. McCloud PI, Darroch JN. An analysis of categorical repeated measurements in the presence of an external factor. *Biometrika* 1995;**82**:47–55.

199. Ukoumunne OC, Gulliford MC, Chinn S, Sterne JAC, Burney PGJ. Methods for evaluating area-wide and organisation-based interventions in health and health care: a systematic review. *Health Technol Assess* 1999;**3**(5).

200. Lindley DV. Making decisions. Second edition. Chichester: Wiley, 1985.

201. Smith JQ. Decision analysis: a Bayesian approach. London: Chapman and Hall, 1988.

202. Bernardo JM, Smith AFM. Bayesian theory. Chichester: Wiley, 1994.

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

* Previous Chair

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