Methods for the analysis of quality-of-life and survival data in health technology assessment

LJ Billingham
KR Abrams
DR Jones
Standing Group on Health Technology

Current members

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

Professor Martin Buxton, Professor of Economics, Brunel University
Professor Francis Creed, School of Psychiatry & Behavioural Sciences, University of Manchester
Professor Charles Florey, Department of Epidemiology & Public Health, Ninewells Hospital & Medical School, University of Dundee

Professor John Gabbay, Director, Wessex Institute for Health Research & Development
Professor Sir John Grimley Evans, Department of Geriatric Medicine, Radcliffe Infirmary, Oxford
Dr Tony Hope, The Medical School, University of Oxford
Professor Richard Lilford, Regional Director, R&D, 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, Director, NHS Centre for Reviews & Dissemination, University of York
Professor Mike Smith, Director, The Research School of Medicine, University of Leeds
Dr John Tripp, Department of Child Health, Royal Devon & Exeter Healthcare NHS Trust

Professor Howard Glennister, 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 Tom Walley, Department of Pharmacological Therapeutics, University of Liverpool
Dr Julie Woodin, Chief Executive, Nottingham Health Authority
Professor Kent Woods (Chair Designate), Regional Director of R&D, NHS Executive, Trent

Past members

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 Howard Glennister, 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

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.ncchta.org). 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, York Publishing Services.

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, York Publishing Services by:

– 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:
York Publishing Services
PO Box 642
YORK YO31 7WX
UK
Email: ncchta@yps-publishing.co.uk
Tel: 0870 1616662
Fax: 0870 1616663
Fax from outside the UK: +44 1904 430868

NHS and 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 contact York Publishing Services at the address above. 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 York Publishing Distribution 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.ncchta.org/htacd.htm). Or contact York Publishing Services (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.
Methods for the analysis of quality-of-life and survival data in health technology assessment

LJ Billingham\textsuperscript{1,2}
KR Abrams\textsuperscript{1}
DR Jones\textsuperscript{1}

\textsuperscript{1} Department of Epidemiology and Public Health, University of Leicester, UK
\textsuperscript{2} Cancer Research Campaign Trials Unit, University of Birmingham, UK

Published July 1999

This report should be referenced as follows:


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).
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 Panel and funded as project number 93/50/04.

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.

Series Editors: Andrew Stevens, Ruairidh Milne and Ken Stein
Editorial Assistant: 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.
Contents

List of abbreviations ............................................. i
Executive summary ............................................. iii

1 Introduction .................................................. 1
   Aims ......................................................... 1
   Rationale for the study ................................... 1
   Introduction to this report ............................... 1
   References .................................................. 2

2 Literature search ............................................. 3
   Introduction ............................................... 3
   Search methodology ..................................... 3
   Results of the search ................................... 5
   Further potential searching ............................. 5
   Conclusions ............................................... 6

3 Background to quality-of-life and survival data ............................................. 7
   The meaning of quality of life .......................... 7
   Measuring quality of life ................................ 7
   Quality-of-life data ...................................... 8
   Survival data ............................................. 11
   References .................................................. 11

4 Background to the illustrative example – the MIC study .................................. 13
   Introduction ................................................. 13
   Background ............................................... 13
   Measuring quality of life ................................ 13
   Missing data in the quality-of-life study ............ 15
   References .................................................. 16

5 Quality-of-life analysis in the presence of informative drop-out: introduction ...... 19

6 Descriptive methods for the analysis of longitudinal quality-of-life data .......... 21
   Introduction ................................................ 21
   Patient profiles ............................................ 21
   Group profiles ............................................ 23
   References .................................................. 27

7 Standard methods for the analysis of longitudinal data .................................... 29
   Introduction ............................................... 29
   Analysis of each time point separately .......... 29
   Summary measures ....................................... 29
   Repeated measures analysis of variance ...... 31
   More complex modelling techniques for longitudinal data .................................. 31
   References .................................................. 32

8 Longitudinal modelling methods that deal with informative drop-out ............... 35
   Introduction ............................................... 35
   Methods ..................................................... 35
   Discussion .................................................. 36
   References .................................................. 36

9 Quality-of-life analysis in the presence of informative drop-out: summary and discussion ............................................. 37
   References .................................................. 37

10 Analysis of survival data adjusting for quality of life: introduction ................. 39

11 Comparison of treatments in terms of survival ............................................. 41
   Introduction to survival analysis ..................... 41
   Survivor and hazard functions ....................... 41
   Comparison of treatments using Kaplan–Meier estimates ................................... 42
   Non-parametric tests for comparing survival ............................................. 43
   References .................................................. 44

12 Comparison of treatments in terms of survival adjusting for covariates ........... 45
   Introduction ............................................... 45
   Stratified survival analysis ......................... 45
   Cox proportional hazards model with fixed covariates ................................... 46
   Cox model with time-dependent covariates .......... 47
   References .................................................. 49

13 Analysis of survival data adjusting for quality of life: summary and discussion ... 51

14 Simultaneous analysis of quality-of-life and survival data: introduction ........... 53

15 Quality-adjusted life years .................................. 55
   Introduction ............................................... 55
   Utilities ..................................................... 55
   QALY methodology ...................................... 57
   Example: calculating QALYs in the MIC study ............................................. 58
   Application of QALYs .................................... 60
   Critical appraisal of QALYs .......................... 61
   Alternative and related methods to QALYs .......... 62
   References .................................................. 63
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>Quality-adjusted survival analysis</td>
<td>71</td>
</tr>
<tr>
<td>17</td>
<td>Multistate survival analysis</td>
<td>85</td>
</tr>
<tr>
<td>18</td>
<td>Incorporating results from a multistate model into a Q-TWiST analysis</td>
<td>97</td>
</tr>
<tr>
<td>19</td>
<td>Modelling quality of life and survival as two simultaneous processes</td>
<td>99</td>
</tr>
<tr>
<td>20</td>
<td>Simultaneous analysis of quality-of-life and survival data: summary and discussion</td>
<td>101</td>
</tr>
<tr>
<td>21</td>
<td>Conclusions and recommendations</td>
<td>103</td>
</tr>
<tr>
<td></td>
<td>Overall summary and discussion</td>
<td>103</td>
</tr>
<tr>
<td></td>
<td>Recommendations</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>106</td>
</tr>
<tr>
<td></td>
<td>Acknowledgements</td>
<td>107</td>
</tr>
<tr>
<td></td>
<td>Bibliography</td>
<td>109</td>
</tr>
<tr>
<td></td>
<td>Appendix 1 Details of search terms used in electronic database searching</td>
<td>145</td>
</tr>
<tr>
<td></td>
<td>Health Technology Assessment reports published to date</td>
<td>147</td>
</tr>
<tr>
<td></td>
<td>Health Technology Assessment panel membership</td>
<td>149</td>
</tr>
</tbody>
</table>
## List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIDS</td>
<td>Bath Information and Data Service</td>
</tr>
<tr>
<td>CI</td>
<td>confidence interval</td>
</tr>
<tr>
<td>CT</td>
<td>chemotherapy arm (MIC study)</td>
</tr>
<tr>
<td>EM</td>
<td>an algorithm consisting of an expectation step and a maximisation step</td>
</tr>
<tr>
<td>EORTC</td>
<td>European Organisation for Research and Treatment of Cancer</td>
</tr>
<tr>
<td>HYE</td>
<td>healthy-years equivalent</td>
</tr>
<tr>
<td>MAL</td>
<td>malaise variable (MIC study)</td>
</tr>
<tr>
<td>MIC</td>
<td>mitomycin, ifosfamide and cisplatin</td>
</tr>
<tr>
<td>MQS</td>
<td>mean quality-of-life score</td>
</tr>
<tr>
<td>PAL</td>
<td>palliative treatment arm (MIC study)</td>
</tr>
<tr>
<td>QAL</td>
<td>quality-adjusted life</td>
</tr>
<tr>
<td>QALY</td>
<td>quality-adjusted life year</td>
</tr>
<tr>
<td>QoL</td>
<td>quality of life*</td>
</tr>
<tr>
<td>Q-TWiST</td>
<td>quality-adjusted TWiST</td>
</tr>
<tr>
<td>REL</td>
<td>time spent with symptoms of disease following disease relapse</td>
</tr>
<tr>
<td>SAVE</td>
<td>saved young life equivalent</td>
</tr>
<tr>
<td>TOX</td>
<td>time spent with toxicity from treatment</td>
</tr>
<tr>
<td>TR</td>
<td>time to disease relapse</td>
</tr>
<tr>
<td>TTO</td>
<td>time trade-off</td>
</tr>
<tr>
<td>TWiST</td>
<td>time spent without symptoms of disease and toxicity of treatment</td>
</tr>
<tr>
<td>TWiST(L)</td>
<td>accumulated TWiST</td>
</tr>
</tbody>
</table>

*Used only in tables and figures
Objectives

Quality of life has become an important issue in health care, especially in studies of chronic diseases. Substantial amounts of quality-of-life data are now being gathered in clinical trials, using a variety of instruments. In longitudinal studies of quality of life in which survival is also an endpoint, patients are generally severely ill and it is common for participants to drop out of the study because of illness or death. In such situations, the drop-out process may depend on the quality of life being experienced, rather than being random; hence the incomplete follow-up of patients is called informative drop-out. This must be appropriately accounted for in any analysis of the data to avoid the introduction of bias.

This study identifies and reviews critically the methods proposed for the analysis of quality-of-life and survival data in health technology assessment, particularly those that assess both these endpoints simultaneously. In this way methodology that requires wider dissemination can be identified together with areas requiring further research. It was not within the remit of this study to address issues related to the meaning, definition and measurement of quality of life.

Methods

The scientific and medical literature was searched for relevant methodological articles. Electronic searches were carried out systematically using Science Citation Index, Social Science Citation Index and the EMBASE database provided by BIDS (Bath Information and Data Service). The searches were supplemented by exploded references, personal collections and handsearching of the journal Quality of Life Research.

Results

Methods for analysing quality-of-life and survival data were found to fall into three broad categories, as described below, according to the research question underlying the study; this in turn depends on the disease and treatments under investigation.

Quality-of-life analysis in the presence of informative drop-out

The use of standard methods for the analysis of longitudinal data is discussed in terms of their application to longitudinal quality-of-life data. All methods, from simple descriptive analysis to complex modelling techniques, will give biased results when informative drop-out is present in the data. Standard methods should therefore be used with caution when analysing longitudinal quality-of-life data. Modelling techniques that deal with informative drop-out have been developed and their application to quality-of-life data is discussed.

Analysis of survival data adjusting for quality of life

In comparing treatments in terms of survival, it is often necessary to adjust for other patient-related factors, known as covariates, that could potentially affect the survival time of a patient. In some situations the survival analysis may need to adjust for baseline measures of quality of life (fixed covariates), while in others, allowance for changing quality of life over time may be required (time-dependent covariates). If assessments of quality of life are infrequent or data are missing for reasons other than death, then it may be difficult to adjust for changes in quality of life with any degree of accuracy. Modelling quality of life and survival as two simultaneous processes may improve the analysis in this situation.

Simultaneous analysis of quality-of-life and survival data

In studies in which quality of life and survival are both important endpoints, it may be advantageous to assess health technologies in terms of these endpoints simultaneously. Three different approaches can be used to achieve this:

• combining quality and quantity of life into a single endpoint and using quality-adjusted survival analysis methods to compare treatments
• using multistate models to model the movement of patients between various health states, defined by levels of quality of life and by death, and exploring how treatments differ in terms of these movements
• considering quality of life and survival as two simultaneous processes and describing the data in terms of two interlinked models.

Quality and quantity of life can be combined into a single endpoint by weighting periods of survival time according to the quality of life experienced. The resulting outcome measures are generally referred to as QALYs (quality-adjusted life years) with special forms known as TWiST (time spent without symptoms of disease and toxicity of treatment) and Q-TWiST (quality-adjusted TWiST). The use of standard survival analysis techniques on the QALY endpoint will generally give biased results because individuals with a worse quality of life will be censored earlier than those with a good quality of life, resulting in informative censoring. Methods of overcoming this problem, including partitioned survival analysis, are discussed. Quality-adjusted survival analysis overcomes problems of informative drop-out due to death and has the potential to be extended to deal with other disease- or treatment-related reasons for drop-out.

Multistate models are defined by categorising the period of follow-up of patients in a trial into a number of different health states defined in terms of levels of quality of life and death. The movement between health states is described by transition rates, which are modelled using the transition times for patients. Various modelling approaches are discussed. The inclusion of death as a health state in the model enables the analysis to deal with informative drop-out due to death and the inclusion of a ‘drop-out’ state could cover other reasons.

The most recently developed, and potentially most powerful, approach to analysing quality-of-life and survival data is to model the longitudinal quality-of-life data and the drop-out process, which includes drop-out due to death, as two simultaneous processes. Such an approach has the advantage of allowing quality-of-life data to be assessed longitudinally while adjusting for informative drop-out. In addition, the inter-relationship between the two can be explored.

Conclusions and recommendations

Obtaining appropriate data
• The method of analysis needs to be decided at the design stage of a study so that appropriate quality-of-life data can be collected. Issues to consider are:
  – the quality-of-life instrument to be used
  – the frequency and timing of quality-of-life assessments
  – the need to minimise non-compliance
  – the collection of additional information, such as reason for drop-out
  – the sample size required.

Choosing the appropriate method
• The choice of method should be based on the research question that the study aims to answer. The advantages and disadvantages of each method should be considered carefully together with the relevance and interpretability of the results to clinicians and patients.
• Methods used to analyse longitudinal quality-of-life data must allow for informative drop-out.

Reporting the analysis
• Methods used should be reported clearly, with details of definitions and assumptions used in the analysis.
• A sensitivity analysis should be carried out to assess the robustness of conclusions to any critical assumptions made in the analysis.

Recommendations for further research
• Further experience in the application of quality-adjusted survival analysis techniques to quality-of-life data is needed to enable a proper evaluation of such methods.
• Further research is needed in order to develop hierarchical models, multistate models and simultaneous modelling methods in their practical application to quality-of-life and survival data using both classical and Bayesian approaches. Consideration should be given to how methods could deal with the multivariate nature of the quality-of-life endpoint.
• A full review of available software for methods that simultaneously analyse quality-of-life and survival data is needed to highlight areas requiring further development.
• Progress in the most rapidly developing areas of multistate survival analysis and simultaneous modelling should be monitored, together with parallel areas of methodological development such as in the field of AIDS research.
Chapter 1

Introduction

Aims

The methods that have been proposed for the analysis of quality of life and survival data in health technology assessment are reviewed, with particular reference to those that assess these two endpoints simultaneously. The key objectives were to:

- identify proposed methodology for the simultaneous analysis of quality and length of life
- review critically the proposed methodology
- illustrate such methodology where possible by application to data from a previously conducted study
- identify on the basis of the review:
  - methodology that requires dissemination within the health research community and the NHS
  - areas of non-existent or deficient methodology
  - areas that require further research.

Rationale for the study

Quality-of-life assessment has become an important issue in healthcare research and the study of new technologies, especially in many chronic diseases. Although survival is usually the standard endpoint for assessing treatments in clinical trials, informed clinical decisions generally require quantification and interpretation of quality-of-life measures, especially with respect to variation over time and the interrelationship with length of life. The role of quality of life will become even more prominent in the future as, for many diseases, improvements in survival due to treatment are either unlikely to be dramatic or likely to be made at the expense of quality of life.

There has been much research into the development of instruments with which to measure quality of life. This has resulted in a plethora of instruments and substantial amounts of quality-of-life data being gathered in trials. Hence there is a need for methods that enable the effective and efficient analysis and interpretation of such data. It is therefore timely that the range of statistical methods which have been proposed for dealing with such data are systematically reviewed and evaluated. There are a number of previous reviews of the analysis of quality of life in clinical trials.1–7

In longitudinal studies of quality of life in which survival is also an endpoint, the patient population will not be stable over time. Patients are generally severely ill and individuals may have incomplete follow-up of quality of life for reasons related to disease or treatment, including death. This dropout process may be informative and can make statistical analysis particularly problematic. There is, therefore, a need to identify appropriate methods which will yield unbiased and clinically relevant assessment of health technologies in such situations.

This study will therefore yield two benefits.

- The identification of existing methodology that has been shown to benefit the assessment of health technologies with respect to both quality and length of life, and thus enable more informed decision-making within the NHS.
- The identification of those areas in which further work is required so that existing methods can either be more appropriately applied or be realistically developed.

Introduction to this report

Coverage

The first four chapters of this report provide an introduction and background. They include details of the methodology used to carry out the review and they provide a background to both types of data encountered in the report – quality-of-life data and survival data. The early chapters also provide a background to the clinical trial from which quality-of-life and survival data were taken to illustrate some of the proposed methodology.

Health technology assessment is the evaluation of any intervention intended to improve health. The investigation of quality of life and survival most often takes place in the context of a randomised clinical trial that compares treatments and the report focuses on this scenario. The main body of the report, in which the proposed methods for
analysing quality-of-life and survival data are discussed, is in three parts according to the definition of the research question of interest. This research question depends on the disease and treatments under investigation.

The first approach, discussed in chapters 5–9, covers studies in which the primary aim is to compare health technologies in terms of quality of life. Survival is also considered but only in terms of the problematical impact it has on the quality-of-life assessment data. Deaths of patients during the course of a study may result in informative drop-out, causing quality-of-life data to be missing. The problems for analysis caused by informative drop-out are discussed.

The second approach, discussed in chapters 10–13, covers studies in which the primary aim is to compare health technologies in terms of survival. Quality of life may also be measured and any survival analysis may need to be adjusted for this.

The third approach, discussed in chapters 14–20, includes studies in which quality of life and survival are both important endpoints for assessing health technologies. Instead of analysing each endpoint separately, it may be more appropriate to analyse both endpoints simultaneously. The focus of the report is on this type of methodology.

In the final chapter of the report the methods discussed are summarised and the implications for the design and conduct of health technology assessment research are considered. Methodology that requires wider dissemination is identified and recommendations for further research are made.

**Format of the report and guidance for readers**

The report can be approached in a variety of ways depending on the needs of the individual reader. Although the report is cross-referenced, the section addressing each approach can be read as a single entity, thus enabling the reader to focus on just that part of the report that relates to the particular problem that they wish to address.

Technical detail has been kept to a minimum in order to make the report accessible to a wide range of readers. For clarification, most key methods are illustrated with worked examples, using data from a real study. However, the analysis presented here is purely to illustrate the methodology and should not be interpreted as a report of the results of this study; these will be presented elsewhere.

It was not within the remit of this study to address issues relating to the meaning, definition and measurement of quality of life. Use of the term, quality of life, in the literature has thus been accepted uncritically. The term, quality of life, is used here to mean anything that purports to measure health-related quality of life or some aspect of it. This does not affect any discussion of the methodological problems of analysing such data when it is assessed longitudinally in conjunction with survival data.

**References**


Chapter 2

Literature search

Introduction

The literature search forms the basis of the review and in this chapter details are presented of the methodology used in the search. The aim was to identify all statistical methodology that has been proposed for the analysis of data for health technology assessment with respect to both quality and length of life. The focus was on methods that analyse quality-of-life data over time in situations in which length of survival is an issue and, in particular, methods that simultaneously analyse quality of life and survival.

Search methodology

Introduction

A systematic and repeatable strategy was devised to search the literature for relevant articles. A complete list of all papers citing relevant statistical methods was not necessary, since enumeration of usage was not an objective. However, the search did seek to identify the full range of methods and so was kept as broad and as thorough as possible within the time constraints of the study.

Statistical methodology is represented, not only in theoretical but also in applied literature. The literature of greatest interest therefore was essentially of a statistical or medical nature. The search, however, was not limited to these areas since methods found in other fields, such as industry, agriculture and education, could be adapted to be applicable in a health context. Journal articles comprised the main body of literature but all types, including books, reports, conference papers and theses, were considered for inclusion.

The search strategy consisted of a variety of different approaches. The main part was undertaken using electronic databases but a considerable number of references were also obtained by other methods, such as handsearching journals, exploding references, and personal recommendations. When identified, references were stored and managed on a database using Reference Manager bibliographic software v. 7 (Research Information Systems, USA, 1995).

Electronic database searching

A variety of electronic databases provided by Bath Information and Data Service (BIDS) were used to search the scientific literature. BIDS Science Citation Index provides access to over 4400 journals in natural, physical and biomedical science and technology, BIDS EMBASE to over 3500 pharmacological and biomedical journals, and BIDS Social Science Citation Index to over 1400 journals in behavioural and social sciences.

The search was kept as broad as possible with articles in any language being included and searches going back in time as far as databases would allow (1980 for EMBASE; 1981 for Science Citation Index and Social Science Citation Index). A broad strategy was maintained by searching articles for relevant words and phrases (search terms), not only in keywords but also in titles and abstracts.

The search terms fell into three main categories; quality of life, survival and methodology (see appendix 1 for details). The terms were chosen to achieve an adequate balance between sensitivity and precision (or specificity). Sensitivity is the proportion of relevant references captured by the search, while precision is the proportion of articles captured by the search which are relevant. Searches need to be sensitive, in order to achieve a comprehensive result overall, and precise, so that the number of references to check for relevance is not impractical.

The search strategy, applicable to each database, was split into three stages as follows.

Stage 1 Journals in the fields of statistics, epidemiology, clinical trials and biometry (see appendix 1 for details on how these were selected) were searched for articles containing quality-of-life search terms.

Stage 2 All journals on the database were searched for articles containing methodology search terms, that is, terms relating to known statistical methodologies used in the simultaneous analysis of quality-of-life and survival data.

Stage 3 All journals on the database were searched for articles containing both quality-of-life and survival search terms.
The first and second stages were devised to retrieve methodological articles. At Stage 1 this was achieved by restricting the search to the more methodological journals, while at the same time allowing any article that had a reference to quality of life. Conversely, at Stage 2 the search allowed any type of journal available on the database but restricted to those articles that specifically used a known relevant methodology. Both stages were devised to produce a relatively small number of references with a high degree of precision. Applying the first two stages of the search strategy to Science Citation Index, Social Science Citation Index and EMBASE produced a manageable number of references, which contained the bulk of the references for the review, including many key papers.

The third stage was devised to make the overall search more comprehensive. It proved, however, to be problematic by producing an unmanageable number of references (approximately 1400 new references from Science Citation Index alone). Attempts were made to make the number of references more manageable by splitting them according to whether or not they contained a longitudinal search term (see appendix 1), with the expectation that papers containing such a term would have a greater chance of being relevant to the review. In practice, however, this did not prove to be helpful since relevant references were found in both categories. Given the time constraints of the study, it was decided to abandon this part of the search strategy.

Other searching methods
The electronic database searching was supplemented by the following methods:

- the journal *Quality of Life Research* was handsearched from its first issue (1992) for relevant articles
- references known by the researchers or colleagues to be relevant to the study were added to the database
- reference lists of articles already identified were checked and added to the database where relevant (‘exploded’ references).

Selection of relevant literature
The criteria thought to make an article ‘relevant’ to the study are defined formally below, together with a description of the process of checking an article for relevance.

**Criteria**
The criteria for judging the relevance of an article were as follows.

- A paper had to include some sort of quality-of-life assessment over time. Papers in which quality of life was studied at a single time point or as a change from baseline, and in which standard non-longitudinal statistical methods of analysis were used, were rejected.
- Papers in which quality of life was assessed longitudinally in circumstances where survival was not an issue were rejected.
- A paper had to use either a known methodology of interest or a new and clearly detailed methodology. Papers that purely discussed quality of life were rejected.
- Papers that described an application for which the methodology was not clearly detailed were rejected.
- Papers that were essentially discussions of issues relating to quality-of-life instruments, including validity and reliability, were rejected.

As a general rule, if there was any doubt as to whether a paper should be included or not, then it was included.

A paper was defined as ‘key’ if it contained all the elements relevant to the study, that is, it described a methodology used to simultaneously assess health technologies in terms of longitudinal quality of life and survival.

**Methodology**
A strategy was devised (see Figure 1) to assess papers for their relevance to the project using the above criteria. For most articles, abstracts were available and these were used initially to assess the article’s relevance. If the relevance was not clear from the abstract and the journal containing it was readily accessible, then it was assessed by skim-reading the whole paper. In some cases it was not until a paper was obtained and read in detail that it became apparent that it was not relevant; such papers were discarded. Many papers were in journals that were not available from the university libraries and departmental collections at Leicester and Birmingham; copies of these papers were obtained from the British Library.

One author (LJB) made the first trawl, discarding the obviously non-relevant references and including the obviously relevant ones. All three authors then checked the abstracts of the more questionable references and if at least one author thought a paper might be relevant then it was included.
Results of the search

The search produced 1127 references in total, of which 361 were included in the report as relevant (see Table 1). Key papers were obtained from all parts of the search strategy.

At Stage 1 of the database search, statistics and biometry journals produced a small number of papers, of which most were key references, whereas the clinical trial and epidemiology journals produced a large number of references, of which the majority were not relevant.

At Stage 2 of the search a large number of references were brought up, of which a large proportion were associated with quality-adjusted life years (QALYs) and, in particular, their use in decision analysis and health economics (see page 61). These papers were of no direct interest and only a selected group of these references (in general, those that were in readily accessible journals) were included.

Further potential searching

Time limitations resulted in only three electronic databases being searched. These provided good coverage of the literature but, for complete comprehensiveness, other databases should be searched. A variety of medical databases were considered (but not formally used) for the review: MEDLINE, CINAHL, CANCERLIT, AIDSLINE and HEALTHPLAN. In addition, other specialised databases were considered, such as the psychological database PsycLIT, the social sciences database ASSIA, and the statistical databases MATHSCI and CIS.

The overall search strategy identified some grey literature, such as reports, conference papers and theses, which were mainly acquired following recommendation by colleagues. A fully comprehensive search would require a more thorough searching of the grey literature. This would include contacting individuals and departments who are known to be working in the field, together with more formal methods, such as using the special electronic database for grey literature, SIGLE.

TABLE 1 Numbers of references obtained from each part of the search strategy

<table>
<thead>
<tr>
<th>Part of search strategy</th>
<th>References retrieved</th>
<th>Relevant references</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIDS search stage 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QoL search terms in statistical journals</td>
<td>40</td>
<td>17</td>
</tr>
<tr>
<td>QoL search terms in biometry journals</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>QoL search terms in clinical trial journals</td>
<td>69</td>
<td>5</td>
</tr>
<tr>
<td>QoL search terms in epidemiology journals</td>
<td>198</td>
<td>10</td>
</tr>
<tr>
<td>BIDS search stage 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methodology search terms in all journals</td>
<td>600</td>
<td>165</td>
</tr>
<tr>
<td>Handsearching Quality of Life Research</td>
<td>29</td>
<td>12</td>
</tr>
<tr>
<td>Recommended references</td>
<td>92</td>
<td>65</td>
</tr>
<tr>
<td>Exploded references</td>
<td>95</td>
<td>83</td>
</tr>
<tr>
<td>Total</td>
<td>1127</td>
<td>361</td>
</tr>
</tbody>
</table>
The electronic searches only went back until 1980. This may well be adequate since quality-of-life research is a relatively new phenomenon and exploded references should identify the most important earlier articles. More formal searching of databases or journals, however, could be carried out to check for all relevant articles published before 1980.

Other approaches to searching electronic databases were considered but were not carried out because of time limitations. An alternative to using search terms would be to search by author, using the names of people active in the area of quality-of-life research. Also, citation searches of key papers found from the existing search could be tried. In addition, an Internet search could be attempted.

The timing of the electronic searches was such that references were restricted to all those published before 1997. Searching is an ongoing process and the search strategy could be re-run to create a more up-to-date reference database. Relevant articles published after the search that have been brought to our attention have been included in our reference database but have not necessarily been referenced in the report. The bibliography (see page 109) highlights key papers that were published too late for inclusion in this report.

Conclusions

The search aimed to identify all methods proposed for the analysis of quality-of-life and survival data. It is not possible to determine if this has been achieved but the breadth of the search strategy used, together with the method of exploding references, should ensure a reasonably complete coverage of relevant material.
Chapter 3

Background to quality-of-life and survival data

The meaning of quality of life

In its most general context, quality of life is a concept incorporating all the factors that might impact on an individual’s life. In health service research it is more usual to consider health-related quality of life, which includes only those factors that affect an individual’s health. There is no general agreement regarding the identification of such factors.\(^1,2\)

WHO defines health as ‘a state of complete physical, mental and social well-being, and not merely the absence of disease and infirmity’.\(^3\) Quality of life is often referred to in these terms. For example, Schumacher and colleagues\(^4\) specify the dimensions comprising quality of life as follows:

- symptoms of disease and side-effects of treatment (e.g. nausea, pain, anorexia)
- physical and functional status (e.g. mobility, self-care, fatigue)
- emotional status (e.g. anxiety, depression, satisfaction with care)
- social functioning (e.g. family interaction, work/recreation, time with friends).

In this study, a pragmatic view is taken and quality of life is accepted as any measure that purports to reflect health-related quality of life or some aspect of it.

Measuring quality of life

There are many issues to consider when attempting to measure quality of life, such as, what questions should be asked, how should responses be recorded, when should questions be asked, of whom should questions be asked and who should do the asking. These, together with other aspects of measuring quality-of-life data in clinical trials, have been extensively reviewed and discussed elsewhere.\(^5,7\)

The quality of life of a patient is usually measured using an instrument in the form of a questionnaire designed for patient completion. The questionnaire generally comprises sets of questions or items relating to the various dimensions of quality of life, such as physical, psychological or social. The format of responses may be ‘yes/no’, a series of ordered categories, or a linear analogue scale.

Many instruments are used in the assessment of quality of life,\(^6,10\) for example, Rotterdam Symptom Checklist, Nottingham Health Profile, Sickness Impact Profile, Hospital Anxiety and Depression Scale, Short Form (SF) 36. Generic quality-of-life instruments measure general aspects of quality of life and are applicable in a wide range of research settings, while non-generic instruments are relevant for a specific disease or treatment. Some questionnaires are dimension-specific, that is, they only ask questions relating to a particular aspect of quality of life (e.g. Hospital Anxiety and Depression Scale). A review of the use made of quality-of-life instruments is currently being undertaken.\(^11\)

Much work has been done to test the validity and reliability of some such instruments. An instrument is valid if it is actually measuring what it is designed to measure and it is reliable if, all things being equal, it measures consistently from one occasion to the next. There has been much discussion regarding the requirements of quality-of-life measures and methods for assessing such requirements, for example, Cronbach alpha coefficients and factor analysis.\(^2,12-15\) The criteria for assessing quality-of-life instruments are reviewed in another NHS R&D Health Technology Assessment report.\(^16\)

Another approach to measuring quality of life is based on the valuations of or preferences for different quality-of-life states. The advantage of this approach is that it yields a single value, often referred to as a utility, as a measure of quality of life. The utility value is a single score ranging from 0 – representing a quality of life equivalent to death, to 1 – representing perfect health. The relationship between descriptive and valuation approaches has been explored.\(^17,18\) This method of measurement is usually used in the context of the QALY (as this is covered later in chapter 15, details of the methodology are included there).
Quality-of-life data

Nature of quality-of-life data

Quality-of-life data is generally longitudinal in nature. Some studies assess quality of life at one time point only or take a baseline measure and a follow-up measure but, generally, quality of life is recorded at more than two time points during the course of a study. There can be any number of time points and these will not necessarily be evenly spaced or consistent across individuals.

Quality-of-life data is generally multivariate in nature. At one extreme, quality of life can be measured by a single global measurement such as the Karnofsky Index,\textsuperscript{19} while at the other, assessment is made by a multitude of items measuring a variety of conceptual dimensions, for example, the Sickness Impact Profile measures, in 136 items, 12 dimensions. Items within each dimension are sometimes combined as a weighted or unweighted sum to create dimension-specific global measures, or sometimes all items from the questionnaire are combined to create an overall global quality-of-life score. Thus, at each time point, quality-of-life data may comprise a single measure or a large set of measures.

Quality-of-life data may take various forms (binary, ordinal or continuous) for which a variety of distributional assumptions are appropriate. Responses to each item on a questionnaire may yield binary data from a yes/no response, ordinal data from a categorical scale or continuous data from a linear analogue scale. The aggregation of items, to give a global measurement of a dimension or of overall quality of life, results in data that is usually treated as continuous, despite the fact that the global measure may take only a restricted set of values from a restricted range. The distribution of quality-of-life variables as measured from a linear analogue scale or as an aggregated global score may or may not be normally distributed. Sometimes data may be transformed to create normally distributed data for analysis but on other occasions (e.g. when there is a heavy preponderance of a particular response value, perhaps corresponding to ‘no problem’) this may be difficult.

Problem of missing data

One of the main problems in analysing longitudinal quality-of-life data is caused by missing data. The validity of the analysis of data with missing values is dependent on the mechanism associated with the missing data. Responses missing at a particular time point, \( t \), may be categorised in three ways, as defined by Little and Rubin:\textsuperscript{20}

(i) **missing completely at random**, when the probability of response at time \( t \) is independent of both the previously observed values and unobserved values at time \( t \)

(ii) **missing at random**, when the probability of response at time \( t \) depends on the previously observed values but not the unobserved values at time \( t \)

(iii) **non-ignorable non-response**, when the probability of response at time \( t \) depends on the unobserved values at time \( t \) and possibly on the previously observed values as well.

There are three different forms of missing data that can arise in a longitudinal quality-of-life study designed to take at least three serial measures of quality of life from each participant:

(i) single missing items from an otherwise complete questionnaire

(ii) intermittent missing whole questionnaires

(iii) missing data resulting from a patient dropping out of the study.

The problems associated with each form of missing data and the ways of handling them are discussed in general below and, in relation to the illustrative example, in chapter 4 (see pages 15–16).

Single missing items

Missing single items mainly cause problems in terms of calculating global scores since a dimension-specific score is based on the values of all items within the dimension. If values for items are missing in the dataset, then they either need to be imputed or the calculation of global scores needs to accommodate them (see page 10).

Problems will also occur if the item with missing values is to be analysed independently as a measure of quality of life. Single missing items may not prove to be a major problem in this case, since it is usually not unreasonable to assume that this sort of missing data is missing completely at random. This assumption would be untenable in situations where consistent non-response to an item suggests that the question is inappropriate or difficult to answer.

Intermittent missing whole questionnaires

An intermittent missing questionnaire occurs if a patient does not complete a questionnaire at the required time point but has completed questionnaires at time points before and after the missing form. It may be possible to assume that this sort of data is missing completely at random, since the patient has not dropped out of the study altogether.
The probability of response may, however, depend on covariates, such as treatment, or may depend on the quality of life experienced at that time, in which case it would be invalid to assume that the missing data mechanism was ignorable.

**Missing data resulting from patient drop-out**

A drop-out is defined as a patient who withdraws from a study before they have completed all planned assessments. Patients may withdraw from a study for a number of reasons, illness, death, cessation of treatment, lack of treatment effect, lost to follow-up, or they may reach the censor date of the study. In studies where survival, as well as quality of life, is an endpoint, patients are generally severely ill; thus drop-out caused by death or illness will be a common occurrence. In such situations, the drop-out process may depend on the unobserved measurements (i.e. those measurements that would have been observed had the patient not dropped out) and the incomplete follow-up of subjects is called **informative drop-out** (other terms used include **informative censoring** and **non-ignorable drop-out**).

Once a patient has dropped out of a study, no more information on quality of life is available from that point onwards. Quality-of-life information could be considered as censored at the date of drop-out. In terms of analysis, drop-outs cause problems because they create missing data which are not just missing at random. The missing data mechanism is likely to depend upon the health status of the patient and is therefore non-ignorable.

One way of dealing with missing data from drop-outs is to impute values to replace the missing data from data that already exist. There are a variety of methods for doing this but it may be difficult if there is a large amount of missing data. Otherwise the method of analysis must allow for informative drop-out. Methods of analysis which simultaneously assess quality-of-life and survival data (see chapters 14–20) overcome the problems associated with drop-out due to death but informative drop-out for other reasons also needs to be considered in any analysis.

**Handling the multivariate nature of the quality-of-life endpoint**

Quality-of-life instruments consist of many items, often grouped into a number of dimensions. Because quality of life can be considered in terms of individual items or in terms of separate dimensions, it is, as an endpoint, potentially multivariate in nature. In some situations it may be desirable to consider each item or dimension as a separate quality-of-life endpoint. In a descriptive analysis, this will only cause problems of presentation and interpretation if the different items or dimensions give conflicting conclusions. If hypothesis testing is involved, analysis of multiple endpoints will lead to the problem of multiple testing, where the probability of a finding false-positives increases as the number of tests performed increases. When quality-of-life measures consist of only a few dimensions, the problem of multiple testing should not be a major one.

In some studies it may be possible to limit the amount of hypothesis testing by specifying in advance a few key quality-of-life measures on which hypotheses will be tested, leaving the remaining variables to be analysed purely descriptively. If this approach is not practical or desirable then the analysis will need to account for the multivariate nature of the quality-of-life endpoint.

There are a variety of ways of handling multiple endpoints in clinical trials. The application of such approaches to quality-of-life data has been discussed and some methods have been applied and compared in the analysis of quality-of-life data. One approach is to combine multiple endpoints to create global scores before analysis. Another approach is the post-analysis combination of results from the univariate analysis of each separate endpoint. Alternatively, a hierarchical approach to the analysis can be used. These are now considered in turn.

**Combining multiple endpoints to create global scores**

For each individual, the values of the items that make up a quality-of-life endpoint can be combined in some way to form a global score. In some cases the items within each quality-of-life dimension may be combined to create dimension-specific global scores, while in others an overall quality-of-life global score may be created either by combining all items on a questionnaire or by combining dimension-specific global scores. Use of a single global quality-of-life score simplifies statistical analysis and should be aimed at when sensible and justifiable. Global scores can be analysed using standard univariate techniques.

Global scores can be calculated using either an unweighted or a weighted sum. It is suggested that unweighted sums should only be used to combine items which are highly positively
correlated. However, a weighted sum may make interpretation difficult and the weights used may be controversial. Weights can be determined either from the data, using scores from a factor analysis, or from decision theory, using utility analysis techniques, or arbitrarily. Olschewski and Schumacher recommend aggregation using data-oriented procedures.

A method for calculating global scores proposed by O’Brien has been applied to quality-of-life data. It is a non-parametric approach and creates global scores from ranks rather than actual data values. The data for all treatment groups are pooled and, for each variable in the multivariate quality-of-life endpoint, the values across all individuals are ranked. A global score is created for each individual by summing the ranks for all variables.

Problems in calculating global scores can occur if data are missing on some of the items within the score. One way to tackle the problem is to impute the missing values. This is only feasible if they are limited in number. If a subject has a missing value for an item then the value could reasonably be imputed from:

- values of the other items within the dimension for that patient
- values of the other items in the patient’s questionnaire
- values of the item on the patient’s questionnaires at time points on either side of the missing value.

If missing values are not imputed then the formula for calculating global scores needs to allow for the number of items involved in the calculation. Using a mean rather than a sum allows accommodation of missing values into the global score since the mean can be calculated for a reduced number of items. Alternatively, expressing the sum as the percentage of the maximum achievable score allows for the possibility of a reduced number of items. Otherwise the global score should be recorded as missing if any item within it is missing.

**Combining results from univariate tests on multiple endpoints**

The simplest method for testing a global null hypothesis of no treatment effect, using the results from multiple univariate tests, is to use a Bonferroni-type adjustment. The p-values from the multiple univariate tests are adjusted by multiplying each p-value by the number of tests carried out. Each endpoint can be assessed using these adjusted p-values or a global null hypothesis can be assessed using the minimum p-value.

This method has been recommended for use with quality-of-life data and has been applied and compared to other methods in a quality-of-life setting. The main drawback with the global null hypothesis approach is that it confines attention to the smallest p-value and may be too conservative.

A parametric method for combining results from multiple univariate t tests, originally proposed by O’Brien but developed by Pocock and colleagues, has been applied to quality-of-life data. The following test statistic can be used to assess a global null hypothesis of no treatment effect:

$$J^T S^{-1} t / (J^T S^{-1} J)^{1/2}$$  

where J is a vector of ones, S is an estimated correlation matrix and t is a vector of t statistics from the separate univariate t tests. The test statistic has an asymptotic standard normal distribution.

The main drawback of these methods is that they do not give an estimate of the treatment effect, they just provide a test statistic.

**Hierarchical approach**

Multilevel models have been advocated for the analysis of data that have a hierarchical structure. Longitudinal data can be thought of as hierarchical data, with level one of the hierarchy being observations over time within a patient and level two the patient. Multilevel models have been used to analyse longitudinal quality-of-life data and their application is discussed in more detail in chapter 7.

The hierarchical approach also provides a means of handling the multivariate nature of the quality-of-life endpoint. The multiple dimensions that constitute quality of life can simultaneously be analysed in a multilevel model by adding an extra level to the standard longitudinal data model, with level one becoming the various quality-of-life dimensions, level two the observations over time and level three the patients. Models are fitted to the data in the usual way using a variety of assumptions.

Multilevel models have an advantage over methods discussed previously in that they provide estimates of the treatment effect as well as test statistics. Treatment effects are estimated for each dimension separately and, if appropriate, an overall summary estimate may be obtained. The model also allows the correlation between dimensions to be estimated. Multilevel models are flexible in...
that they can cope with situations where some of the dimension scores may be missing for some patients. However, the application of multilevel models to quality-of-life data in general is problematic since the method assumes the missing data mechanism is ignorable, which is not generally true of quality-of-life data.

**Survival data**

**Nature of survival data**

Measurements of the time between two events, an initial occasion and an endpoint of interest, are known as survival data. In a clinical trial, the initial occasion will usually be defined as the same event for all individuals, such as date of randomisation, date of starting treatment or date of diagnosis. The endpoint of interest will depend on the nature of the disease and treatments under investigation but is often death or relapse from a period of disease remission.

Survival data are different from other types of continuous data because over the period of a study the endpoint of interest is not necessarily observed in all subjects. This may occur because:

(a) some patients are lost to follow-up, that is, they are not followed to the end of the study and, when last seen, have not experienced the event of interest, or

(b) the event has not occurred in some patients by the time the study closes for analysis.

Such data are referred to as censored survival times and are different from missing data in that they provide a lower bound for the actual non-observed survival times. Any analysis carried out on survival data should use statistical methods that do not disregard censored data and, indeed, make the fullest possible use of it to avoid loss of information.

This study is specifically interested in the analysis of quality-of-life and survival data where survival generally means time to death. Thus in the context of this report, unless specified otherwise, survival will refer to time to death from a fixed origin, typically randomisation in a clinical trial. Most findings, however, will also be applicable to survival defined in other ways such as, for example, time from randomisation to relapse or death (whichever occurs first).

**Informative censoring**

Most analytical methods used for survival data with censored observations are only valid if censoring is non-informative. This means that the censoring is not related to any factors associated with the actual survival time, that is, the actual survival time, \( t \), of an individual is independent of any mechanism which causes that individual’s survival time to be censored at time \( c \), where \( c < t \). When the censoring mechanism is not independent of survival time, informative censoring occurs and standard methods used for survival analysis are invalidated.

Informative censoring is a particular problem when analysing quality-of-life and survival data simultaneously and will be discussed in more detail in the context of each proposed statistical methodology.

**References**


33. Beacon HJ. The statistical analysis of self assessed quality of life data in cancer clinical trials [PhD thesis]. London School of Hygiene and Tropical Medicine, University of London; 1996.

Introduction

The methods identified by the review and discussed in this report are illustrated by application to data from a previously conducted study, referred to here as the MIC study. This illustrative example is, in fact, the second of two concurrent Phase III trials of MIC (mitomycin, ifosfamide and cisplatin) chemotherapy which were conducted in patients with non-small cell lung cancer. The results from both trials are reported elsewhere and this report focuses on the quality-of-life and survival data from the second trial (MIC2) only.

The general background to the trial is presented below, including the characteristics of patients in the quality-of-life study. The measurement of quality of life in the MIC study is described together with details of missing data.

Background

In 1988, a randomised Phase III trial was initiated at the Cancer Research Campaign Clinical Trials Unit, Birmingham, to evaluate the role of chemotherapy in the treatment of non-small cell lung cancer in patients with extensive stage disease. Patients were randomly allocated to receive either standard palliative treatment (PAL arm), usually radiotherapy, or MIC chemotherapy, to a maximum of four courses, followed by palliative care (CT arm). The aim of the study was to compare treatments in terms of both survival and quality of life. Quality of life was an important endpoint in the study because both treatments were considered largely palliative and MIC chemotherapy was considered by some clinicians to be highly toxic.

The trial closed in March 1996, by which time 359 patients had been randomised into the study. For practical reasons associated with the availability of the research nurse, the quality-of-life component of the study was carried out only on a subset of trial patients, essentially consisting of patients treated at three main oncology centres. Quality-of-life data were collected for 109 patients from the trial, 67 on the CT arm and 42 on the PAL arm.

### Table 2 Characteristics of patients in the MIC quality-of-life study

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>CT arm (n = 67)</th>
<th>PAL arm (n = 42)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>47 (70%)</td>
<td>32 (76%)</td>
</tr>
<tr>
<td>Female</td>
<td>20 (30%)</td>
<td>10 (24%)</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>62</td>
<td>66</td>
</tr>
<tr>
<td>Range</td>
<td>42–75</td>
<td>49–75</td>
</tr>
<tr>
<td>Inter-quartile range</td>
<td>56–68</td>
<td>61–71</td>
</tr>
<tr>
<td>Histology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Squamous</td>
<td>35 (52%)</td>
<td>26 (62%)</td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td>28 (42%)</td>
<td>10 (24%)</td>
</tr>
<tr>
<td>Large cell undifferentiated</td>
<td>2 (3%)</td>
<td>1 (2%)</td>
</tr>
<tr>
<td>Not known</td>
<td>2 (3%)</td>
<td>5 (12%)</td>
</tr>
<tr>
<td>WHO performance status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>20 (30%)</td>
<td>7 (17%)</td>
</tr>
<tr>
<td>1</td>
<td>29 (43%)</td>
<td>17 (40%)</td>
</tr>
<tr>
<td>2</td>
<td>17 (25%)</td>
<td>18 (43%)</td>
</tr>
<tr>
<td>Not known</td>
<td>1 (2%)</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>

The sex, age, histology and performance status of the patients on entry to the quality-of-life study are given in Table 2.

The palliative group, in comparison to the chemotherapy group, had a slightly greater proportion of males, was on average an older group and had a greater proportion of squamous tumours. In addition, the group appeared to have poorer performance status at baseline compared with the chemotherapy group. These differences in patient characteristics may need to be considered in any comparison of treatments.

Measuring quality of life

**Instrument**

Quality of life was assessed using questionnaires completed by the patients with help of a dedicated quality-of-life research nurse. The questionnaire was designed specifically for the trial but was based on the EORTC QLQ-LC13, the lung cancer module.
Background to the illustrative example – the MIC study

of the quality-of-life questionnaires designed by the European Organisation for Research and Treatment of Cancer (EORTC). ²

The questionnaire consisted essentially of 11 questions evaluating specific physical and psychological aspects of a patient’s quality of life, plus a more general question on malaise (see Table 3). All questions related to how the patient had been feeling over the previous 3 weeks. Responses were on a four-level ordered categorical scale (‘None’, ‘A little’, ‘Quite a bit’ and ‘Very much’) and for analytical purposes have been coded from 0 to 3, respectively.

The questionnaire also asked about the patient’s performance status using WHO categories and asked how they felt in comparison to the time of the previous questionnaire; however, these questions will not be considered for the purposes of this report.

Two measures of quality of life from the MIC study, one ordinal and one continuous, are used here to demonstrate the various methodologies: these are, respectively, the malaise question (MAL) and the mean quality-of-life score (MQS). The MQS was calculated from the answers to the 12 questions presented in Table 3. If a patient did not respond to all 12 items, then the mean was calculated as the mean of the reduced number of non-missing values. In total, 399 MQSs were calculated and 11% of these were calculated from incomplete data (see page 15).

Timing of assessments

In the MIC study, the quality of life of patients was assessed only during their treatment period. There were several reasons for this decision, which was made when the study was designed:

(i) the main aim of the quality-of-life part of the study was to assess both treatments in terms of their immediate impact on quality of life
(ii) the survival time of patients in the study in general is short (median survival time for all patients in trial was 5.6 months) and thus long-term quality of life was not thought to be an issue when comparing treatments
(iii) the collection of data during this period was most practicable in that the patients were attending clinic for treatment and were therefore readily accessible.

The study was designed so that patients completed questionnaires on entry to the trial and then every 3 weeks thereafter, completing five forms in total on the CT arm and four forms in total on the PAL arm. In theory, all questionnaires should thus have been completed within 12 weeks of entering the trial; however, in reality the timing of the questionnaires varied considerably from this planned administration (see Table 4). On average, the baseline questionnaire was completed 1 week after entry to the trial and then subsequent questionnaires were completed every 3 weeks after this.

The analysis of the data needed to focus on the period during which quality-of-life data was collected and applicable. A cut-off point defining

<table>
<thead>
<tr>
<th>Item</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cough</td>
<td>Do you have a cough?</td>
</tr>
<tr>
<td>Severe dyspnoea</td>
<td>Do you get breathless on mild activity like dressing?</td>
</tr>
<tr>
<td>Moderate dyspnoea</td>
<td>Do you get breathless when walking on the flat?</td>
</tr>
<tr>
<td>Mild dyspnoea</td>
<td>Do you get breathless on stairs or walking uphill?</td>
</tr>
<tr>
<td>Haemoptysis</td>
<td>Have you coughed blood?</td>
</tr>
<tr>
<td>Pain</td>
<td>How much pain are you getting?</td>
</tr>
<tr>
<td>Appetite</td>
<td>Have you noticed any loss of appetite?</td>
</tr>
<tr>
<td>Anxiety</td>
<td>Have you been worrying?</td>
</tr>
<tr>
<td>Depression</td>
<td>Have you been depressed?</td>
</tr>
<tr>
<td>Dysphagia</td>
<td>Have you any difficulty swallowing?</td>
</tr>
<tr>
<td>Nausea</td>
<td>Did you feel sick during or since your last treatment? (CT arm)</td>
</tr>
<tr>
<td></td>
<td>Have you been feeling sick? (PAL arm)</td>
</tr>
<tr>
<td>Malaise</td>
<td>Have you been feeling generally ill?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cough</td>
<td>Do you have a cough?</td>
</tr>
<tr>
<td>Severe dyspnoea</td>
<td>Do you get breathless on mild activity like dressing?</td>
</tr>
<tr>
<td>Moderate dyspnoea</td>
<td>Do you get breathless when walking on the flat?</td>
</tr>
<tr>
<td>Mild dyspnoea</td>
<td>Do you get breathless on stairs or walking uphill?</td>
</tr>
<tr>
<td>Haemoptysis</td>
<td>Have you coughed blood?</td>
</tr>
<tr>
<td>Pain</td>
<td>How much pain are you getting?</td>
</tr>
<tr>
<td>Appetite</td>
<td>Have you noticed any loss of appetite?</td>
</tr>
<tr>
<td>Anxiety</td>
<td>Have you been worrying?</td>
</tr>
<tr>
<td>Depression</td>
<td>Have you been depressed?</td>
</tr>
<tr>
<td>Dysphagia</td>
<td>Have you any difficulty swallowing?</td>
</tr>
<tr>
<td>Nausea</td>
<td>Did you feel sick during or since your last treatment? (CT arm)</td>
</tr>
<tr>
<td></td>
<td>Have you been feeling sick? (PAL arm)</td>
</tr>
<tr>
<td>Malaise</td>
<td>Have you been feeling generally ill?</td>
</tr>
</tbody>
</table>

**TABLE 3 Questions from the MIC quality-of-life questionnaire**

**TABLE 4 Timing of questionnaires (in weeks) from date of entry to trial**

<table>
<thead>
<tr>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planned</td>
<td>0</td>
<td>3</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>Median</td>
<td>0.9</td>
<td>4.0</td>
<td>7.2</td>
<td>10.3</td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
<td>1.9</td>
<td>5.4</td>
<td>8.7</td>
</tr>
<tr>
<td>Maximum</td>
<td>6.7</td>
<td>9.0</td>
<td>26.0</td>
<td>27.0</td>
</tr>
</tbody>
</table>
this period needed to be established. Of the 399 completed questionnaires, 41 questionnaires were not completed within the 12 weeks from date of entry to the trial. If a 12-week cut-off was chosen for the analysis then the information from these questionnaires would be lost. Investigation of the distribution of times of the questionnaires indicated that 18 weeks from entry to trial would be the optimal cut-off value. This value enabled the majority of the questionnaires to be incorporated in the analysis, with information from only four (three participants) being lost, while at the same time not creating a situation in which individuals had large time spans with no quality-of-life information. It should be noted that 37 participants (34%) died during this 18-week period.

Missing data in the quality-of-life study

The number of returned questionnaires in the MIC study diminishes with time. A reduction in numbers from one time point to the next results either from intermittent missing whole questionnaires or from individuals drop-out of the study (see Table 5). For example, with the second questionnaire on the CT arm, only 63 questionnaires from an expected 67 were returned; three were intermittently missing and one was missing because the patient had dropped out of the study, leaving 66 patients left in for the third questionnaire. These types of missing quality-of-life data were discussed in general in chapter 3 (see page 8). A third type of missing data, single missing items from otherwise complete questionnaires, was also covered in this chapter and all three are discussed here in the context of analysis and interpretation of quality-of-life data from the MIC study.

**Single missing items**

For some returned questionnaires in the study, responses were not given to all questions. The extent of this problem was not great (see Table 6), with 89% of questionnaires being completed fully and only 3% having more than one item missing. The maximum number of missing items on any one questionnaire, five out of 12, occurred just once. This type of missing data caused problems in calculating MQSs (see page 14) and when data were missing on a questionnaire, the calculation of the mean was based on the reduced number of non-missing items. This assumes that the missing data is ignorable, that is, the reason for it being missing is not related to the quality of life at that time.

<table>
<thead>
<tr>
<th>Number of missing items</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>356 (89)</td>
</tr>
<tr>
<td>1</td>
<td>33 (8)</td>
</tr>
<tr>
<td>2</td>
<td>8 (2)</td>
</tr>
<tr>
<td>3–5</td>
<td>2 (1)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>399</strong></td>
</tr>
</tbody>
</table>

Because this report analyses the malaise item independently, the missing data for this specific item causes problems. Missing values for the MAL variable could be considered either as intermittent missing data or as drop-out data depending on the questionnaire from which the item was missing. The malaise item was missing on eight questionnaires. In four of these, the missing item could be classed as intermittent, in that MAL values were available on questionnaires both before and after that with the missing value. These missing values were tackled in the same way as if the whole questionnaire was missing (see below). The remaining four missing values occurred in the final questionnaires of three patients (two values were on consecutive questionnaires belonging to one participant) and, for the purposes of this report, were treated in the same way as missing data resulting from drop-out (see below). In general, the nature of this type of missing data is different from drop-out data since, with this type of data, the patient is still participating in the study; treating them as drop-outs is, thus, questionable.
Intermittent missing whole questionnaires
In the MIC study, five patients had intermittent missing whole questionnaires, with one having two consecutive questionnaires missing. In the analysis of the data, the values from the questionnaire prior to the intermittent missing one were assumed to carry over until the questionnaire after the missing one. In this way intermittent missing questionnaires were effectively ignored. This assumes that the reason for them being missing was not related to the patient’s quality of life at that time.

Missing data resulting from drop-out
In the MIC study, a drop-out is defined as a participant who did not complete a final questionnaire (fifth on CT arm and fourth on PAL arm). Participants dropped out at various times during the study (see Table 5).

There are four categories of drop-out:

- drop-out directly caused by death (i.e. patient died within 3 weeks of last completed questionnaire before the next planned assessment)
- drop-out not due to death (i.e. patient was alive at next planned assessment) and patient died before the analysis cut-off time (i.e. within 18 weeks of entry to trial)
- drop-out not due to death (i.e. patient was alive at next planned assessment) and patient died after the analysis cut-off time (i.e. after 18 weeks from entry to trial)
- drop-out patient last known to be still alive (after 18 weeks).

In the MIC study, 46 patients dropped out of the study at some point. They were balanced across the two treatment arms: 43% (29/67) of patients on the CT arm and 40% (17/42) on the PAL arm. The type of drop-out, in terms of time of death relative to last returned questionnaire, appeared to be slightly different on the two arms (see Table 7 and Figures 2 and 3), with almost half of the drop-outs on the CT arm (14/29) surviving the 18-week period from entry to trial but only three surviving the 18-week period of analysis on the PAL arm.

In some analyses of the MIC data, values for missing data caused by drop-out were imputed. Several approaches to imputing this missing data were considered (see Table 8). Each have their disadvantages and it was decided, in situations where imputation was necessary, to use the last value carried forward approach. The impact of different methods for imputation should be investigated in a sensitivity analysis.

References
FIGURE 2  Drop-outs on CT arm of the MIC study: timing of last questionnaire and death in relation to study entry time (NB: patient 106 was still alive when last seen, 186 weeks after study entry)

FIGURE 3  Drop-outs on PAL arm of the MIC study: timing of last questionnaire and death in relation to study entry time (NB: patient 97 died on same day as last questionnaire; patients 131 and 184 completed only one questionnaire, on day of study entry)
### Background to the illustrative example – the MIC study

**TABLE 8** Possible approaches to imputing missing data in the MIC study

<table>
<thead>
<tr>
<th>Approach</th>
<th>Description</th>
<th>Problems</th>
</tr>
</thead>
</table>
| Last value carried forward    | Assume drop-outs stay in the same health state, as measured at the last assessment, until date of death or censor date, whichever comes first. | • If last recorded measure of QoL was ‘good’, it may be invalid to assume patient remains in a ‘good’ state until death.  
• If dropped out early, assumption of steady state needs to be applied to long period. |
| Worst value carried forward   | Assume drop-out caused by ill health and thus drop-outs move into the poorest health state (if they are not already in that state) from the time of drop-out onwards. | • Choice of time at which patient is moved to ‘poor’ health state may not be obvious – time of next planned assessment could be used.  
• If drop-out is not caused by ill health, as may well be true for those patients who die a considerable time after dropping out, they may be allocated to ‘poor’ state too early.  
• Bias may result since assumption only applicable to those who dropped out while in ‘good’ state. |
| Linear decrease over time     | Assume QoL decreases linearly from time of drop-out until death.             | • Bias may result since assumption only applicable to those patients who dropped out while in ‘good’ state.  
• Assumption of linearity may be questioned.  
• Method difficult to apply to individuals with censored survival times. |
In some studies, the main criterion by which health technologies are assessed is quality of life. When the disease under study is potentially fatal, such as cancer, survival will also be an issue, whereas with long-term chronic diseases, such as arthritis, this will not be the case. This review considers situations where both quality of life and survival are of interest. Quality of life is usually the main endpoint of interest in cancer clinical trials investigating treatments which are purely palliative in nature and, within this context, the relief of symptoms is weighed against the adverse effects of treatment. It will also be the main endpoint in trials of new treatments which have been developed to reduce treatment toxicity compared with standard treatments, without any detrimental effect on survival. In these types of studies, survival is an issue but mainly through the problems it causes in interpreting the quality-of-life endpoint.

In studies of seriously ill patients who are near to death, the collection of quality-of-life data over time is problematic. Patients drop out of the study because of illness or death and this results in patient attrition, that is, a reduction in the number of subjects in the quality-of-life study over time. The drop-out process is related to the health status of the patient and is therefore likely to be informative (see page 9); thus the missing data mechanism in such studies is non-ignorable. Ironically, the situations in which quality of life is most likely to be the main endpoint, such as clinical trials of palliative treatments for cancer patients, are the situations in which the problem may well be most serious.

The aim of this part of the report (chapters 5–9) is to discuss methods for analysing longitudinal quality-of-life data. Although the methods discussed here relate to longitudinal data collected over three or more time points, which may be fixed or varying, some could be satisfactorily and usefully applied to data sets with measurements at just two time points. However, in situations where data is collected at two time points, the analysis is usually in terms of change in quality of life and, as with data collected at just one time point, standard non-longitudinal statistical methods can be used. Informative drop-out may still cause problems in these circumstances and cautious interpretation of the results is needed.

The analysis of quality-of-life data should begin descriptively to give the investigator and reader insight into the data. The methods that can be used to describe quality-of-life data over time are discussed in chapter 6, where the problems of interpretation caused by informative drop-out are highlighted and methods that attempt to account for the problems are described.

Quality-of-life data collected over time are often analysed using standard methods of longitudinal data analysis. These methods all assume that the missing data mechanism is ignorable. This could be appropriate in studies in which survival is not an issue but, where there is informative drop-out, analysis using these longitudinal methods could lead to invalid conclusions. Although the focus of this report is on the analysis of quality-of-life data in situations in which survival is also an issue, chapter 7 is dedicated to describing some of the standard methods of longitudinal data analysis. Despite the general inappropriateness of these methods in situations of informative drop-out, the methods are described briefly with a discussion of their potential problems and possible use with quality-of-life data. In chapter 8 methods for modelling longitudinal quality-of-life data are described which attempt to deal with the problem of informative drop-out. A summary and discussion of quality-of-life analysis in the presence of informative drop-out is presented in chapter 9.
Introduction

The interpretation of longitudinal quality-of-life data can be difficult and an initial exploratory analysis often gives an insight into the data before any formal testing or modelling is carried out. Descriptive methods of analysis do not have the problems associated with multiple testing and, hence, the quality-of-life data can be explored as extensively as desired. The variety of ways of exploring the data both at the patient level and the treatment group level are discussed below. In general, the clearest way to describe data is graphically but, at the group level, it may be preferable to tabulate the results. Descriptive methods for analysing quality-of-life data have been reviewed elsewhere.1,2

Patient profiles

The examination of each patient’s quality-of-life data over time individually can be very helpful. It may reveal a consistent pattern across patients and will highlight errors, outliers and patterns of missing data.

Patient profiles can be examined by plotting individual patient scores over time. This technique is most useful for continuous data but may also be used for ordinal data. Carlens and colleagues3 and Nou and Aberg4 plotted each patient’s quality of life over time and called them vitagrams. Profiles for each patient can be plotted as a set of mini-graphs or they can be overlaid on one graph. Individual profiles can be overlaid for each treatment group separately and then used to compare patterns within each treatment group.

Individual patient profiles of MQSs over time are plotted and overlaid for each treatment group in the MIC study (see Figures 4 and 5). These show that the data vary widely at all time points and no obvious patterns, in terms of change in quality of life over time, are apparent.

The main problem with patient profile data is that it is often impractical to display data from large numbers of patients. One solution is to plot a simple random sample of the patients in a study,2 another may be to categorise and plot the data on a Lexis diagram, as described below.

Lexis diagrams, or variations of such (as used for the MIC data in Figures 6 and 7), describe individual patient data and allow data for reasonably large numbers of patients to be displayed at once.2 Each patient profile is represented by a line over time which is solid or dotted depending on the value of a binary quality-of-life variable at that time. The patient profiles are usually ordered by date of entry to study. If desired, a Lexis diagram can be extended to incorporate a higher level categorical variable by using several different types of line but the diagram becomes difficult to interpret as the number of categories increases. The technique is most appropriate for binary data but can also be used for ordinal data with few categories. Continuous and ordinal level data can be adapted for a Lexis diagram by grouping the values into, ideally, two levels.

Patient profiles showing time spent with malaise (MAL = 1, 2 or 3) or without malaise (MAL = 0) were plotted for the MIC study (see Figures 6 and 7) by assuming the value of MAL measured at each questionnaire point carries over to the next assessment. Values at the last questionnaire (fifth for CT group and fourth for PAL group) were carried over to the time when the next assessment would have taken place. The graphs do not show any clear patterns in terms of change in quality of life over time.

Plots of individual patient profiles may reveal patterns of missing data. Intermittent missing data will be apparent as gaps in the lines, while a profile ending early will signify a patient drop-out. If a patient dies during the course of the quality-of-life study then this can be plotted as a symbol at the appropriate time. Patient profiles can be grouped and overlaid according to differing lengths of follow-up and also, possibly, differing reasons for shortened follow-up, giving a possible insight into the association of the drop-out process with previous quality of life (see page 25 for further details on this approach relating to group profiles).
FIGURE 4 Patient profiles of MQS over time for the CT arm of the MIC study

FIGURE 5 Patient profiles of MQS over time for the PAL arm of the MIC study
After examining individual patient profiles of quality of life, it is necessary to summarise the experience of patients in each treatment group to enable the treatments to be compared more clearly in relation to their effect on quality of life over time. Plots of group profiles over time, overlaid on the same graph, enable a clear comparison of groups. Informative drop-out can cause problems in interpretation and possible approaches for analysis are discussed below.

### Summary measures to plot

There are two main types of group summary statistic that can be plotted over time, an average measure of quality of life or a proportion with a certain level of quality of life. The choice depends on the type of quality-of-life measure being summarised. In situations where the population is not stable over time, there are several choices regarding which subjects to include in the calculation of the summary measure at each time point and this is discussed below, followed by examples from the MIC study.

**FIGURE 6** Patient profiles of malaise over time for the CT arm of the MIC study (---, malaise; - - - -, no malaise)

<table>
<thead>
<tr>
<th>Patient number</th>
<th>Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>357</td>
<td>1</td>
</tr>
<tr>
<td>353</td>
<td>2</td>
</tr>
<tr>
<td>349</td>
<td>3</td>
</tr>
<tr>
<td>345</td>
<td>4</td>
</tr>
<tr>
<td>331</td>
<td>5</td>
</tr>
<tr>
<td>327</td>
<td>6</td>
</tr>
<tr>
<td>321</td>
<td>7</td>
</tr>
<tr>
<td>185</td>
<td>8</td>
</tr>
<tr>
<td>179</td>
<td>9</td>
</tr>
<tr>
<td>172</td>
<td>10</td>
</tr>
<tr>
<td>164</td>
<td>11</td>
</tr>
<tr>
<td>160</td>
<td>12</td>
</tr>
<tr>
<td>159</td>
<td>13</td>
</tr>
<tr>
<td>153</td>
<td>14</td>
</tr>
<tr>
<td>144</td>
<td>15</td>
</tr>
<tr>
<td>141</td>
<td>16</td>
</tr>
<tr>
<td>139</td>
<td>17</td>
</tr>
<tr>
<td>132</td>
<td>18</td>
</tr>
<tr>
<td>127</td>
<td>19</td>
</tr>
<tr>
<td>120</td>
<td>20</td>
</tr>
<tr>
<td>119</td>
<td>21</td>
</tr>
<tr>
<td>116</td>
<td>22</td>
</tr>
<tr>
<td>115</td>
<td>23</td>
</tr>
<tr>
<td>113</td>
<td>24</td>
</tr>
<tr>
<td>112</td>
<td>25</td>
</tr>
<tr>
<td>108</td>
<td>26</td>
</tr>
<tr>
<td>104</td>
<td>27</td>
</tr>
<tr>
<td>103</td>
<td>28</td>
</tr>
<tr>
<td>102</td>
<td>29</td>
</tr>
<tr>
<td>101</td>
<td>30</td>
</tr>
<tr>
<td>98</td>
<td>31</td>
</tr>
<tr>
<td>90</td>
<td>32</td>
</tr>
<tr>
<td>86</td>
<td>33</td>
</tr>
<tr>
<td>84</td>
<td>34</td>
</tr>
<tr>
<td>81</td>
<td>35</td>
</tr>
<tr>
<td>80</td>
<td>36</td>
</tr>
<tr>
<td>78</td>
<td>37</td>
</tr>
<tr>
<td>76</td>
<td>38</td>
</tr>
<tr>
<td>74</td>
<td>39</td>
</tr>
<tr>
<td>72</td>
<td>40</td>
</tr>
<tr>
<td>67</td>
<td>41</td>
</tr>
<tr>
<td>65</td>
<td>42</td>
</tr>
<tr>
<td>60</td>
<td>43</td>
</tr>
<tr>
<td>59</td>
<td>44</td>
</tr>
<tr>
<td>57</td>
<td>45</td>
</tr>
<tr>
<td>51</td>
<td>46</td>
</tr>
<tr>
<td>50</td>
<td>47</td>
</tr>
<tr>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>45</td>
<td>49</td>
</tr>
<tr>
<td>44</td>
<td>50</td>
</tr>
<tr>
<td>43</td>
<td>51</td>
</tr>
<tr>
<td>42</td>
<td>52</td>
</tr>
<tr>
<td>41</td>
<td>53</td>
</tr>
<tr>
<td>40</td>
<td>54</td>
</tr>
<tr>
<td>39</td>
<td>55</td>
</tr>
<tr>
<td>38</td>
<td>56</td>
</tr>
<tr>
<td>37</td>
<td>57</td>
</tr>
<tr>
<td>36</td>
<td>58</td>
</tr>
<tr>
<td>35</td>
<td>59</td>
</tr>
<tr>
<td>33</td>
<td>60</td>
</tr>
<tr>
<td>32</td>
<td>61</td>
</tr>
<tr>
<td>31</td>
<td>62</td>
</tr>
<tr>
<td>30</td>
<td>63</td>
</tr>
<tr>
<td>29</td>
<td>64</td>
</tr>
<tr>
<td>28</td>
<td>65</td>
</tr>
<tr>
<td>27</td>
<td>66</td>
</tr>
<tr>
<td>26</td>
<td>67</td>
</tr>
<tr>
<td>25</td>
<td>68</td>
</tr>
<tr>
<td>24</td>
<td>69</td>
</tr>
<tr>
<td>23</td>
<td>70</td>
</tr>
<tr>
<td>22</td>
<td>71</td>
</tr>
<tr>
<td>21</td>
<td>72</td>
</tr>
<tr>
<td>20</td>
<td>73</td>
</tr>
<tr>
<td>19</td>
<td>74</td>
</tr>
<tr>
<td>18</td>
<td>75</td>
</tr>
<tr>
<td>17</td>
<td>76</td>
</tr>
<tr>
<td>16</td>
<td>77</td>
</tr>
<tr>
<td>15</td>
<td>78</td>
</tr>
<tr>
<td>14</td>
<td>79</td>
</tr>
<tr>
<td>13</td>
<td>80</td>
</tr>
<tr>
<td>12</td>
<td>81</td>
</tr>
<tr>
<td>11</td>
<td>82</td>
</tr>
<tr>
<td>10</td>
<td>83</td>
</tr>
<tr>
<td>9</td>
<td>84</td>
</tr>
<tr>
<td>4</td>
<td>85</td>
</tr>
</tbody>
</table>

---

Group profiles

Software for Health Technology Assessment: Vol. 3: No. 10

23
Descriptive methods for the analysis of longitudinal quality-of-life data

**Averages**

The mean or median quality-of-life scores in each group can be plotted over time (examples of means are given in O’Brien, et al., Figure 3 and in Anderson, et al., Figure 1; examples of medians are given by the MRC Lung Cancer Working Party, 1996, Figure 2). Bars at each time point representing 95% confidence intervals (CIs) for the mean or median should be included. This is the most useful way to represent continuous data. For ordinal data, the use of means is not theoretically correct but may be considered if the ordinal scale is long.

**Proportions**

If the data are binary then the proportion with a symptom or side-effect can be plotted over time. Bars at each time point representing 95% CIs for the proportion should be included. For ordinal data, the use of means is not theoretically correct but may be considered if the ordinal scale is long.

If the data are ordinal or continuous then the proportion reaching or exceeding a certain level of quality of life over time can be plotted (for example, see Figures 2 and 3 in the report of the MRC Lung Cancer Working Party, 1992, Figure 2 in their 1991 report, and Figure 3 in their 1996 report). Again, bars representing 95% CIs for the proportion should be included at each time point. Detail on the severity of the symptom or side-effect will be lost.

**Inclusion criteria for plot**

Summarising the quality of life of patients in each treatment group over time is complicated by the problem of missing data, especially when informative drop-out occurs. Cautious interpretation of the data is needed in these situations. There are several choices regarding which subjects to include in the calculation of the summary measure over time. Summary measures can be calculated and plotted either for the whole sample of patients in the study, or for the subgroup of patients who completed all assessments, or for subgroups of patients according to different lengths of follow-up and reasons for drop-out.

**All patients**

If the whole sample of patients is used then, because of drop-out due to death and illness,
the sample reduces over time and this should be made clear in any graph or table by specifying the number contributing to the descriptive measure at each time point. Interpretation of such data needs caution because a comparison of results at different time points compares different groups of patients. Subjects contributing to summary measures at later time points are likely to be the ‘healthier’ members of the original group. In situations in which quality of life is measured on a 0–1 scale, where 0 represents quality of life equivalent to death, a more realistic description may be obtained by including patients who drop out because of death, which gives them a quality-of-life score of 0.6

Complete case analysis

If the analysis is restricted to just those patients with complete data, then not only will it greatly reduce the sample size but it will also result in biased estimates, unless missing data is missing completely at random. Those included are not likely to be a representative subset of the overall sample, because patients who complete all assessments must be survivors and compliers and, hence, are probably a ‘healthy’ subgroup of the overall sample. The longer the follow-up period, the greater the level of attrition and the more reduced, and possibly more biased, the sample for analysis will be.

Subgroups according to differing lengths of follow-up and reasons for drop-out

Patients can be split into subgroups according to differing lengths of follow-up. Subgroups can be formed according to number of completed assessments, or a more detailed break-down based on different reasons for drop-out, such as censoring, death or lost to follow-up, could be used. Quality of life over time can be compared across subgroups to establish its association with the drop-out process. If there are no obvious differences between the subgroups then it may be valid to combine the data for analysis. A quality-of-life study needs a large number of participants for subgroups to contain adequate numbers of subjects for this type of analysis. This type of approach is used by de Stavola and Christensen in their method for dealing with informative drop-out in longitudinal data modelling (see chapter 8).

Group profiles for the MIC data

All patients

Summary measures for MQS and MAL in the MIC study were calculated for all patients in the study. The numbers of patients contributing to the summary measure at each time point have been specified in the graphs to highlight the problems of informative drop-out and aid cautious interpretation.

Treatment group means together with their associated 95% CIs are plotted over time for MQS (see Figure 8). A possibly naïve interpretation of these plots, ignoring the effect of missing data, may be as follows. MQS reduces slightly (i.e. quality of life improves) over the first three time points in the CT group and increases slightly (i.e. quality of life deteriorates) in the PAL group, with 95% CIs suggesting that the groups significantly differ at the third time point. After this point, the curves move back towards each other again and quality of life in both treatment groups seems comparable.

The proportions of patients with malaise (i.e. MAL = 1, 2 or 3) are plotted over time for each treatment group in the MIC study (see Figure 9). Bars representing 95% CIs for the proportions are included. This shows the extent of malaise in the two groups is fairly high (at about the 60% level) and reasonably comparable over time. In both groups, the extent of malaise reduces initially but then increases again. As with MQS, this interpretation could be questionable since it ignores the effect of patient drop-out.

Complete case analysis

Plots of MQS and malaise were plotted over time only for those patients who completed the first four questionnaires (see Figures 10 and 11). For MQS, only 43 patients on the CT arm had complete data, with 21 dropping out before completing the fourth

![Figure 8: Treatment group means (and 95% CIs) over time for MQS in the MIC study (---, CT; ---, PAL)](image-url)
questionnaire and three having intermittent missing questionnaires, while on the PAL arm, 25 had complete data, with 17 dropping out. For malaise, a further four patients on the CT arm and one on the PAL arm were excluded because they had missing values for malaise on one of the first four questionnaires. These plots contain data for those patients who did not drop out of the study and are thus likely to represent the ‘healthier’ patients from the MIC study. The plots, however, are not dissimilar to those for all patients (see Figures 8 and 9).

**Subgroups according to differing lengths of follow-up**

The patients in the MIC study (from both treatment arms) were split into subgroups according to the number of the last completed questionnaire (see Table 9). The mean MQS for each subgroup was plotted over time (see Figure 12). This shows that quality of life in those who dropped out after one questionnaire was worst at baseline and, in those who dropped out after two or three questionnaires, quality of life over time deteriorated slightly before dropping out. Quality of life in those who did not drop out during the first four questionnaires appears to be reasonably stable. Although the plot needs to be interpreted with caution, because the numbers in some of the subgroups are small, it suggests that there may be some association between drop-out and quality of life. This means that for a valid analysis of quality-of-life data, the drop-out process should probably not be ignored.
References


2. Beacon HJ. The statistical analysis of self-assessed quality of life data in cancer clinical trials [PhD thesis]. London School of Hygiene and Tropical Medicine, University of London; 1996.


Introduction

The methods discussed in this chapter are standard approaches to analysing longitudinal data and all assume that the mechanism giving rise to missing data is ignorable. The application of these methods to longitudinal quality-of-life data is problematic because, in this situation, the missing data are generally a result of informative drop-out and, hence, the missing data mechanism is non-ignorable. Methods discussed here should therefore be applied with caution to quality-of-life data, as they may result in invalid conclusions.

For continuous measures, such as those derived from a linear analogue scale or an aggregated global score, the appropriateness of parametric or non-parametric methods depends on the normality of the data distribution. It could be argued that since the continuous measures have truncated distributions, parametric methods will always be inappropriate. It may be possible to normalise a distribution using a transformation, such as log or square root, and then compare treatments using a parametric approach on the transformed variable.

Analysis of each time point separately

If quality of life is assessed at fixed time points during the study then the data can be analysed by considering each time point separately. At each time point, there may be three different research questions of interest.

1. To what extent do treatment groups differ in terms of quality of life at that time point?
2. To what extent do treatment groups differ in terms of change in quality of life relative to a previous time point (usually baseline)?
3. Is the change in quality of life between the given time point and a previous time point statistically significant?

All these research questions can be tackled using simple, standard statistical methods. The method of analysis depends on the type of variable used to measure quality of life (binary, ordinal, continuous non-normal or continuous normal) and the number of treatment groups being compared. Examples of quality-of-life studies, in which the data are analysed at each time point can be found in the literature.

There are three major problems with this simple approach to longitudinal data analysis:

- fixed time points are required
- the longitudinal nature of the data is ignored
- multiple analyses are involved, which may result in differences being regarded as significant by chance.

In addition to these problems, the approach is not generally suitable for quality-of-life data since it will not overcome the problem of informative drop-out. The patients at each time point will be a subset of the original sample of patients recruited to the study and, the further from baseline the time point, the more biased the sample will be, in that they will be the most ‘well’ patients from the original group.

Summary measures

Using summary measures, as promoted by Matthews and colleagues, is the simplest method for analysing longitudinal data. It reduces the repeated measures over time on an individual to a single summary measure, which can then be analysed using standard statistical methods, for example, for treatment group comparisons. The choice of summary measure needs to be clinically meaningful and will depend on the nature of the measure under investigation, together with the disease and treatment under study. In quality-of-life studies, the calculation of summary measures in patients who drop out may be difficult and if the drop-out process is informative, the analysis of the summary measure may give biased results.
**Simple summary measures for longitudinal quality-of-life data**

A wide range of summary measures could be chosen to represent longitudinal quality-of-life data. For example, the maximum score reached over time may be appropriate, or change in quality of life between two time points. Alternatively, the slope representing change over time for each individual could be computed.\(^7\)

Summary measures may be difficult to calculate when informative drop-out is present in the data; however, it may be possible to accommodate drop-outs into the analysis by imputing appropriate quality-of-life values. For example, in situations where quality of life is measured on a 0–1 scale, with 0 representing quality of life equivalent to death, time points with missing quality-of-life data caused by death may be allocated values of 0. This enables patients who drop out of the study because of death to be included in the analysis\(^7\) and may reduce bias caused by informative drop-out.

A standard summary measure used for longitudinal data is *area under the curve*. For each individual, the area under the curve of their repeated quality-of-life measures over time can be calculated and used as a summary measure.\(^7\)\(^8\) If patients do not have full follow-up then their summary measure will be censored, suggesting that methods of survival analysis may be appropriate. Censoring, however, will be informative and standard methods will give biased results. An alternative estimator has been proposed but is still biased.\(^7\) An outcome measure that combines quality-of-life and survival data, the QALY, is a special form of this summary measure and is discussed in detail later (see chapters 15 and 16).

Simple binary indicator summary measures can be used to define palliation.\(^10\)–\(^12\) For example, an indicator may be set to show whether an individual experienced a decrease in the level of severity of a particular symptom or in overall quality of life at any time compared with baseline or, alternatively, whether the individual experienced total disappearance of a symptom at any time. Other summary measures include duration of palliation and percentage of patient survival time during which there was palliation.\(^11\)\(^12\) These summary measures will only be valid if all patients have died, otherwise they should be restricted to a set follow-up time for which all patients in the study have been followed. Again, calculation of these summary measures may be problematic in the presence of informative drop-out.

**Time to the occurrence of a quality-of-life related event**

Longitudinal quality-of-life data can be summarised by a single value representing the time to the occurrence of a quality-of-life related event. The most widely used clinical endpoint, which in some situations will be quality-of-life related, is relapse-free survival, that is, the time from study entry to disease relapse. Nabholtz and colleagues\(^15\) used time to first occurrence of an important clinical adverse event or disease progression as a quality-of-life oriented endpoint. Rosenman and Choi\(^14\) used the Karnofsky index as a global measure of quality of life and used the time until first occurrence of a Karnofsky index of less than 60 as the summary measure on which to compare treatments. Hopwood and colleagues\(^15\) suggest time to first improvement or time to first worsening of quality of life from baseline as a summary measure, with times for patients not achieving such targets being censored. Time to palliation of various symptoms, in those patients with the symptom present pre-treatment, has also been used as a summary measure for treatment comparison.\(^16\)

Once a quality-of-life oriented endpoint has been defined, standard survival analysis techniques (see chapter 11) can be used to analyse the data. One advantage of this method is that patients who do not achieve the endpoint because of death or censoring would still be included in the analysis as a censored data point, thus dealing with the problem of informative drop-out. However, because the censoring mechanism may be related to the ‘survival’ time, standard survival analysis techniques may be invalidated by informative censoring (see page 11). Another advantage of this summary measure is that it can be used in situations where quality of life has been assessed at varying time points. However, if quality of life is assessed at only a few widely spaced time points, then the summary measure will be very crude.

A problem with this approach is that the potential for change, whether in terms of improvement or worsening, depends on the baseline value. For example, the worse a patient is at baseline, the greater the potential for improvement, so patients who do not experience a symptom at baseline are not able to improve and are thus excluded from the analysis. This will only be a problem in the comparison of treatments if the treatment groups differ with respect to baseline symptoms. The fact that a treatment may prevent a symptom from starting may be important and would not be assessed by this type of endpoint.
Repeated measures analysis of variance

A conventional repeated measures analysis of variance\(^\text{17}\) may be an appropriate statistical model to use in situations where quality of life is measured at fixed time points and will be most appropriate when the number of time points is reasonably small. The method assumes the data to be normally distributed, with the covariance matrix the same for all treatment groups. Generalisations on this simple model are possible.\(^\text{18}\) Using the model it can be determined if quality of life changes over time and if treatment groups vary in terms of changes in quality of life, while taking account of within-patient correlations.

Patient drop-out in quality-of-life studies results in missing measurements and causes the data to be unbalanced. One strategy to enable the use of repeated measures analysis of variance in this situation is to undertake a complete case analysis, in which treatment comparisons are made only in patients with complete data. Problems of reduced numbers for analysis and potential bias make this approach problematic and so an alternative strategy, which allows all available data to be included, may be preferable. This latter approach is based on the assumption that missing data are missing at random and, if this is not the case, then a method of analysis that allows for informative drop-out needs to be considered (see chapter 8).

Complete case analysis

A complete case analysis will not generally be applicable to quality-of-life data since it is only valid if missing data are missing completely at random. If there are very few subjects with complete data then a complete case analysis would be based on a very small sample. Even if the reduced sample is reasonably large, it is not likely to be representative of the whole sample. Those patients who survived for the duration of the study and were well enough throughout to complete questionnaires would generally be the ‘healthier’ participants and thus estimates of mean quality of life over time are likely to be biased (upwards). If the change in quality of life over time is independent of the level of quality of life, then, although estimates at each time will be biased, the estimates of the change may be unbiased.

In situations where quality of life is measured on a 0–1 scale, where 0 represents quality of life equivalent to death, patients who drop out because of death can be included in the analysis by giving them a quality-of-life score of 0.\(^\text{19}\) This will minimise the problem of reduced numbers and may provide more realistic conclusions.

Analysis based on assumption of missing data being missing at random

The analysis suggested by Zwinderman\(^\text{20}\) for quality-of-life data with missing measurements is based on the assumption that the missing data are missing at random, that is, non-response depends only on past measurements and not on future ones. This assumption enables the mean change pattern to be estimated using just the observed data.

The mean at time point, \(t + 1\), is estimated as:

\[
m_{t+1} = m_t + d_{t+1}
\]

where \(m_t\) is the estimated mean at time point \(t\) and \(d_{t+1}\) is the mean difference in measurements at time points \(t\) and \(t + 1\) for those patients who have measurements at both time points.

This method assumes that the change in quality of life over time is independent of the level of quality of life. Allen-Mersh and colleagues\(^\text{21}\) assessed longitudinal quality-of-life data using a repeated measures analysis of covariance based on Zwinderman’s approach.

Zwinderman\(^\text{20}\) discusses various computer programs in relation to this type of analysis. If the data are in the format of one record per patient containing all repeated measurements then BMDP5V\(^\text{22}\) automatically uses all available data. The GLM procedure in SAS/STAT\(^\text{23}\), however, cannot deal with the missing data and performs a complete case analysis only, unless the data are transformed so that they consist of one record per patient per repeated measure. Brooks and colleagues\(^\text{8}\) used BMDP5V to carry out a repeated measures analysis of variance on quality-of-life data and thus, intentionally or unintentionally, assumed missing data were missing at random.

More complex modelling techniques for longitudinal data

Models for longitudinal quality-of-life data examine the patterns of change over time and the factors, such as treatment, that may have some influence on these patterns. The modelling needs to account for the correlation among values within each subject which is present in longitudinal data. There is much literature on modelling longitudinal data\(^\text{20,24,25}\) and such

...
methods are considered here in terms of their application to quality-of-life data.

Longitudinal quality-of-life studies usually give rise to unbalanced data. This may be a result of assessment times varying from subject to subject or because of missing data. The impact of different types of missing data mechanisms on longitudinal data analysis has been discussed. Analysis of longitudinal quality-of-life data requires modelling techniques that both capture the dynamic nature of the data and cope with the unbalanced structure. Techniques also need to be able to handle all types of data, since quality-of-life measures could be binary, ordinal or continuous.

Some of the more complex modelling approaches that satisfy these requirements fall into two main categories: marginal models (sometimes called population-averaged models) and random effects models (sometimes called subject-specific models). Both approaches are based on the theory of generalised linear models and allow the analysis of binary, ordinal or continuous response variables. As such, the expected response is related to the explanatory variables in the model via a link function. The choice of link function depends on type of response variable; logit or probit link for binary responses, cumulative logit link for ordinal responses, log link for counts and identity link for continuous data.

In a marginal model, the regression coefficients estimate ‘population-averaged’ effects, that is, they give the estimated effect of explanatory variables on the response of the population as a whole. Random effects models allow each individual to have their own regression coefficients; these subject-specific effects are assumed to have been drawn from an overall population distribution, which is often the primary focus of the analysis.

Another difference between the two classes of model is the way in which they incorporate within-subject correlation. In the marginal modelling approach, the effect of explanatory variables on the marginal distributions of response and the within-subject correlation are modelled separately, with the latter treated as a nuisance factor. In the random effects approach, within-subject correlation is accommodated by incorporating a subject-specific random effect into the model.

For normally distributed data, parameter estimates can be obtained by using maximum likelihood methods. Laird and Ware computed maximum likelihood estimates in random effects models using the EM algorithm. An alternative for parameter estimation is to use a Bayesian approach using Gibbs sampling.

For categorical responses, full likelihood estimation of regression parameters is generally not feasible. One alternative is to use a weighted least-squares approach. There are two main limitations to this approach: large numbers of subjects are required and continuous explanatory variables cannot be accommodated. Another alternative is the generalised estimating equations method, developed for marginal models and based on the concept of quasi-likelihood. Generalised estimating equations offer one way of estimating regression parameters together with their variances, while taking within-subject correlation into account, and can be used for both marginal and random effects models. Bayesian analysis can also be used to estimate parameters.

Hierarchical or multilevel models are becoming a standard method in the analysis of longitudinal data where the lowest level of the hierarchy consists of the observations over time within each subject and second level units are the subjects. They have been applied to quality-of-life data.

Modelling techniques for longitudinal data have been discussed in general, in relation to normally distributed responses, in relation to non-normally distributed responses, and in relation to categorical responses. Software for implementing the sort of modelling techniques discussed here are available, such as PROC MIXED in SAS/STAT and MLn. Examples of their application to quality-of-life data is limited, possibly because of their complexity. The main problem with applying these models to quality-of-life data is that they assume the missing data mechanism is ignorable. In particular, the generalised estimating equations approach requires missing data to be missing completely at random. When there is informative drop-out, standard modelling techniques will give biased parameter estimates and modelling techniques that allow for informative drop-out should be considered (see chapter 8).

References


36. Beacon HJ. The statistical analysis of self assessed quality of life data in cancer clinical trials [PhD thesis]. London School of Hygiene and Tropical Medicine, University of London; 1996.


Introduction

In longitudinal studies of quality of life with informative drop-out, the standard methods of longitudinal data analysis discussed in chapter 7 are frequently invalid, since they assume that the missing data mechanism is ignorable. Using standard methods in situations of informative drop-out may give biased results. A variety of methods for analysing longitudinal data in the presence of informative drop-out have been proposed but there is little evidence of their application to quality-of-life data. This chapter provides a brief overview of such methods.

Methods

A variety of methods have been proposed for modelling the change in a continuous variable over time while accounting for informative drop-out, and Little\textsuperscript{1} provides a review of such methods. Methods are generally based on random effects models, although an approach based on marginal models has also been considered.\textsuperscript{2} The techniques focus on a classical approach but a Bayesian approach to longitudinal data models with informative drop-out has been discussed.\textsuperscript{3}

Most methods are developed from a linear random effects model,\textsuperscript{4,6} in which it is assumed that each subject’s measurements follow a linear regression with random intercept and slope. The average rate of change can be estimated using likelihood-based approaches or using either weighted or unweighted averages of the individual least-squares slope estimates. In the presence of informative drop-out, likelihood-based approaches and weighted averages have been shown to be seriously biased and, although unweighted averages were generally found to be unbiased, they are too inefficient to merit consideration.\textsuperscript{4,6,9} Thus, informative drop-out needs to be accounted for when modelling longitudinal quality-of-life data with such data missing.

Assuming a linear random effects model for the response variable over time, Wu and Carroll\textsuperscript{4} model the relationship between the probability of drop-out and the path of the response variable, as described by the intercept and slope parameters, using a probit model. Parameters were estimated using pseudo-maximum likelihood. They suggest that Cox proportional hazards and logistic models could also be appropriate.

Wu and Bailey\textsuperscript{5,6} developed methods that account for informative drop-out without actually modelling the drop-out process. They proposed a conditional linear model for the individual least-squares estimated slopes, assuming them to be linear functions of total follow-up time. Two methods for estimating the population rate of change, based on a weighted least-squares approach were suggested and compared to pseudo-maximum likelihood estimates; linear minimum variance unbiased estimates and linear minimum mean squared error estimates. Mori and colleagues\textsuperscript{7} developed the method further by using empirical Bayes methodology to estimate rate of change, enabling estimation of individual as well as population slopes.

The problem with the probit drop-out model approach and the conditional linear model approach is that at least two measurements per subject are needed to enable least-squares estimation of individual slopes. A further limitation of the conditional linear model approach is that the method assumes the potential follow-up time is the same for all subjects and thus application of such methods to studies with staggered patient entry times may be invalid.

Schluchter\textsuperscript{8} developed the conditional linear model approach to allow for staggered patient entry and individuals with single measurements. The method assumes that the individual intercept, slope and log survival time follow a trivariate normal distribution and the maximum likelihood estimates of model parameters are calculated using the EM algorithm. Disadvantages of this approach are that it is computationally complex and, hence, not easily accessible and that large amounts of data may be required. A multilevel version of Schluchter’s model has been applied
Longitudinal modelling methods that deal with informative drop-out
to quality-of-life data.\textsuperscript{10} Related methods based on
general rather than linear random effects models
have also been proposed\textsuperscript{11,12} but these suffer from
the same limitations.

The method of multilevel modelling has been
adapted to deal with informative drop-out by
analysing sequentially overlapping portions of
the follow-up information.\textsuperscript{13} This is comparable
to the descriptive method discussed earlier in
which patients are split into subgroups according
to differing lengths of follow-up (see page 25).
A set of sequential time points, $T_1, T_2, \ldots, T_3, \ldots, T_K$,
which are meaningful to the study are determined.
All those patients who are still in the trial at time
$T_i$ form the risk set for that time point and the
follow-up information for each of those patients
consists just of the data up to time $T_i$. In this way,
overlapping risk sets are created with patients
with long overall follow-up contributing to most
risk sets while patients who dropped out early
contribute to the first few only. Multilevel models
are fitted to each of the $K$ risk sets. An overall
multilevel model, including a third level repre-
senting the last risk set to which a patient belonged,
can be fitted to evaluate the extent of informative
drop-out in the data and provide a direct com-
parison of time profiles of patients with different
follow-up times.

Discussion

Longitudinal data models that deal with informative
drop-out may be appropriate methods for
analysing quality-of-life data but there is little
evidence of their application in this field. The
proposed methods have focused on continuous
responses and thus may not be applicable to
quality-of-life measures that are categorical or
highly skewed continuous variables. De Stavola
and Christensen\textsuperscript{13} claim that their multilevel
model approach can include categorical variables.
Beacon\textsuperscript{10} suggests that the conditional linear
models proposed by Wu and Bailey\textsuperscript{5,6} and, in
particular, the trivariate normal model proposed by
Schluchter\textsuperscript{8} are the most suitable for application
to quality-of-life data. Further work is needed to
fully investigate the use of these methods in
quality-of-life assessment, especially in relation
to the application to categorical responses.
Although the methods discussed here allow for
informative drop-out, simultaneous modelling
methods, discussed later in chapter 19, in which
both the quality-of-life and the drop-out process
are explicitly modelled, may provide a potentially
more powerful approach.

References
1. Little RJA. Modelling the drop-out mechanism
in repeated measures studies. J Am Stat Assoc
2. Diggle P, Kenward MG. Informative drop-out in
3. Best NG, Spiegelhalter DJ, Thomas A, Brayne CEG.
Bayesian analysis of realistically complex models.
4. Wu MC, Carroll RJ. Estimation and comparison
of changes in the presence of informative right
censoring by modelling the censoring process.
5. Wu MC, Bailey KR. Estimation and comparison
of changes in the presence of informative right
censoring: conditional linear model. Biometrics
6. Wu MC, Bailey K. Analysing changes in the presence
of informative right censoring caused by death and
7. Mori M, Woodworth GG, Woolson RF. Application
of empirical Bayes inference to estimation of rate
of change in the presence of informative right
8. Schluchter MD. Methods for the analysis of
informatively censored longitudinal data.
9. 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
10. Beacon HJ. The statistical analysis of self assessed
quality of life data in cancer clinical trials [PhD
dissertation], London School of Hygiene and Tropical
Medicine, University of London; 1996.
11. de Gruttola V, Tu XM. Modelling progression of
CD4-lymphocyte count and its relationship to
12. Pawitan Y, Self S. Modelling disease marker
13. de Stavola BL, Christensen E. Multilevel models
for longitudinal variables prognostic for survival.
Lifetime Data Anal 1996;\textbf{2}:529–47.
In quality-of-life studies where survival is also an issue, patients are often severely ill, and drop-out caused by illness or death is a common occurrence. This drop-out process may be informative but the situation may be complex, involving drop-outs of several types. Standard analytical methods for longitudinal data assume that the missing data mechanism is ignorable, and thus use of these methods for longitudinal quality-of-life data may give biased results and invalid conclusions. However, if the drop-out rates and reasons for drop-out are balanced across treatment arms then between-treatment comparisons may remain unbiased.1,2

The analysis of longitudinal quality-of-life data should begin descriptively to give insight to the data before any formal testing or modelling is carried out. Graphs of individual patient profiles allow data to be examined at the most basic level, while treatment group profiles enable a clearer comparison of treatments in relation to their effect on quality of life over time. Interpretation of group profiles needs caution when informative drop-out is present, as group summary measures may well be biased.

An initial approach to the formal analysis is to use summary measures, in which treatment comparison is based on a single value summarising the quality-of-life data over time. The method enables the use of simple standard statistical techniques and thus has the advantage of being easy to apply. It is also reasonably flexible in terms of the type of data to which it can be applied and gives results that are clinically meaningful. In replacing a set of repeated measures with a single summary measure, however, the method does not fully capture the dynamic nature of quality-of-life data. Furthermore, calculation of summary measures may be difficult and analysis may be biased if informative drop-out is present.

If quality-of-life data are collected at a reasonably small number of fixed time points and have an underlying normal distribution, then a simple repeated measures analysis of variance may be appropriate. Strategies that enable application of the method to quality-of-life data with missing measurements are available3 but these methods assume the missing data mechanism is ignorable, which will not of course be the case if drop-out is informative.

More complex modelling techniques, such as random coefficient and marginal models, allow greater flexibility in terms of being able to cope with unbalanced data, varying time points and non-normality, but their application to quality-of-life data may be difficult and problematic because of the complexity of the models.4 Examples of the application of more complex modelling techniques to quality-of-life data is limited.5–7 Although these techniques model change over time, they do not explicitly model the missing data mechanism and thus are not valid when informative drop-out is present.

Methods which assume data are missing at random, rather than missing completely at random,5,7 go some way to dealing with the missing data problem in longitudinal quality-of-life studies but will still give biased results in situations of informative drop-out. There are no easy ways to test if missing data are missing at random or are informative, and assumptions have to be based on subjective judgement. The assumptions should be assessed as part of a sensitivity analysis.

Modelling techniques that deal with informative drop-out have been developed but are complex and focus on continuous response data. The practical application of such methods may be difficult since appropriate software is not readily available. The application of such methods to quality-of-life data has, as yet, not been fully investigated and further work is needed to assess their use in this context. Simultaneous modelling methods (discussed later in chapter 19) which explicitly model both the quality-of-life and the drop-out process may provide an approach that is potentially more powerful.

References


6. Beacon HJ. The statistical analysis of self assessed quality of life data in cancer clinical trials [PhD thesis]. London School of Hygiene and Tropical Medicine, University of London; 1996.

In many studies, especially cancer clinical trials, treatments are assessed primarily in terms of survival, usually defined as time from entry to trial until death. Survival is the main endpoint in studies evaluating potentially curative treatments or treatments which are expected to prolong life significantly. There may also be an interest in quality of life as a secondary endpoint, especially in situations where treatments have side-effects.

In comparing treatments in terms of survival, it is often necessary to adjust for patient-related factors that could potentially affect the survival time of a patient. These covariates may include demographic variables such as age and sex, physiological variables such as white blood cell count, or disease-related variables such as type of tumour or stage of disease at entry to the study. One such patient-related factor to consider as a covariate may be quality of life, either in terms of a baseline measure or in terms of its changing values over time. In chapters 11–13 methods of analysing quality-of-life and survival data are dealt with, in which the aim of the study is to compare treatments in terms of survival while adjusting for quality of life. The analysis also allows the effect of quality of life on survival (that is, as a predictor) to be explored.

In chapter 11, some background to the analysis of survival data is given and the standard methods of analysis used to compare treatments in terms of survival without any covariate adjustment are outlined. In chapter 12, methods of analysis are discussed which enable a comparison of treatments while adjusting for the effects of covariates. Inclusion of both fixed and time-dependent covariates are considered and the particular application of these methods to situations with quality of life as a covariate is also discussed. The findings of this section of the review are summarised and discussed in chapter 13.
**Chapter I I**

**Comparison of treatments in terms of survival**

**Introduction to survival analysis**

Survival data differs from other types of continuous data in that the actual survival time may not necessarily be observed for all subjects in the study, thus giving rise to censored data (see chapter 3). Any methods used to analyse survival data must be able to deal with censored data.

There are many books and papers that discuss statistical methods for survival analysis. Parmar and Machin\(^1\) provide non-technical coverage, while Collett\(^2\) and especially Kalbfleisch and Prentice\(^3\) provide the reader with more theoretical detail.

**Survivor and hazard functions**

Survival data is generally described and modelled in terms of two related functions, the survivor function and the hazard function. The survivor function, \(S(t)\), represents the probability that an individual survives from the time origin to some time beyond \(t\), and is given by:

\[
S(t) = \Pr(T > t) = 1 - F(t) = 1 - \int_0^t f(u)\,du \tag{3}
\]

where \(T\) is the random variable representing survival time, and the distribution of survival times is described by \(f(t)\), the probability density function, and \(F(t)\), the associated cumulative distribution function.

The hazard function, \(h(t)\), is the probability that an individual dies at time \(t\), given that they have survived up to that time. It represents the instantaneous death rate for an individual surviving to time \(t\). It is linked to the probability density function and survivor function by:

\[
h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} \log S(t) \tag{4}
\]

The survivor function and hazard function can be estimated from observed data. If the form of \(f(t)\) is not specified then non-parametric procedures can be used, otherwise parametric models can be fitted to the data.

**Some parametric forms for survival data**

In some situations it may be appropriate to assume a distribution for \(f(t)\), the probability density function of the survival time. The most common distributions used to model survival data are the exponential and Weibull distributions, the exponential being a special form of Weibull distribution. Only these two distributions are considered in this report but, in general, other distributions such as log-normal, log-logistic and gamma may be more appropriate.\(^2\)

If survival times have an exponential distribution then

\[
f(t) = \lambda \exp(-\lambda t) \tag{5}
\]

the survivor function is given by

\[
S(t) = \exp(-\lambda t) \tag{6}
\]

and the hazard function is given by

\[
h(t) = \lambda \tag{7}
\]

Thus an exponential distribution assumes that the hazard rate is constant over time.

If the constancy of the hazard rate is not a valid assumption, then a Weibull distribution may be a more appropriate distribution for survival times. In this case,

\[
f(t) = \lambda \gamma t^{\gamma-1} \exp(-\lambda t^\gamma) \tag{8}
\]

the survivor function is given by

\[
S(t) = \exp(-\lambda t^\gamma) \tag{9}
\]

and the hazard function is given by

\[
h(t) = \lambda \gamma t^{\gamma-1} \tag{10}
\]

The parameters \(\gamma\) and \(\lambda\) determine the shape and scale of the hazard function and are thus called the shape and scale parameters, respectively. In the special case of \(\gamma = 1\), the distribution of survival times is exponential and the hazard is constant. If \(\gamma > 1\), the hazard increases with time and if \(0 < \gamma < 1\) then the hazard decreases monotonically.
The suitability of these parametric models can be assessed using a log-cumulative hazard plot. This is a graph of \( \log(-\log[S(t)]) \) against \( \log t \) where \( S(t) \) is the survival function estimated by the Kaplan–Meier method (see below). If a Weibull model is appropriate, the line will be approximately straight, with the slope giving an estimate of the shape parameter for the distribution. If an exponential model is appropriate, the slope of the line will be approximately 1.

The parameters for a model are estimated from the survival data using the method of maximum likelihood. In SAS/STAT, the LIFEREG procedure fits parametric models to survival data.

### Comparison of treatments using Kaplan–Meier estimates

The survivor function can be estimated non-parametrically from observed data, both censored and uncensored, using the Kaplan–Meier method. This method is also called the product-limit method and is based on maximum likelihood estimation. Suppose deaths occur at times \( t_1 < t_2 < ... < t_j < ... < t_n \), then the Kaplan–Meier estimate of the survivor function is given by

\[
\hat{S}(t) = \prod_{j=1}^{k} \left(1 - \frac{d_j}{n_j}\right) \text{ for all } t_k \leq t
\]

where \( n_j \) is the number of individuals alive just before time \( t_j \) and \( d_j \) is the number of deaths at time \( t_j \). Survival times censored at time \( t_j \) are assumed to occur immediately after the death time when computing values of \( n_j \). CIs for the survivor function can be calculated using a variety of different methods.

The calculation of Kaplan–Meier estimates is based on the assumption that the deaths of individuals in the sample occur independently of one another. This allows the probabilities of surviving from one interval to the next to be multiplied together to give the survivor function. It should also be noted that the Kaplan–Meier method gives the maximum likelihood estimate of the survivor function only if deaths and censoring are independent. Thus, for unbiased Kaplan–Meier estimates, it is necessary for the censoring mechanism to be non-informative.

Kaplan–Meier estimates of the survivor function \( S(t) \) can be plotted against time \( t \) as a survival curve. The survival curve is a stepped plot with the survivor function dropping instantaneously at each time of death and remaining level between successive death times. These provide a useful summary of the data and can be used to determine summary statistics such as median survival time. Survival curves for different treatment groups can be used to compare treatments descriptively in terms of survival. A more formal comparison of survival curves can be made using various non-parametric tests (see page 43). The SAS/STAT LIFETEST procedure calculates Kaplan–Meier estimates and plots survival curves.

### Survival analysis of the MIC data

Kaplan–Meier estimates of the survivor function for each treatment group in the MIC study were calculated for just those patients in the quality-of-life study and plotted as survival curves (Figure 13). Summary statistics for each treatment group were calculated from these survival curves (Table 10).

100
75
50
25
0
% surviving

<table>
<thead>
<tr>
<th>Numbers at risk</th>
<th>Time (months)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT 67</td>
<td>0</td>
</tr>
<tr>
<td>PAL 42</td>
<td>0</td>
</tr>
</tbody>
</table>

FIGURE 13 Overall survival for patients in the MIC quality-of-life study (——, CT; - - - , PAL)

<table>
<thead>
<tr>
<th>TABLE 10 Summary statistics of survival for the MIC study</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT arm ((n = 67))</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>Number of deaths</td>
</tr>
<tr>
<td>Median (95% CI)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>% surviving 1 year</td>
</tr>
<tr>
<td>(95% CI)</td>
</tr>
</tbody>
</table>
For those individuals in the quality-of-life study, survival appeared to be better in the CT group than in the PAL group.

The 18-week period from entry to trial has been selected as the time on which the analysis will focus, since it was during this time that quality-of-life data was collected (see page 14). Survival within 18 weeks of entry to the trial can be examined using Kaplan–Meier estimates. The survival curves of each treatment group were plotted (Figure 14). These are just the overall survival curves (from Figure 13) magnified and cut-off at 18 weeks to highlight that period. The summary statistics for survival were taken from the curves (Table 11). Within the first 18 weeks of entry to trial, the CT group appeared to have a better survival than the PAL group.

Non-parametric tests for comparing survival

Survival in two or more groups of patients can be compared using a non-parametric test such as the log-rank test, also called the Mantel–Cox test. This is the most widely used method of comparing survival curves.

The method essentially calculates at each death time, for each treatment group, the expected number of deaths under the null hypothesis of no difference between groups. These are then summed to give the total expected number of deaths in each treatment group, say \( E_i \), for treatment group \( i \). The log-rank test compares the observed number of deaths in each treatment group, say \( O_i \), for treatment group \( i \), to the expected number by calculating the test statistic

\[
X^2 = \sum_{i=1}^{g} \frac{(O_i - E_i)^2}{E_i}
\]  

and comparing it to a chi-square distribution with \((g - 1)\) degrees of freedom, where \( g \) is the number of treatment groups.

In the situation where two groups are being compared, the log-rank test is testing the null hypothesis that the ratio of the hazard rates in the two groups is equal to 1. The hazard ratio is a measure of the relative survival experience in the two groups and is estimated by

\[
HR = \frac{O_1 / E_1}{O_2 / E_2}
\]

where \( O_i / E_i \) is the estimated hazard rate in group \( i \). A CI for the hazard ratio can be calculated. Other non-parametric tests which are sometimes used to compare groups in terms of survival are the Mantel–Haenszel test and the Wilcoxon test. These are more suitable than the log-rank test when the assumption of proportional hazards is not valid for the alternative hypothesis. The assumption of proportional hazards can be assessed by means of a log-cumulative hazard plot (see page 46).

Survival analysis of the MIC data

Hazard ratios were calculated and log-rank tests performed to compare the two treatment groups in the MIC study in terms of both survival and survival within 18 weeks (Table 12). There was strong evidence to suggest that chemotherapy...
Comparison of treatments in terms of survival

reduced the risk of death compared with standard palliative treatment both overall and within 18 weeks of entry to trial.

### References


---

**TABLE 12** Hazard ratios and log-rank test for survival and survival within 18 weeks for the MIC study

<table>
<thead>
<tr>
<th></th>
<th>CT vs PAL hazard ratio (95% CI)</th>
<th>Results from log-rank test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survival</td>
<td>0.56 (0.37, 0.86)</td>
<td>$\chi^2 = 9.06, p = 0.0026$</td>
</tr>
<tr>
<td>Survival within 18 weeks</td>
<td>0.40 (0.20, 0.79)</td>
<td>$\chi^2 = 8.07, p = 0.0045$</td>
</tr>
</tbody>
</table>
Chapter 12
Comparison of treatments in terms of survival adjusting for covariates

Introduction

In comparing treatments in terms of survival, it is often necessary to adjust for patient-related factors, known as covariates, that could potentially affect the survival time of a patient. Covariates may be continuous measures, such as tumour size, or they may be ordinal variables, such as, for example, performance status rated as poor, fair, and good. They may also be binary measures, such as the presence/absence of a symptom.

Covariates that keep the same value for the duration of a study are called fixed or time-independent, while those with possibly changing values over time are called time-dependent. Fixed covariates can be adjusted for in a survival analysis either using stratified survival analysis or using a Cox proportional hazards model. Time-dependent covariates can also be adjusted for in a Cox model.

Stratified survival analysis

The simplest way to incorporate covariates into a survival analysis is to use a stratified survival analysis. This compares survival between treatment groups within each level of a covariate. The log-rank test methodology can be used to test for any differences between treatments, and comparisons of treatments within each stratum are combined to give an overall comparison of treatments that has been adjusted for the effect of the covariate.

Stratified survival analysis has two main limitations. Firstly, the method is only applicable to fixed covariates and, secondly, the covariate has to be in a categorical format with, in practice, very few levels. Thus it may be preferable, especially when several variables are involved, to use a Cox model (see page 46).

For situations with quality of life as a covariate, a stratified survival analysis could be used if the quality-of-life measures are available at the start of the analysis and remain unchanged throughout, such as, for example, baseline measures. The measure of quality of life may need to be categorised so that it consists of a few levels containing a reasonable number of subjects.

Stratified survival analysis of the MIC data

The overall survival and survival within 18 weeks were compared for both treatment groups in the MIC study using a stratified survival analysis. Stratification according to two different measures of baseline quality of life taken from the first questionnaire were considered. Using MAL, baseline quality of life was categorised as no malaise (MAL = 0) and malaise (MAL = 1, 2 or 3), and using MQS, baseline quality of life was categorised as above and below the median value of MQS (i.e. MQS ≥ 0.75 and MQS < 0.75).

Although baseline malaise does not appear to affect survival (see Table 13), there is evidence of a relationship between the baseline values of MQS and overall survival (see Table 14). The hazard rates suggest that the risk of death is reduced if the baseline value of MQS is below the median, that is, if the baseline quality-of-life measure is ‘good’. In a comparison of treatments,

<table>
<thead>
<tr>
<th></th>
<th>( n )</th>
<th>Observed deaths</th>
<th>Expected deaths</th>
<th>Hazard rate (observed/expected)</th>
<th>( \chi^2 )</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survival</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No malaise at baseline</td>
<td>42</td>
<td>41</td>
<td>41.48</td>
<td>0.99</td>
<td>0.009</td>
<td>0.92</td>
</tr>
<tr>
<td>Malaise at baseline</td>
<td>67</td>
<td>66</td>
<td>65.52</td>
<td>1.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Survival within 18 weeks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No malaise at baseline</td>
<td>42</td>
<td>16</td>
<td>14.51</td>
<td>1.10</td>
<td>0.25</td>
<td>0.62</td>
</tr>
<tr>
<td>Malaise at baseline</td>
<td>67</td>
<td>21</td>
<td>22.49</td>
<td>0.93</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Comparison of treatments in terms of survival adjusting for covariates

stratification by these covariates does not greatly affect the results of the log-rank tests, with \( p \)-values from the stratified log-rank tests being similar to those from the unstratified test (see Table 15).

**Cox proportional hazards model with fixed covariates**

**Background**

Survival data can be modelled using the Cox proportional hazards model.\(^1\) This is a regression model suitable for survival data, which allows the variation in survival, as expressed by the hazard function, to be explained by certain factors. By incorporating treatment with other covariates in the model, the differences in survival between treatment groups can be investigated while adjusting for the other covariates. Here the model is considered when covariates are fixed but the model can be extended to include time-dependent covariates (see page 47). The variables in the model can be continuous or categorical – in the latter case they take the form of dummy variables.

In the Cox proportional hazards model, the hazard function \( h(t) \) is modelled as:

\[
h(t) = h_0(t) \exp (\beta^T x)
\]

where \( h_0(t) \) is the underlying baseline hazard function, \( x \) is a vector of fixed covariates and \( \beta \) is a vector of regression coefficients. Parameters are estimated by maximising a partial likelihood function.\(^2\)

The model does not assume any particular form of probability distribution for the survival times and thus the underlying baseline hazard is allowed to be arbitrary. It does, however, have one underlying assumption, that of proportionality of hazards and, because of this restriction, the model is usually referred to as semi-parametric. If the proportional hazards assumption is valid, the ratio of the hazards in the subgroups defined by the covariate values remains approximately constant over time.

The assumption of proportional hazards can be assessed by means of a log-cumulative hazard plot where, using Kaplan–Meier estimates of the survivor function, \( \log(-\log S(t)) \) is plotted against \( \log t \) for each subgroup. If the assumption of proportional hazards is valid, the graphs will be parallel.\(^3\)

Nested models can be compared by checking the change in the value of \(-2 \log (\text{likelihood})\) against a chi-square distribution with degrees of freedom equal to the difference in the number of parameters being estimated.\(^3\) Most standard statistical software packages, including the PHREG procedure in SAS/STAT software,\(^4\) can be used to fit Cox regression models with fixed covariates.

**Application to quality-of-life data**

The quality of life of a patient on entry to a study may be related to the survival of that patient and could be included in a survival model as a fixed covariate. Including baseline quality of life, together with treatment as covariates in the model, enables the difference in survival between treatment groups to be assessed, while accounting for the baseline effect of quality of life. Measures of quality of life may be continuous or categorical.

---

**TABLE 14** Survival analysis of MIC data by baseline MQS

<table>
<thead>
<tr>
<th>Survival</th>
<th>( n )</th>
<th>Observed deaths</th>
<th>Expected deaths</th>
<th>Hazard rate (observed/expected)</th>
<th>( \chi^2 )</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MQS &lt; 0.75</td>
<td>50</td>
<td>48</td>
<td>59.74</td>
<td>0.80</td>
<td>5.38</td>
<td>0.02</td>
</tr>
<tr>
<td>MQS ( \geq 0.75 )</td>
<td>59</td>
<td>59</td>
<td>47.26</td>
<td>1.25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Survival within 18 weeks</th>
<th>( n )</th>
<th>Observed deaths</th>
<th>Expected deaths</th>
<th>Hazard rate (observed/expected)</th>
<th>( \chi^2 )</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MQS &lt; 0.75</td>
<td>50</td>
<td>14</td>
<td>18.35</td>
<td>0.76</td>
<td>2.05</td>
<td>0.15</td>
</tr>
<tr>
<td>MQS ( \geq 0.75 )</td>
<td>59</td>
<td>23</td>
<td>18.65</td>
<td>1.23</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 15** Stratified and unstratified survival analysis for treatment comparison in the MIC study (new results in bold)

<table>
<thead>
<tr>
<th>Survival</th>
<th>( \chi^2 )</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unstratified</td>
<td>9.06</td>
<td>0.0026</td>
</tr>
<tr>
<td>Stratified by baseline malaise</td>
<td>8.70</td>
<td>0.0032</td>
</tr>
<tr>
<td>Stratified by baseline MQS</td>
<td>8.54</td>
<td>0.0035</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Survival within 18 weeks</th>
<th>( \chi^2 )</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unstratified</td>
<td>8.07</td>
<td>0.0045</td>
</tr>
<tr>
<td>Stratified by baseline malaise</td>
<td>8.09</td>
<td>0.0044</td>
</tr>
<tr>
<td>Stratified by baseline MQS</td>
<td>8.25</td>
<td>0.0041</td>
</tr>
</tbody>
</table>
Cox regression analysis of MIC data with quality of life as fixed covariate

A Cox model was fitted to the MIC survival data with two fixed covariates, treatment and baseline quality of life. Treatment was included as a binary variable with 1 representing chemotherapy and 0 representing standard palliative care. As with the stratified survival analysis (see page 45), two different measures of baseline quality of life taken from the first questionnaire were included. Using MAL as before, quality of life was included as a binary variable with 0 representing no malaise (MAL = 0) and 1 representing malaise (MAL = 1, 2 or 3). MQS was used as an alternative measure of quality of life and was included in the model as a continuous variable, rather than as a categorical variable in the stratified analysis.

The regression coefficients and hazard ratios for the quality-of-life covariates suggest that there is no relationship between baseline malaise and survival (see Table 16) but there is evidence of a relationship between baseline values of MQS and overall survival (see Table 17). Hazard rates suggest that increasing values of baseline MQS (i.e. worsening quality of life) are associated with an increased hazard of death. These results are comparable to the stratified analysis (see page 45). Regression coefficients and hazard ratios (see Tables 16 and 17) together with likelihood ratio tests (see Table 18) show that after adjusting for baseline values of malaise and MQS, there is still a significant treatment effect on both survival and survival within 18 weeks. The results show that chemotherapy reduces the hazard of death.

### Table 16: Cox regression analysis with baseline malaise as fixed covariate

<table>
<thead>
<tr>
<th></th>
<th>( \beta ) (95% CI)</th>
<th>Hazard ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Survival</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malaise</td>
<td>0.04 (–0.35, 0.43)</td>
<td>1.04 (0.70, 1.54)</td>
</tr>
<tr>
<td>Treatment (adjusted for malaise)</td>
<td>–0.60 (–1.00, –0.20)</td>
<td>0.55 (0.37, 0.82)</td>
</tr>
<tr>
<td><strong>Survival within 18 weeks</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malaise</td>
<td>–0.15 (–0.80, 0.50)</td>
<td>0.86 (0.45, 1.65)</td>
</tr>
<tr>
<td>Treatment (adjusted for malaise)</td>
<td>–0.91 (–1.56, –0.26)</td>
<td>0.40 (0.21, 0.77)</td>
</tr>
</tbody>
</table>

### Table 17: Cox regression analysis with baseline MQS as fixed covariate

<table>
<thead>
<tr>
<th></th>
<th>( \beta ) (95% CI)</th>
<th>Hazard ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Survival</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MQS</td>
<td>0.79 (0.36, 1.22)</td>
<td>2.20 (1.43, 3.40)</td>
</tr>
<tr>
<td>Treatment (adjusted for MQS)</td>
<td>–0.55 (–0.95, –0.15)</td>
<td>0.58 (0.39, 0.86)</td>
</tr>
<tr>
<td><strong>Survival within 18 weeks</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MQS</td>
<td>0.45 (–0.15, 1.05)</td>
<td>1.56 (0.86, 2.85)</td>
</tr>
<tr>
<td>Treatment (adjusted for MQS)</td>
<td>–0.85 (–1.51, –0.19)</td>
<td>0.43 (0.22, 0.83)</td>
</tr>
</tbody>
</table>

### Table 18: Cox regression analysis with fixed covariates compared with stratified and unstratified survival analysis for treatment comparison in the MIC study (new results in bold)

<table>
<thead>
<tr>
<th></th>
<th>( \chi^2 )</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Survival</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unstratified</td>
<td>9.06</td>
<td>0.0026</td>
</tr>
<tr>
<td>Stratified by baseline malaise</td>
<td>8.70</td>
<td>0.0032</td>
</tr>
<tr>
<td>Stratified by baseline MQS</td>
<td>8.54</td>
<td>0.0035</td>
</tr>
<tr>
<td>Adjusted for baseline malaise</td>
<td>8.40</td>
<td><strong>0.0038</strong></td>
</tr>
<tr>
<td>Adjusted for baseline MQS</td>
<td>6.86</td>
<td><strong>0.0088</strong></td>
</tr>
<tr>
<td><strong>Survival within 18 weeks</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unstratified</td>
<td>8.07</td>
<td>0.0045</td>
</tr>
<tr>
<td>Stratified by baseline malaise</td>
<td>8.09</td>
<td>0.0044</td>
</tr>
<tr>
<td>Stratified by baseline MQS</td>
<td>8.25</td>
<td>0.0041</td>
</tr>
<tr>
<td>Adjusted for baseline malaise</td>
<td>7.53</td>
<td><strong>0.0061</strong></td>
</tr>
<tr>
<td>Adjusted for baseline MQS</td>
<td>6.41</td>
<td><strong>0.0113</strong></td>
</tr>
</tbody>
</table>

Cox model with time-dependent covariates

**Background**

A time-dependent covariate is a variable which may explain survival differences; its value for any individual may vary over time. Unlike fixed covariates, which remain constant over the study period, time-dependent covariates are assessed throughout a patient’s follow-up period. The assessment times are not necessarily at regular intervals and often differ between patients. There are often problems in collecting complete data on time-dependent covariates and, in some circumstances, the data may be censored. Time-dependent covariates could be continuous, ordinal or binary. In particular, a time-dependent binary covariate may represent the occurrence of an event, that is, an indicator variable whose value remains at zero until the event occurs, at which time it becomes 1.
The Cox proportional hazards model has been discussed above in relation to fixed covariates but the model can be extended to incorporate time-dependent covariates. The Cox model for the hazard function can be written as

\[ h(t) = h_0(t) \exp(\beta^T x + \delta^T z(t)) \]  \hspace{1cm} (15)

where \( h_0(t) \) is the underlying baseline hazard, \( x \) is a vector of fixed covariates with associated regression coefficients vector \( \beta \) and \( z(t) \) is a vector of time-dependent covariates with associated regression coefficients vector \( \delta \). The variables in \( z(t) \) represent successive measures of a binary, ordinal or continuous covariate.

When time-dependent variables are included in the model the relative hazard, \( h(t) / h_0(t) \) becomes time-dependent and so the model ceases to be a proportional hazards model.

**Estimating model parameters**

Suppose the death times of the \( n \) individuals in a study are \( t_1 < t_2 < \ldots < t_i < \ldots < t_n \), where some of these death times may be censored, and suppose \( R(t_i) \) is the set of individuals at time \( t_i \) who are at risk of death. To fit the Cox model with a time-dependent covariate \( z(t) \), the value of the covariate \( z(t_i) \) at each uncensored death time \( t_i \) is needed for all individuals in the risk set \( R(t_i) \). The value of the covariate at any time \( t \) is usually taken to be the last recorded value prior to \( t \). In situations where there are recorded values either side of the time of interest, then it may be preferable to use either the value at the closest time or, for continuous variables, a linearly interpolated value.

Nested models can be compared by checking the change in value of \(-2\log(\text{likelihood})\) against a chi-square distribution with degrees of freedom equal to the difference in the number of parameters being estimated. In SAS/STAT, the PHREG procedure can be used to fit survival models with time-dependent covariates, although the agreg function in S-PLUS may be more straightforward to use.

**Application to quality-of-life data**

During a study, a patient may experience changes in quality of life as time passes. This could be described either by a changing quality-of-life score or by the movement in and out of various quality-of-life health states. The change in score or pattern of movement between states may help to explain survival differences and should be considered for inclusion in any survival model as a covariate. The change in quality of life over time is clearly a time-dependent covariate.

Consideration of models that include a time-dependent covariate representing quality of life and a treatment term may provide useful information on the effectiveness of the treatment. If the model includes a quality-of-life term but the estimate of treatment effect is not clearly different from zero, this could indicate that the treatment affects quality of life, which in turn affects survival, but the treatment has no additional effect.

**Cox regression analysis of MIC data with quality of life as time-dependent covariate**

In a survival analysis of the MIC study, quality of life was included as a time-dependent covariate in two different ways. Patients were successively assessed for quality of life over time for 18 weeks from entry to study and the values of both MQS and MAL were included as time-dependent covariates in a Cox model of survival within 18 weeks. MQS was included in terms of the changing value over time while malaise was included in terms of the movement between two different health states: no malaise (MAL = 0) and malaise (MAL = 1, 2 or 3). Time spent with no malaise was allocated the value 0 and with malaise was allocated the value 1. Treatment was included in the model as a binary variable, with 1 representing chemotherapy and 0 representing standard palliative care.

The analysis required, for the whole survival time of a patient, a value of quality of life to be available continuously over the 18-week analysis period. To enable this, two assumptions regarding the quality-of-life data were made (see page 58 for further details of these assumptions).

- Changes in quality of life were assumed to occur midway between assessments and patients were assumed to remain in steady state between changes.
- Patients were assumed to remain in steady state from time of last assessment to either death or 18 weeks (whichever came first).

The S-PLUS program was used to perform the analysis. The data were transformed so that it consisted of multiple lines per patient, with each line representing the time a patient spent either in different states of malaise or with different values of MQS. Each line of data consisted of the times of entry to and exit from the state relative to date of entry to study, an indicator variable representing whether the patient had died or not at the end of the interval, and a variable representing treatment.

The regression coefficients and hazard ratios for the quality-of-life covariates suggest that there is
no relationship between the changing values of malaise and survival within 18 weeks (see Table 19) but there is evidence of a relationship between the changing values of MQS and survival within 18 weeks (see Table 20). Hazard rates suggest that increasing values of MQS (i.e. worsening quality of life) are associated with an increased hazard of death. Regression coefficients and hazard ratios (see Tables 19 and 20), together with likelihood ratio tests (see Table 21), show that after adjusting

for the effect of changing values of both malaise and MQS over time, survival within 18 weeks is still significantly different between treatments.

References


### Table 19

<table>
<thead>
<tr>
<th>Survival within 18 weeks</th>
<th>Malaise β (95% CI)</th>
<th>Hazard ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malaise</td>
<td>0.72 (–0.01, 1.45)</td>
<td>2.05 (0.99, 4.25)</td>
</tr>
<tr>
<td>Treatment (adjusted for malaise)</td>
<td>–0.94 (–1.59, –0.29)</td>
<td>0.39 (0.20, 0.75)</td>
</tr>
</tbody>
</table>

### Table 20

<table>
<thead>
<tr>
<th>Survival within 18 weeks</th>
<th>MQS β (95% CI)</th>
<th>Hazard ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MQS</td>
<td>0.99 (0.44, 1.54)</td>
<td>2.68 (1.55, 4.66)</td>
</tr>
<tr>
<td>Treatment (adjusted for MQS)</td>
<td>–0.70 (–1.37, –0.03)</td>
<td>0.50 (0.25, 0.97)</td>
</tr>
</tbody>
</table>

### Table 21

<table>
<thead>
<tr>
<th>Survival within 18 weeks</th>
<th>Unstratified</th>
<th>Stratified by baseline malaise</th>
<th>Stratified by baseline MQS</th>
<th>Adjusted for baseline malaise</th>
<th>Adjusted for baseline MQS</th>
<th>Adjusted for changing malaise</th>
<th>Adjusted for changing MQS</th>
</tr>
</thead>
<tbody>
<tr>
<td>χ²</td>
<td>8.07</td>
<td>8.09</td>
<td>8.25</td>
<td>7.53</td>
<td>6.41</td>
<td>8.02</td>
<td>4.30</td>
</tr>
<tr>
<td>p-value</td>
<td>0.0045</td>
<td>0.0044</td>
<td>0.0041</td>
<td>0.0061</td>
<td>0.0113</td>
<td>0.0046</td>
<td>0.0381</td>
</tr>
</tbody>
</table>
Chapter 13

Analysis of survival data adjusting for quality of life: summary and discussion

Treatment comparison in terms of survival may, in some circumstances, need to adjust for the effects of quality of life. Standard survival analysis techniques can be used for this purpose. Quality of life may be included as a covariate, either in terms of baseline values or in terms of changing values over time. The quality-of-life measure may be binary, ordinal or continuous, and survival times may have a specified or unspecified underlying distribution.

If adjustment for baseline quality-of-life measures is necessary, then a stratified survival analysis can be used. Although the method is simple, it has limited use; the stratification variable must be categorical and have a small number of levels, and the method is unable to cope with more than a few covariates. Categorisation of a continuous quality-of-life measure for use in a stratified survival analysis may result in a loss of information; a modelling approach, which can incorporate any number and type of covariates, may be more appropriate.

Cox proportional hazards regression models can be used to adjust for baseline measures of quality of life in a survival analysis, by incorporating the values as a fixed covariate. Cox models can also be used to adjust for quality of life in terms of change over time by including the changing values as a time-dependent covariate. Modelling in this way also enables investigation of possible interactions between treatment and quality of life, which may be particularly important if a treatment, for example, causes early toxicity.

If assessments of quality of life are infrequent or data are missing for reasons other than death, then it may be difficult to adjust for changing quality of life with any degree of accuracy. The analysis may be improved by modelling quality of life and survival as two simultaneous processes, as discussed later in chapter 19. In this approach, values of quality of life incorporated into a survival analysis as a covariate are estimated from the model for quality-of-life data over time fitted to all patients.
Chapter 14

Simultaneous analysis of quality-of-life and survival data: introduction

Health technology assessment often requires evaluation in terms of a trade-off between quality and length of life. In comparing the effectiveness of a new treatment against a standard treatment in a randomised clinical trial, benefits in terms of survival often have to be weighed up against the extent of undesirable side-effects. Alternatively, a new treatment may have no extra benefit in terms of survival but it may result in improved quality of life through improved relief of symptoms or reduced toxicity.

When quality of life and survival are analysed as separate endpoints, it is often difficult to assess the balance between the two in selecting the optimal therapy for patients, especially when the two endpoints indicate conflicting treatment preferences. It may therefore be preferable to assess the effect of a treatment on quality of life and survival simultaneously, and the statistical methods appropriate for that purpose are discussed in chapters 15–20.

The concept of QALYs is introduced in chapter 15. This is a composite outcome measure in which the survival time of a patient is scaled down according to the quality of life they experience. The quality-adjusted survival time is then used to compare treatments in a quality-adjusted survival analysis, discussed in chapter 16, which includes the TWiST (time spent without symptoms of disease and toxicity of treatment) concept and Q-TWiST (quality-adjusted TWiST) methodology.

An alternative approach to the simultaneous analysis of quality-of-life and survival data, discussed in chapter 17, is multistate survival analysis. Rather than analysing quality-adjusted survival times, this method models the movement of patients between a finite number of health states defined in terms of quality of life. Some relatively new methodology is outlined in chapter 18, which combines this approach with the previous quality-adjusted survival analysis approach by incorporating the results from a multistate model into a Q-TWiST analysis.

The direct modelling of quality of life and survival is considered as two simultaneous processes in chapter 19, in particular, how the two interrelate. Such modelling allows for the comparison of quality of life conditional upon survival. A more general approach, which considers modelling the drop-out process rather than survival, is also discussed.

The findings of this part of the study are summarised and discussed in chapter 20.
Chapter 15
Quality-adjusted life years

Introduction

QALYs are a measure of health status and can be used as an alternative to survival time in some areas of health technology assessment. They account for both quality and duration of survival in a single outcome measure and express health status in terms of well years of life.

The QALY methodology consists of down-weighting periods of survival time for which patients experience sub-optimal levels of quality of life. A special type of weight, often used in QALY calculations, is the utility, which was mentioned briefly in chapter 3 as a means of measuring quality of life. An introduction to utilities follows below. The QALY methodology is illustrated with data from the MIC study.

QALYs were first proposed in the USA by Fanshel and Bush and have been developed in the UK by Alan Williams and colleagues at York University. They are extensively used in health economics, forming the basis for cost–utility analysis, and Torrance was at the forefront of the methodology in this field. They are also widely used in the field of decision analysis. These QALY applications are discussed below. In this report, the main application of the QALY concept is in chapter 16, where the assessment of health technologies in terms of quality-adjusted survival analysis is discussed. The advantage of QALYs lies in their simplicity but they have been the subject of much criticism (see page 61) and a number of alternatives have been proposed.

Utilities

Each health state experienced by an individual during the course of their disease and treatment is associated with a quality of life. Each can be assigned a value that measures the preference of the individual for a health state relative to other states. The value reflects the quality of life in that health state. The value generally lies between 0, representing death, and 1, representing perfect health, although it can be negative, representing health states judged to be worse than death. Technically, these values are utilities when they are measured under conditions of uncertainty but they may approximate to utilities when measured under other conditions.

Once the health states experienced by patients in a study have been defined, then the utilities for these states need to be determined. Methods of measuring health state utilities have been discussed and Torrance, in particular, provides a comprehensive review. There are three broad methods available to researchers for determining utilities for a study; using subjective judgement, using values in the literature or measuring the values. The values can be measured by using a holistic approach in which the health state being valued is considered as a complete entity or by using a decomposed approach which values a health state indirectly by considering particular aspects of a health state.

Determining utilities using subjective judgement

The researchers themselves can make subjective assessments as to the utilities of various health states in their study or they can call on clinicians or other experts for their opinion on the values. This is a quick and inexpensive way to determine utilities but will not necessarily provide realistic measures. Sensitivity analysis should be carried out to determine the robustness of the conclusions to the choice of utility values.

Taking utilities from the literature

The literature can be searched for utility values from previously conducted studies that could be used in the present study. The subjects and health states for which utilities were calculated in a published study should be appropriate for the current need. One very commonly used study is that by Rosser and Kind, whose valuations are at the forefront of the work on QALYs undertaken in the UK.

Rosser and Kind valuation matrix

The matrix of utilities derived by Rosser and Kind is based on a classification system derived by Rosser and Watts and values paired combinations of eight categories of disability (no disability to unconscious) and four categories of distress (none, mild, moderate and severe). The matrix represents the average valuation by 70 people (20 patients,
20 nurses, 20 healthy volunteers and ten doctors), given on a scale of 0, representing death, to 1, representing perfect health, with two negative values representing states worse than death.

If patients in a study are categorised into the same health states as the Rosser and Kind valuation matrix, then the utilities from the matrix can be used in QALY calculations. One approach is to use clinical judgement to place patients into the categories, while another approach is to use the Health Measurement Questionnaire, as developed at the University of York. The questionnaire has been specially devised so that responses given by individuals can be used to place them into one of the Rosser and Kind categories. A further approach may be to reprocess quality-of-life data collected by other means to categorise patients.

There has been some discussion regarding the appropriateness of using the Rosser and Kind valuation matrix. Criticisms relate to the study being out-of-date, the sample on which valuations were measured being small and unrepresentative of the general population, and the scale being insensitive. There has also been criticism regarding the use of the Health Measurement Questionnaire before proper testing and validation of the instrument has been carried out. The use of the Rosser and Kind valuation matrix has been largely superseded by others such as the EuroQol.

**Measuring utilities: the holistic approach**

The most accurate way to measure utility values is by questioning a sample of individuals and eliciting their relative preferences for health states. The holistic approach considers a health state as a complete entity. The subjects could be the patients themselves, health professionals or the general public. The choice of subjects may not be important since large differences in utilities between different groups of people are unusual, although some differences have been observed.

In a clinical trial scenario, patients are the most appropriate subjects when measuring the utility of their condition. If, for example, health professionals or the general public are asked to measure the utility of a health state then a clear description needs to be given, since, unlike a patient, they are not actually in that health state and have not experienced it. The description should be abbreviated in terms of the physical, emotional and social functioning of the condition and should include age at onset, duration of the state, exact prognosis for what follows the state and whether it applies to them or someone else. The general public may need a different level of detail to that given to health professionals.

Having selected the individuals on whom utilities will be measured, an appropriate approach to measuring utilities needs to be chosen. There are three main methods that will be discussed here, although other methods such as ratio scaling may also be used. The rating scale is a direct method for eliciting subject preferences while the standard gamble and time trade-off (TTO) are indirect methods.

**Rating scale**

The rating scale method involves presenting subjects with a line on a page, with one end of the line clearly defined as the most preferred health state and the other end as the least preferred. The subjects are asked to place health states on the line, such that the order and spacing represent preferences and differences in preferences as perceived by the subject.

**Standard gamble**

This is the classical method of measuring cardinal preferences and is the technique used in economic theory to measure utilities. It is based directly on the utility theory presented by von Neumann and Morgenstern.

The subject is presented with two scenarios that they must choose between:

(i) treatment resulting in either normal healthy life for $t$ years with probability $p$ or immediate death with a probability of $1-p$

(ii) a certain health state for life ($t$ years).

The probability $p$ is varied until the respondent is indifferent to the two choices, thus giving the utility value, $p$, for that state. There is evidence to suggest that the standard gamble method overestimates utilities for health states and an improved version has been proposed.

**TTO**

This method was developed specially for use in health care by Torrance and colleagues. The subject is asked how much time, $x$, in a state of perfect health he/she considers equivalent to a period, $t$, in his/her current health state (worse than perfect health). The choice of the length of period $t$ is problematic. In practice, the subject chooses between:

(i) a certain health state for time $t$ followed by death

(ii) healthy for time $x < t$ followed by death.
Time \( x \) is varied until the respondent is indifferent to the choices and \( x/t \), the TTO-score, ranging from 0 to 1, provides an estimate of the utility for that state.

When the TTO-score assessed for time \( t \) is applied to other periods then constant proportional trade-off is assumed. For example, if an individual considers 16 years in perfect health equivalent to 20 years in a reduced health state, then it is assumed they would also consider 4 years in perfect health to be equivalent to 5 years in the reduced health state. Studies have shown that this assumption may not always be valid.\(^{27,35} \) A violation of the proportionality assumption invalidates the calculation of a TTO-score and subsequent application of this score to periods of length other than \( t \) in QALY calculations (see page 62). For further discussion regarding the interpretation of the TTO score as a measure of utility, see the paper by Stalmeier and colleagues.\(^{35} \)

**Comparison of methods**

The standard gamble method is the gold standard for measuring utilities but is impractical, especially when a large number of health states need to be measured, and can be expensive because of the need for interviewers. The TTO is similar to the standard gamble but is easier to use, and the rating scale method is the simplest of all. The TTO and rating scale methods, however, do not in general measure utility directly since they do not involve probabilities, and utilities are preferences measured under conditions of uncertainty. Preference values, \( v \), from methods such as TTO and rating scale, can be converted to utilities, \( u \), using a power function: \(^{36,37} \)

\[
(1 - u) = (1 - v)^\alpha
\]

(16)

where \( \alpha \) reflects the risk attitude of the individuals \((0 < \alpha < 1 \) reflects risk-averse, \( \alpha > 1 \) reflects risk-seeking, and \( \alpha = 1 \) reflects risk-neutral) and may be estimated by measuring values and utilities on the same health states. The TTO method only measures the true utility when people are risk-neutral \((\alpha = 1)\) and the utility function is linear in time. Studies have shown that the TTO method tends to underestimate utilities since individuals are generally risk-averse rather than risk-neutral.\(^{38} \)

The rating scale, standard gamble and TTO methods have been compared with each other and with other methods.\(^{29,39-43} \) The different methods can give very different results and this can then affect any analysis that incorporates them.\(^{29,40} \) Torrance\(^{41} \) recommended researchers to use the TTO technique if they could afford it and, otherwise, to use the rating scale with power curve correction (equation 16).

**Measuring utilities: the decomposed approach**

The decomposed approach to measuring utilities is based on multi-attribute utility theory,\(^{37} \) in which health status is described in terms of a set of core attributes, each broken down into a series of levels describing the range of functioning (as shown by Barr and colleagues,\(^{44} \) for example). Each different combination of levels, one from each attribute, represents a unique health state.

Utility values for each subject in the study are estimated by measuring the health status of a patient using a questionnaire, and converting their responses into a utility by combining them in a predetermined formulation. A variety of questionnaires have been devised to measure preferences using a decomposed approach, including the Health Utility Index, the Quality of Well-being Scale, the Rosser Index and the EuroQol\(^{26} \) (now called the EQ-5D). A comparison of these instruments is reported in another issue of *Health Technology Assessment*.\(^{45} \)

**QALY methodology**

**The QALY model**

To calculate QALYs, years of life are multiplied by a fraction, the quality-adjustment fraction, which expresses the impairment in quality of life experienced during this time. The quality-adjustment fraction, ranging from 0, representing quality of life equivalent to death, to 1, representing perfect health, may represent the utility and can be derived in several ways (see above).

If the patient experiences or is expected to experience a series of health states \( s_i \) \((i = 1 \text{ to } n)\), with different levels of quality of life measured by utilities \( u_i \) \((i = 1 \text{ to } n)\), and the time spent in each state \( s_i \) is given by \( t_i \) \((i = 1 \text{ to } n)\), then the conventional approach to calculating QALYs is to sum the weighted times spent in the different states. This gives the following standard form for the QALY model:

\[
\text{QALY} = \sum_{i=1}^{n} u_i t_i
\]

(17)

In graphical terms, if the quality of life of an individual, represented by utility values, is plotted over time, then the QALYs for that individual are
calculated as the area under this curve (see below and also Figures 16 and 17 later for examples from the MIC data).

**Discounting**

QALYs are sometimes discounted to adjust for the fact that immediate benefits are generally valued more highly than later ones. The benefits of future life years are converted into present values by discounting. A discount rate of 5% per annum is used in most studies. If discounting is at a rate of \(r\)% per year, then the discount factor for year \(n\) is given by:

\[
\text{Discount factor}_n = \left(1 + \frac{r}{100}\right)^{-n} \quad (18)
\]

For example, suppose an individual lives for 5 years with a quality-adjustment fraction of 0.8, then the total number of non-discounted QALYs will be 4 (i.e. \(5 \times 0.8\)). If discounting at 5% per year then the total number of discounted QALYs will be 3.46 (i.e. \((1.05^{-1} \times 0.8) + (1.05^{-2} \times 0.8) + \ldots + (1.05^{-5} \times 0.8)\)).

Gudex and Kind\(^{14}\) expand further on discounting in their explanation of QALY methodology using Rosser and Kind valuations.

**Model assumptions**

Glasziou and colleagues\(^{46}\) discuss the assumptions on which the QALY model is based, which are as follows:

- **utility independence**: the utility value for a health state does not depend on the time spent in that state
- **context independence**: the utility value assigned to a health state is independent of previous or future quality of life or the amount of remaining life
- **risk neutrality**: all life years are valued equivalently.

The risk neutrality assumption means that time is included in the model as a linear term, \(t\), rather than as a non-linear function of time, that is, \(f(t)\). More general models, that include some sort of discounting or risk-adjustment, have also been suggested.\(^{46-48}\)

**Practical issues**

Given longitudinal quality-of-life data and survival times for a group of subjects, there are two approaches to calculating QALYs. The first is to use the serial measurements as utility values over time and thus calculate QALYs for each individual. This is straightforward if the instrument for measuring quality of life yields a utility-type value, otherwise the quality-of-life scores need to be transformed so that they are on a 0–1 scale, with 0 representing death and 1 representing perfect health. The second approach is to use the quality-of-life data to categorise a patient at any time into one of a number of health states. The utility values for these health states are determined separately and can be combined with the time spent in each health state to form QALYs for each patient.

If the quality of life for an individual changes between two consecutive assessments then, for both approaches to QALY calculation, a technique for determining the exact time of the change is required. Three possible techniques are as follows (see Figure 15):

(a) assume that quality of life changes linearly between assessments
(b) assume that quality of life is maintained from one assessment to the next
(c) assume that quality of life changes at the midpoint between assessments.

The appropriate choice may depend on the type of quality-of-life measurement, the timing of the assessments and the period to which the questions relate. Ganiats and colleagues\(^{49}\) compared the first two techniques and concluded that either the appropriate choice for a study should be decided at the protocol stage or a sensitivity analysis should be carried out to establish the extent to which the conclusions are affected by the technique chosen.

**Example: calculating QALYs in the MIC study**

In the MIC study, quality-of-life data is available only for the first 18 weeks from study entry (see chapter 4) and so only the survival time for this limited period can be adjusted for quality of life. Assessments were made at several distinct time points during this period and various assumptions are needed in order to translate the quality-of-life information so that it is available over continuous time (see below). Sensitivity analysis could be used to assess the impact of the assumptions on the conclusions of any analysis. The two approaches to calculating QALYs, based on these assumptions, are described below.

**Assumptions regarding quality of life over time**

**Time between study entry and first assessment**

Quality of life during the time between study entry and first assessment was assumed to be that measured at the first assessment, that is, the first
value was carried backwards to date of study entry. This is not an unreasonable assumption, firstly because the questionnaire related to the previous 3-week period and, secondly, because the time between study entry and first assessment was reasonably short (median, 1 week; range, 0–6.7 weeks).

**Time between last assessment and either death or 18 weeks**

Quality of life during the time between last assessment and either death or 18 weeks (whichever came first) was assumed to be that measured at the last assessment, that is, the last value was carried forward. This is not an unreasonable assumption if the period is short but, for those who drop out of the study for reasons other than death, this assumption may be problematic (see Table 8). Other approaches, such as worst value carried forward and linear decrease over time were considered. Each has its own problems, especially when dealing with drop-outs (see Table 8).

**Times of changes in quality of life**

If quality of life changes between two consecutive assessments then the change is assumed to occur at the midpoint between the two assessment times. This is assumption (c) discussed above under ‘Practical issues’. Quality of life is assumed to remain constant between changes. Intermittent missing quality-of-life values are handled by considering consecutive **non-missing** assessments when determining changes in quality of life. Thus, if there was a change in quality of life between assessment \( m \) taken at \( t_m \) and assessment \( n \) taken at \( t_n \) where \( m < n \), then the time of change was taken as:

\[
\text{Time of change} = t_m + 0.5 (t_n - t_m) \tag{19}
\]

Alternatives for estimating the exact times of change were considered. The actual assessment date when the change in quality of life was measured could have been used, that is, assumption (b) discussed above under ‘Practical issues’. However, this option takes no account of the fact that the questionnaire was retrospective. To accommodate this, an alternative was considered in which time of change was chosen as the date 3 weeks before the date when the change in quality of life was measured. For various reasons, this was problematic and it was considered that, in reality, the patient would probably not recall their quality of life that far back. The midpoint option therefore seemed the most realistic and the effect of this choice on the analysis could be investigated in a sensitivity analysis.
Calculating QALYs

Two different approaches, as discussed earlier, were considered for calculating QALYs in the MIC study. One approach was to use MQSs to estimate utility over time and calculate QALYs. The other approach was to use MAL to divide survival time into periods of differing quality of life and use selected utility values to calculate QALYs.

In order to use the MQS as a utility, transformation of the score was required. The maximum possible score of 3, occurring when all symptoms and side-effects were severe, was not deemed to be equivalent to death and so a score of 4 was allocated to death. The minimum possible score was 0, which occurred when there were no symptoms or side-effects, reflecting perfect health. The following transformation reverses the score and places it on a 0 to 1 scale:

\[
MQS' = 1 - \frac{MQS}{4}
\]  

with MQS' = 0 assumed to be the quality-of-life score for death. The problem with this approach in the MIC study is that quality of life was not measured as a utility. In particular, the responses were categorical and quality of life was not assessed relative to death. Thus the use of the transformed score as a utility is subjective and controversial, and should be viewed only as an illustrative example.

As an alternative, MAL was used to divide each patient’s survival time within 18 weeks into periods of ‘good’ (MAL = 0) and ‘poor’ (MAL = 1, 2 or 3) quality of life. To illustrate this approach, utility values for these two health states were chosen arbitrarily as 1 and 0.8, respectively. The choice of utility values is subjective but alternative values can be considered as part of a sensitivity analysis. This approach to QALY calculation was chosen as the preferred option for the MIC study, since it was more flexible with regard to utilities.

As an illustration, the QALYs calculated from each approach were compared for two individuals in the MIC study, one who died within the 18-week period (see Figure 16) and one who died after 18 weeks (see Figure 17). The values of MQS’ were plotted over time, together with the division of survival time based on MAL. The area under each of these curves gives the QALYs within 18 weeks for those individuals. The plots in Figure 16 gave QALYs of 13.58 for MQS’ and 16.21 for MAL, while those in Figure 17 gave QALYs of 12.21 for MQS’ and 15.27 for MAL. The QALYs based on MAL are used in chapter 16 to illustrate a quality-adjusted survival analysis.

Applications of QALYs

QALYs are widely used in the fields of health economics and decision analysis and a brief outline of such applications follows. Examples are given and further examples can be found in the bibliography (see page 109). QALYs can also be used as the outcome measure in an assessment of treatments in a clinical trial. This forms the basis
for quality-adjusted survival analysis, discussed in chapter 16.

**The use of QALYs in health economics**

Cost–utility analysis assesses health technologies in terms of cost per QALY. For example, a hip replacement costs £750 per QALY gained and a heart transplant costs £5000 per QALY gained.2 Different strategies for treating a group of patients can be compared in terms of this outcome measure, to aid determination of the optimal treatment.50–56 The use of QALYs in this context has been discussed.57–59 Cost–utility analysis can also be used to assess the benefits of health interventions such as disease prevention programmes,60 screening programmes61,62 and healthcare guidelines.63 Further examples can be found in the decision-analysis field (see below).

A major use of cost–utility analysis is in the prioritisation of strategies for treating different groups of patients, aiding decisions regarding resource allocation. Gudex21 provides a detailed example of the use of QALYs for this purpose. The use of QALYs in resource allocation has been discussed,2,64–68 although there is much controversy regarding such applications (see below).

**The use of QALYs in decision analysis**

Decision analysis is a technique used to aid clinical decision making and is sometimes seen as an alternative to other approaches, such as clinical trials and meta-analysis, for health technology assessment. It provides a means of synthesising all existing knowledge on an intervention and quantifying the net benefits. Decision analysis is not only useful for evaluating optimal treatment for patients, but also for identifying the most important variables in decision-making processes.69

Life expectancy has been the most popular outcome measurement for decision analyses but QALYs are also now being widely used as an alternative.70 Using QALYs as an outcome measure requires estimates of life expectancy, together with appropriate quality-of-life adjustments. Cost–utility analysis is a specialised form of decision analysis using cost per QALY as an outcome measure. There are many examples in the literature in which QALYs have been used in a decision analysis to evaluate various health technologies.71–113

Weinstein and Feinberg69 provide a definitive guide to the methodology of clinical decision analysis, while others discuss the use of Markov models in this context.114,115 The basic model for a decision analysis, which defines the clinical situation for which a decision is needed, is usually represented by a decision tree. For each treatment alternative, the sequence of possible events are represented by a series of decision nodes (events over which clinician has some control) and chance nodes (events over which there is no control). Transition probabilities are allocated to each chance event, using existing knowledge from previous studies, routinely collected data or subjective expert opinion. In the same way, utility values need to be assigned to each health state to reflect the quality of life in that state.

If the outcome measures for the decision analysis are QALYs, then estimates of life expectancy are combined with appropriate utility values. Life expectancy can be estimated using methods such as declining exponential approximation to life expectancy (or DEALE) and Gompertz models,116,117 or methods based on Markov models, such as Markov cohort simulation and Monte Carlo simulation.114,115 Quality-adjusted life expectancy can be calculated for each treatment option, with the treatment with the maximum value assumed to be optimal.

Sensitivity analyses are usually carried out to explore how possible ranges of values in terms of transition probabilities, utilities and costs may affect the conclusions. Considering the effects of changing more than one variable at a time, in a two- or three-way sensitivity analysis, for example,89 may be more realistic but can become complex. Another complex approach, probabilistic sensitivity analysis, uses probability density functions for each variable to account for uncertainty.89 Receiver–operator-characteristic curves can also be used in a sensitivity analysis to identify the most important variables for a decision problem.112

**Critical appraisal of QALYs**

There has been much discussion about QALYs relating to both the theoretical and practical difficulties associated with their calculation, and controversy remains regarding their usage. A selection of references discussing QALYs are given here and further references can be found in the bibliography (page 109).

**Theoretical and practical difficulties**

A major difficulty with QALYs is the derivation of appropriate quality-adjustment fractions. QALYs need to be theoretically and methodologically correct but they also need to be based on ‘good’
quality (i.e. accurate) data. The basic concept that quality of life can be measured in cardinal numbers on a ratio scale could itself be questioned but, assuming that quality of life is measurable, there is still the difficulty that quality-adjustment fractions need to account for the many different ways in which quality of life can be impaired.

Quality-adjustment fractions need to be derived from an appropriate and reliable source. The extensive use of the Rosser valuation matrix as quality-adjustment fractions in QALYs is questioned, but on judgements from a small sample of arbitrarily chosen respondents and were never intended to be used for calculating QALYs. A further difficulty in the derivation of quality-adjustment fractions is the requirement for them to satisfy 'reciprocal commensurability' between duration and quality of survival (called constant proportional trade-off on page 57), that is, $x$ years of life at $x^{-1}$ quality should be equivalent to 1 year of life at unimpaired quality. There is some empirical evidence that validates this concept.

There has been much discussion regarding the validity of the assumptions underlying the QALY model (see page 58), especially by authors advocating an alternative measure (see below). Carr-Hill examines in detail the theoretical assumptions underlying the QALY procedure. There is some empirical evidence to show that none of the assumptions hold. In particular, empirical evidence from Richardson and colleagues casts doubt on the validity of the additive assumption in the usual QALY model. The advantage of using HYEs is that they do not rely on the restrictive assumptions of the QALY model (see page 58) and they allow attitudes towards risk to be incorporated. There has been much debate regarding these claims, together with comparisons between HYEs and QALYs (see Bibliography for further references).

HYEs are calculated using a two-stage scheme involving two standard gamble questions (see page 56 for information on the standard gamble procedure). It has been argued that this two-stage procedure is unnecessary since it is equivalent to a straightforward TTO (see page 56 for information on the TTO technique). One of the major problems with HYEs, admitted by Mehrez and Gafni in their original paper, is their feasibility. Individuals have to participate in much longer and more complex interviews, and the costs associated with this, together with the willingness of participants, may be prohibitive. The completeness, validity and reliability of the answers given in such a situation are questionable.

**Alternative and related methods to QALYs**

**Healthy-years equivalents**

HYEs were originally proposed by Mehrez and Gafni as an alternative measure to QALYs. Like QALYs they combine two outcomes of interest: quality and quantity of life. QALYs were criticised for only partially incorporating patient preferences, with the utility approach used to obtain quality-adjustment fractions for each state separately in a patient’s health profile. HYEs were proposed as superior since they obtain the utility for the whole health profile and, therefore, fully represent the individual’s preferences.

The advantages of using HYEs are that they do not rely on the restrictive assumptions of the QALY model (see page 58) and they allow attitudes towards risk to be incorporated. There has been much debate regarding these claims, together with comparisons between HYEs and QALYs (see Bibliography for further references).
Saved young life equivalent
Nord\(^{158}\) proposed an alternative to QALYs for comparing health technologies that would overcome some of the problems associated with them. He suggested comparisons in terms of units of saving a young life and restoring the young life to full health, referred to as the saved young life equivalent (SAVE). It was chosen in the belief that most people would regard it as the maximum benefit that a single individual could obtain. Outcomes from interventions, described as fully as desired, could then be valued in terms of numbers of SAVES. Nord proposed that SAVES, which measure social value, should be used in addition to QALYs rather than to replace them.

This proposed measure was criticised for not being adequately defined, particularly in terms of what was meant by ‘young’ and also for disadvantaging older or disabled patients.\(^{159-161}\) Nord defended his measure, saying that SAVES allow society to decide on the importance of age in valuing interventions.\(^{162}\) The SAVE procedure is a specific form of a more general approach called the ‘person-trade-off approach’, proposed by Nord as an alternative to the QALY.\(^{163}\)

Quality-adjusted lives
Another alternative to the QALY is the quality-adjusted life (QAL), where treatments are assessed in terms of numbers of lives saved rather than length of life. Stevenson and colleagues\(^{164}\) assessed neonatal intensive care for low birthweight babies using QALs as well as QALYs but only presented results based on QALs, claiming that the two measures gave similar results in this context. The method assessed neonatal intensive care purely in terms of numbers of surviving babies but each life was weighted according to the level of disablement of the child.

Healthy life expectancy
Healthy life expectancy (or health expectancy) is an indicator of the health status of a population, combining mortality and morbidity into a single index, and is used in an epidemiological context. It has been defined as the number of further years of life in good health that someone of a specified age can, on average, expect to enjoy, given the age-specific rates of mortality and morbidity prevailing in the population.\(^{165}\)

Healthy life expectancy is known under various different names depending on the definition of ‘healthy’. Examples include disease-free life expectancy, disability-free life expectancy and quality-adjusted life expectancy. There has been some discussion regarding terminology used in the field of healthy life expectancy.\(^{166,167}\)

There are three different methods for calculating healthy life expectancy: the Sullivan method, the double decrement method and the multistate method. These are described and compared by Barendregt and colleagues.\(^{168}\) All three methods are variations on the standard procedure for calculating life expectancy using life tables and each has different data requirements. The Sullivan method is the most widely used because it is the simplest and is the least demanding in terms of data requirements but it can be unreliable.\(^{168}\)

Healthy life expectancy can be used to assess the impact of chronic diseases on the health of a population,\(^{169,170}\) to assess the impact of an intervention on the health of a population,\(^{171}\) or to examine the underlying trend in the health of a population.\(^{172}\) If healthy life expectancy is to be used as a measure by which to compare populations then standardisation, in terms of definitions and instruments measuring morbidity, is required.\(^{165,173}\) Conceptual and ethical issues regarding the use of healthy life expectancy have also been raised.\(^{174,175}\)

References
Quality-adjusted life years


Chapter 16

Quality-adjusted survival analysis

Introduction

Quality-adjusted survival analysis can be used as an alternative to survival analysis in assessing health technologies in studies in which quality of life and survival are both important endpoints. Adjustments are made to account for the quality of life experienced during the survival time using approaches based on the QALY model (see chapter 15).

There are two main approaches to quality-adjusted survival analysis depending on the level of aggregation of the data. The subject-based approach combines quality-of-life and survival data at the patient level, thus creating a single endpoint for each subject on which to compare health technologies, while the population-based approach aggregates the quality-of-life and survival data at a (pre-defined) group level.

The first part of this chapter covers both subject-based and population-based approaches to quality-adjusted survival analysis using QALYs. A special form of QALY, known as TWiST is described; the analysis of this endpoint is discussed in terms of a subject-based approach. An extension of the TWiST concept, Q-TWiST, is also described; here the analysis takes the population-based approach.

Quality-adjusted survival analysis using a QALY model

Subject-based approach

In any study assessing health technologies, QALYs can be calculated for each subject by weighting periods of survival time according to the quality of life experienced during these periods (see chapter 15). The sum of these weighted survival times creates a single endpoint for each subject for which health technologies can then be compared.

If survival times are known for all the patients in a study (i.e. none are censored), then standard techniques for dealing with continuous data can be used.1 The sample mean and standard deviation of the QALYs for the patients in each treatment arm could be used to provide an estimate of the mean treatment difference in QALYs, together with an appropriate CI. The hypothesis of no difference between treatments in terms of QALYs can be tested using a t test, for example, if distributional assumptions were satisfied.

If the data for some patients are censored, then a standard survival analysis using QALYs for each individual as an endpoint rather than actual survival time may seem appropriate. However, there will be a problem of informative censoring (see page 11). Survival time for patients with poor quality of life will receive a lower weighting than that for patients with good quality of life. Patients with poor quality of life will thus accumulate QALYs at a slower rate and will, therefore, be censored earlier on the QALY timescale than those with good quality of life. This will give biased Kaplan–Meier estimates and invalidate the log-rank test, as well as other standard survival analysis techniques, as a means of comparing treatments.

In the situation of informative censoring, there are two options available for a valid analysis.

1. The censoring date for the analysis could be set as the smallest censored survival time value, thus restricting the analysis to a period during which all subjects have full follow-up and eliminating censoring. This is only feasible if the smallest censored value is quite large, otherwise a considerable number of events may be lost, thus reducing the statistical power of the analysis.

2. A population-based approach such as that based on the QALY model or the Q-TWiST model may be used. Using this approach, survival analysis is applied to unweighted survival times, for which censoring is non-informative, and the quality-of-life weightings are then applied to treatment group averages obtained from this unbiased survival analysis.

Example using the MIC data

Quality-of-life data was restricted to the 18 weeks from entry to the study, hence a quality-adjusted survival analysis was only possible over this period. An unadjusted survival analysis (see page 44) showed a statistically significant difference in survival within 18 weeks of entry to trial (log-rank test $\chi^2 = 8.07, p = 0.005$).
Quality-adjusted survival analysis using QALYs within 18 weeks of study entry were calculated for each individual (see page 58). All were followed-up for at least 18 weeks; hence, there were no censored survival times within this period. Quality-adjusted survival analysis using QALYs for each patient as an endpoint is thus straightforward since, with no censored data, informative censoring is not a problem. Restricting the analysis of the MIC data to 18 weeks from study entry, because of the availability of quality-of-life data, has essentially resulted in overcoming the informative censoring problem in a way equivalent to the first option discussed above.

With no censored data, treatment comparisons can be made using either standard techniques for continuous measures or standard survival techniques without concern about bias. Since the data distribution was not normal, standard survival analysis techniques were used to compare treatments. Kaplan–Meier survival curves for quality-adjusted survival (see Figure 18) showed chemotherapy to be superior to palliative therapy but the difference was not of statistical significance (log-rank test $\chi^2 = 1.33, p = 0.25$). It should be noted that the steps on these Kaplan–Meier curves do not represent deaths, as they would for a standard survival endpoint, but show reductions in the proportion of patients achieving increasing levels of quality-adjusted survival time.

In conclusion, although treatments differed in terms of survival within 18 weeks, they were not found to differ in terms of quality-adjusted survival. This is based on the use of a value of 0.8 to weight periods with poor quality of life. Sensitivity analysis should be used to establish the robustness of conclusions to the choice of weighting factor used to calculate the QALY endpoint for each patient.

**Population-based approach**

The subject-based approach to quality-adjusted survival analysis using QALYs can be problematic and lead to biased results in situations with censored survival times. It may therefore be preferable to use a population-based approach. This method combines the quality-of-life and survival data at the population or group level rather than at the subject level.

Essentially, a quality-adjusted survival curve for a population is formed by plotting, against time $t$, the product of the mean quality of life of patients living at time $t$ and the probability of surviving to time $t$. The area under this quality-adjusted survival curve gives the mean quality-adjusted survival for the population. Beacon$^2$ refers to this method as the integrated quality–survival product and illustrates the method using quality-of-life data.

In situations in which quality of life is measured at $k$ discrete time points, $t_1, t_2, ... t_k$, the following function is an estimator of the expected quality-adjusted survival time (QAS).$^5$

\[
QAS = \sum_{i=1}^{k} \left( \frac{(Q_i + Q_{i+1})}{2} \right) \left( \frac{(S_i + S_{i+1})}{2} \right) (t_{i+1} - t_i) \quad (21)
\]

where $Q_i (i = 1–k)$ is an estimate of the mean quality of life at time $t_i$ and $S_i (i = 1–k)$ is an estimate of the probability of survival to time $t_i$. In this model, the quality-of-life and survival probabilities in the interval between two time points are assumed to be the average of the values at the time points at each end of the interval.

Hwang and colleagues$^3$ suggest that survival probabilities could be estimated using standard methods such as the life table method, Kaplan–Meier estimates, or by using parametric models; they also suggest estimating the mean quality of life at the given times using kernel smoothing methods. Calculation of the standard error of the mean quality-adjusted survival time is mathematically complex and here they suggest using bootstrap methods$^4,5$ (see page 77).

Ganiats and colleagues$^6$ identified and compared three methods for estimating the mean quality of life and survival probabilities for input into...
equation 21. The simplest method is to use the mean of the quality-of-life data at time $t_i$ and the proportion of subjects surviving this time as the crude survival estimate. A refinement of this method is to use Kaplan–Meier estimates of the probability of survival to time $t_i$ instead of the crude estimates. A third suggested method is to use maximum likelihood estimates of the quality of life and the probability of survival at each time point.

Each method has both advantages and disadvantages. In summary, although the maximum likelihood method is theoretically the most appealing, it is technically the most difficult to implement and software is not readily available. Otherwise the method using Kaplan–Meier estimates is generally preferable to the crude method but has the disadvantage that calculation of the CIs for the mean quality-adjusted survival time is not straightforward, requiring bootstrap methods (see page 77).

### TWiST

#### Introduction

A special QALY endpoint for comparing therapies, which incorporates both length and quality of survival into a single measure, has been developed by Gelber and colleagues in a subject-based approach to quality-adjusted survival analysis. The endpoint which they devised, TWiST, is a measure of the ‘good’ quality time experienced by the patient. It was originally developed to assess treatments for breast cancer and has also been used in the assessment of treatments for ovarian cancer.

#### Defining TWiST

TWiST is calculated for each patient by subtracting from overall survival those periods during which treatment or disease reduces their quality of life. This is equivalent to calculating QALYs for a patient using utility values of 0 for times with symptoms (REL) and toxicity (TOX), and utility values of 1 otherwise.

Suppose that for patient $i$, $TR_i$ is the time from start of treatment to symptomatic disease relapse and $TOX_i$ is the amount of time spent (not necessarily in consecutive periods) with toxicity prior to relapse (see Figure 19): then

$$TWiST_i = TR_i - TOX_i$$  \hspace{1cm} (22)

If the follow-up for patient $i$ is such that neither $TR_i$ nor $TOX_i$ have been observed, then TWiST for this patient will be censored and calculated as

$$TWiST_i = U_i - OTOX_i$$  \hspace{1cm} (23)

where $U_i$ is the length of follow-up for patient $i$ and $OTOX_i$ is the value of $TOX_i$ observed during follow-up. The follow-up for this patient may be such that the actual value of $TOX_i$ has been observed but $TR_i$ has not, in which case TWiST for this patient will be censored and can be calculated as

$$TWiST_i = U_i - TOX_i$$  \hspace{1cm} (24)

where $U_i$ is the length of follow-up for patient $i$.

The definitions of time with symptoms of disease (REL) and time with toxicity of treatment (TOX) can be adjusted for different clinical situations depending on disease and treatment under study. Defining the untoward events that can occur and determining the importance attached to each one, in terms of the amount of time subtracted from total survival, are of paramount importance in creating a meaningful TWiST measure.

![FIGURE 19 An example of TWiST](image-url)
The definitions of TWiST in the literature have, in general, been based on clinical criteria rather than patient-based measures of quality of life. However, quality-of-life data collected via questionnaires at repeated assessments over time could be used to define TWiST. For example, Glimelius and colleagues used average quality-of-life scores from various parts of their questionnaire to define TWiST for their study. They assigned a quality weighting factor of 1 to the survival time during which a patient had either an unchanged high quality of life and no signs of symptomatically progressive disease or improvement in quality-of-life estimates without being hospitalised; all other survival time was assigned a weighting factor of 0. In further examples, survival time spent with ‘normal’ quality-of-life scores (as defined by the quality-of-life instrument) is used as the definition of TWiST.

**Comparison of treatments using TWiST**

**Redefining the endpoint for treatment comparisons**

Standard methods of survival analysis, such as Kaplan–Meier, log-rank tests or Cox regression models (see chapters 11 and 12), could be used on the TWiST endpoint to compare treatments. However, as with the subject-based approach to quality-adjusted survival analysis using QALYs (see page 71), standard survival techniques will be invalid because of informative censoring (see page 11). For example, two patients may have been followed-up for the same length of time and both be censored but the one with longer duration of toxicity will be censored earlier than the one with shorter duration. This informative censoring results in the Kaplan–Meier estimates of TWiST being overestimates of the true (uncensored) TWiST.

In order to reduce (or preferably eliminate) the amount of censoring, thus overcoming the problem of informative censoring, Gelber and colleagues suggested using accumulated TWiST rather than TWiST as the endpoint for treatment comparisons. Accumulated TWiST, TWiST(L), is defined as the amount of TWiST observed within L time units from the start of treatment. If the cut-off time, L, is chosen such that all patients are followed-up beyond that time (i.e. \( L \leq U_i \) for all \( i \)), then TWiST(L) is not censored and the standard Kaplan–Meier method could be used to estimate the survivor function of TWiST(L) without bias. In addition to sensitivity analysis on the choice of L, estimating TWiST(L) for various values of L may show over time how the delay in relapse justifies earlier toxic effects.

**Alternative methods of comparing treatments**

As an alternative to comparing treatments using mean TWiST(L), Gelber and colleagues suggested that treatments could be compared with respect to TWiST(L) using quantile distance plots. The quantile distance function, which represents the horizontal distance between the survivor functions for two treatments, is plotted against the population percentile.

Willemse and colleagues suggested assessing treatments using the ratio between TWiST and progression-free survival, the TWiST index. This measures the proportion of progression-free time with ‘good’ quality of life and can be used as a kind of cost–benefit ratio. Interpretation may be difficult in situations in which some patients have censored progression-free survival times.

**TWiST analysis of MIC data**

In the MIC study, MAL was used to define TWiST. Periods within the first 18 weeks where no malaise was experienced were defined as TWiST. For each patient, these periods were summed to give a total value for TWiST. As with the QALY endpoint, the analysis of the TWiST endpoint in the MIC data is straightforward, since there are no censored survival times within the 18-week analysis period. Restricting the analysis of the MIC data to 18 weeks from study entry, for reasons of quality-of-life data availability, is effectively
equivalent to analysing accumulated TWiST(L) within 18 weeks, i.e. TWiST(18).

With no censored data, treatment comparisons can be made using either standard techniques for continuous measures or standard survival techniques without concern for bias. Since data were not normally distributed, standard survival analysis techniques were used to compare treatments. The area under the Kaplan–Meier survival curves (see Figure 20) give estimates of the mean TWiST(18) in each treatment group (mean = 6.6 weeks for CT arm and 6.5 weeks for PAL arm). The treatments were clearly very similar in terms of TWiST(18) (log-rank test \(\chi^2 = 0.04, p = 0.85\)). Thus, although treatments differed significantly in terms of survival within 18 weeks (see page 44), they did not differ with respect to TWiST(18).

Critical appraisal of TWiST and suggested extensions
The TWiST model is a simplistic way of incorporating quality-of-life into a survival-type endpoint. Brunner\(^{15}\) criticised the model for many reasons including the fact that the amounts of time deducted from overall survival to create the TWiST endpoint are arbitrary and that the model does not account for the quality-of-life experienced during these times. Brunner considered that the TWiST model obscured rather than clarified the problem of weighing up the gains and losses in quality of life associated with treatment, and advocated subjective judgement in preference.

Gelber and colleagues\(^{7}\) suggested several generalisations to their TWiST methodology:

- QALYs could be determined for individual patients, based on personalised weightings
- constant proportions of TOX and REL could be added to TWiST to avoid equating these periods to death
- application of a discount factor to adjust the value of future gains relative to the present.

The second suggestion forms a basis for the Q-TWiST methodology (see below). The TWiST model may be preferred to Q-TWiST, since it avoids the subjective quantification of times with symptoms and toxicity.\(^8\) In addition, the TWiST methodology is useful for clinical situations in which toxic therapies are administered to patients who are free of symptomatic disease.\(^8,9\) In particular, TWiST may also be useful for solving the dilemma of treatment selection when disease-free survival differences are statistically significant but overall survival differences are not, since it deals with the fact that extensions to disease-free time may be at the expense of treatment toxicity.\(^7\)

Q-TWiST
Introduction
The Q-TWiST endpoint is a natural extension of the quality-of-life oriented endpoint TWiST and an adaptation of the concept of QALYs. TWiST methodology is extended so that periods spent with toxicity or relapse are included in the analysis but are weighted to represent their quality value relative to TWiST. Thus, overall survival is scaled downwards by arbitrarily giving survival during treatment or symptoms a reduced value.

Q-TWiST was originally developed and used to assess the effects of adjuvant therapy in women with breast cancer.\(^{1,16-18}\) It has since been modified where necessary for use with other cancers, such as lung cancer,\(^{19}\) lymphoma\(^{20}\) and rectal cancer,\(^{21}\) and in other diseases, such as AIDS.\(^{22}\) Further work is being done to develop Q-TWiST so that it can be applied to neurological diseases such as multiple sclerosis\(^{25}\) and epilepsy.\(^{24}\)

Glasziou and colleagues\(^1\) provide a generalised and more mathematical background to the Q-TWiST methodology introduced by Goldhirsch and colleagues.\(^{16}\) Gelber and colleagues\(^{25}\) provide an updated overview of the Q-TWiST method that includes all recent developments and extensions to the method.

There are three main steps in applying the Q-TWiST methodology:
Quality-adjusted survival analysis

(i) defining the Q-TWiST model, which includes defining the quality-of-life oriented health states and their associated weights
(ii) partitioning the overall survival time into the defined health states in order to calculate the average time spent in each state
(iii) comparing the treatments using Q-TWiST, calculated from a weighted sum of the average time spent in each state.

Defining the Q-TWiST model

Defining health states
During the course of a clinical trial it is assumed that the patient passes through a series of progressive health states that differ in terms of quality of life. The length of time spent in each health state may be affected by the treatment being received by the patient. It is a key assumption that patients progress through the health states in sequence but that any state may be skipped. The first step in a Q-TWiST analysis, therefore, is to define these health states so that they are clinically meaningful and such that they will highlight specific differences between the treatments being compared.

In general, any number and type of health states can be defined, say $H_1, H_2, \ldots, H_k + 1$ where $H_k + 1$ usually represents death. In the original breast cancer application for which the method was developed, the following health states were defined:

- **TOX**: time having subjective toxic side-effects
- **TWiST**: time without symptoms or toxicity
- **REL**: time following systemic relapse (including time spent recovering from treatment for local recurrence).

The clinical criteria which defined those sections of a patient’s follow-up time that would fall into these health states were fully specified.

Applications of Q-TWiST tend to use either the same health states as the original breast cancer example, with slightly different definitions or a slightly modified version. Rosenthal and colleagues used the same TOX, TWiST and REL health states in their lung cancer study, as did Gelber and colleagues in their adjuvant study of rectal cancer. Gelber and colleagues used slightly modified states in their Q-TWiST analysis of symptomatic HIV patients. In this study, patients could move from TWiST to a state where first adverse events were experienced and then on to a state of disease progression. Feldstein extended the original breast cancer states, adding a further ‘recovery’ state. Gelber and colleagues suggest that in some trials, it may be necessary to define a second period of toxicity in addition to TOX, to represent the late toxic effects of treatments on a patient’s quality of life. The Q-TWiST methodology is not always immediately applicable to other diseases, since clear-cut health states may be difficult to define. Health states in most Q-TWiST applications are defined using clinical criteria but patient-assessed quality-of-life data has been used.

The Q-TWiST model

The model used for Q-TWiST is a QALY one (see page 57). In general, weights (which may be utilities) taking values between zero and one, inclusive, are allocated to each health state. These represent the quality values of each health state relative to TWiST, with 0 indicating a state as bad as death and 1 indicating perfect health. Quality-adjusted survival is then defined as the weighted sum of the time spent in each health state. The assumptions for a QALY model (see page 58) are applicable to the Q-TWiST model.

One particular form of the model is represented by

$$Q-TWiST = u_{TOX} \cdot TOX + u_{TWiST} \cdot TWiST + u_{REL} \cdot REL$$

where $TOX$ is the time spent with toxicity resulting from treatment, $REL$ is the time spent with symptoms of disease, and $u_{TOX}$ and $u_{REL}$ are the utilities associated with these periods of survival.

Discounting in the Q-TWiST model

The Q-TWiST analysis can be adapted so that a greater emphasis is placed on the earlier years. Discounting may be included by applying an appropriate transformation to the times in the data and then using these transformed times in the subsequent partitioned survival analysis.

The transformation used by Goldhirsch and colleagues is such that each time $t$ is replaced by

$$t_{\text{new}} = 1 - (1 + r)^{-t}$$

where $r$ is the required discount rate.

Partitioned survival analysis

It may seem appropriate to calculate Q-TWiST for each individual patient and then use a standard survival analysis on this new quality-of-life oriented endpoint to compare treatments. However, as with TWiST, Q-TWiST is related to the censoring mechanism; patients with poor quality of life are
The date of exiting each successive health state is regarded as an endpoint and Kaplan–Meier estimates are calculated for the time from a fixed origin, such as date of randomisation, to each endpoint. If the exit time from one health state is censored at time \( t \) for a patient, then all subsequent exit times will be censored at time \( t \). If a state \( H_i \) is skipped, then entering and exiting times are set equal and the exit time for \( H_i \) will be the same as for \( H_{i-1} \). Kaplan–Meier survival curves corresponding to each transition time can be overlaid on one graph to show the partitioning of overall survival. These are called partitioned survival plots and separate graphs should be produced for each treatment group (see Figures 21 and 22 relating to the MIC data).

For survival time \( T \), the area under a survival curve defined by the survivor function, \( S(t) \), provides an estimate, of the mean survival time, \( E(T) \), given by

\[
E(T) = \int_0^\infty S(t) \, dt
\]  

where each \( t_i \) (from \( i = 1 \) to \( i = L - 1 \)) is a death time, with \( t_0 \) defined to be zero and \( t_L \) defined to be the chosen upper time limit. The mean survival time given by the SAS Institute in their LIFETEST procedure\(^7\) gives the area under the survival curve.

Differences between successive restricted means for time from randomisation to exiting each health state give the restricted mean duration in each state. The restricted mean quality-adjusted survival is estimated by combining the restricted mean durations as a weighted sum according to the Q-TWiST model. Restricted means based on the product limit method are asymptotically unbiased and normally distributed.\(^{28} \) Consequently, statistical inferences for quality-adjusted survival can be based on the asymptotic normality of the estimates and require the calculation of standard errors of the estimates. The variance for quality-adjusted survival can be estimated from the vector of utility weights and the variance–covariance matrix for the mean times in each state.\(^1 \)

Glasziou and colleagues\(^1 \) found no simple expression for the covariance terms when dealing with restricted means and, hence, estimated the variance–covariance matrix using the bootstrap method.\(^{4,5} \) This means creating a new sample of patients, \( N \), by repeatedly sampling with replacement from the \( N \) individuals in the trial. This process is repeated thousands of times to obtain a whole series of new data sets. Restricted means for times spent in each state are calculated for each data set to produce an empirical sampling distribution, called a bootstrap sampling distribution, for the statistic. The variances and covariances computed from these values are used as the variance–covariance estimates. Gelber and colleagues\(^7 \) used variance–covariance estimates based on a series of 1000 new data sets.

censored earlier than those with better quality of life. This problem of informative censoring (see chapter 3) will give biased Kaplan–Meier estimates of the survival function and invalidate any standard survival analysis techniques. Partitioned survival analysis, a population-based approach proposed by Glasziou and colleagues,\(^1 \) provides a means of handling the problem of informative censoring. Overall survival is partitioned into the time spent in each health state and the mean duration in each state for each group are combined as a weighted sum according to the Q-TWiST model. Weighting the time spent in each health state at the group level, rather that at the individual level, avoids the need to weight censored survival times and thus overcomes the problem of informative censoring.

The date of exiting each successive health state is regarded as an endpoint and Kaplan–Meier estimates are calculated for the time from a fixed origin, such as date of randomisation, to each endpoint. If the exit time from one health state is censored at time \( t \) for a patient, then all subsequent exit times will be censored at time \( t \). If a state \( H_i \) is skipped, then entering and exiting times are set equal and the exit time for \( H_i \) will be the same as for \( H_{i-1} \). Kaplan–Meier survival curves corresponding to each transition time can be overlaid on one graph to show the partitioning of overall survival. These are called partitioned survival plots and separate graphs should be produced for each treatment group (see Figures 21 and 22 relating to the MIC data).

For survival time \( T \), the area under a survival curve defined by the survivor function, \( S(t) \), provides an estimate, of the mean survival time, \( E(T) \), given by

\[
E(T) = \int_0^\infty S(t) \, dt
\]  

The areas under the survival curves for successive endpoints can therefore be estimated and used to compute the areas between the curves giving estimates of the mean duration of each health state.\(^{26} \)

If censored data exist, the entire survival curve cannot be estimated; thus these areas can only be calculated if a specified time from randomisation is chosen as the upper time limit for the analysis, that is, the upper limit of the integral will be this time limit rather than infinity. This is usually the upper limit of observation or is based on the follow-up time of the study cohort; it should be chosen so as to reduce or, ideally, eliminate censoring. The median follow-up time, as calculated from the censoring distribution, is often used as the time limit but will only be appropriate if it adequately reduces the amount of censoring.

Mean times from randomisation to exiting each health state, restricted to the upper time limit, are calculated from the area beneath each estimated survivor function from zero to the chosen finite limit. In practice, the area under a survivor function is estimated by summing the rectangular areas under the Kaplan–Meier curve using the following formula:

\[
\text{Area} = \sum_{i=0}^{L} \hat{S}(t_i) (t_{i+1} - t_i)
\]  

where each \( t_i \) (from \( i = 1 \) to \( i = L - 1 \)) is a death time, with \( t_0 \) defined to be zero and \( t_L \) defined to be the chosen upper time limit. The mean survival time given by the SAS Institute in their LIFETEST procedure\(^7\) gives the area under the survival curve.

Differences between successive restricted means for time from randomisation to exiting each health state give the restricted mean duration in each state. The restricted mean quality-adjusted survival is estimated by combining the restricted mean durations as a weighted sum according to the Q-TWiST model. Restricted means based on the product limit method are asymptotically unbiased and normally distributed.\(^{28} \) Consequently, statistical inferences for quality-adjusted survival can be based on the asymptotic normality of the estimates and require the calculation of standard errors of the estimates. The variance for quality-adjusted survival can be estimated from the vector of utility weights and the variance–covariance matrix for the mean times in each state.\(^1 \)

Glasziou and colleagues\(^1 \) found no simple expression for the covariance terms when dealing with restricted means and, hence, estimated the variance–covariance matrix using the bootstrap method.\(^{4,5} \) This means creating a new sample of patients, \( N \), by repeatedly sampling with replacement from the \( N \) individuals in the trial. This process is repeated thousands of times to obtain a whole series of new data sets. Restricted means for times spent in each state are calculated for each data set to produce an empirical sampling distribution, called a bootstrap sampling distribution, for the statistic. The variances and covariances computed from these values are used as the variance–covariance estimates. Gelber and colleagues\(^7 \) used variance–covariance estimates based on a series of 1000 new data sets.
Comparing treatments in terms of Q-TWiST

In some situations it may be possible to compare Q-TWiST for treatments using specific utility values. These may be chosen arbitrarily if no patient-derived information is available or they could be based on a TTO or standard gamble study, for example. In a study of patients with small-cell lung cancer, Rosenthal and colleagues\(^1\) justified their choice of utility coefficients for their final conclusions \((u_t = 0.75\) and \(u_r = 0.25\)) as those which they perceived to most closely resemble the clinical experience of the patients. They compared their Q-TWiST results with those from a study by Goodwin and colleagues,\(^2\) who obtained their utility coefficients \((u_t = 0.57\) and \(u_r = 0.15\)) from a proxy group of patients and health professionals.

Other endpoints on which treatments are assessed, such as overall survival, disease-free survival and TWiST, correspond to Q-TWiST with extreme values of \(u_t\) and \(u_r\) (see Table 22).

### Table 22 Endpoints corresponding to various utility coefficients in a Q-TWiST analysis

<table>
<thead>
<tr>
<th>TOX ((u_t))</th>
<th>TWiST</th>
<th>REL ((u_r))</th>
<th>Endpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Overall survival</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>Disease-free survival</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>TWiST</td>
</tr>
</tbody>
</table>

In most cases, utility weights will be unknown. One approach in this situation is to compare treatments using a spectrum of \(u_t\) and \(u_r\) values. Rosenthal and colleagues\(^3\) examined quality-adjusted survival for a range of utility coefficients and found that the utility coefficient for TOX was far more influential than that for REL.

Another approach is to carry out a threshold utility analysis, a form of sensitivity analysis, in which trial data are used to determine the utility values which would give no difference between treatments, that is, when restricted mean quality-adjusted survival times for both treatments are equal. When there are two unknown utility coefficients, the set of values that give equal quality-adjusted survival is described by a straight line on a two-dimensional plot. CIs can be calculated for this ‘threshold line’ by finding pairs of utility coefficient values for which the lower and upper bounds of the CI for the treatment effect equals zero.\(^4\)

### Q-TWiST analysis of MIC data

The main problem with carrying out a Q-TWiST analysis on the MIC data is defining progressive health states. The MIC data do not conform to the textbook categories of TOX, TWiST and REL for two reasons. First, the patients in the MIC study have extensive disease and are receiving palliative treatments. They are therefore unlikely to experience time totally free of the symptoms of disease and, thus, do not experience TWiST. This is likely to be a common problem in many quality-of-life studies. Second, the quality-of-life data are only collected during the treatment stage of their survival; hence, during the period of analysis (i.e. 18 weeks from entry to trial) they are constantly in the TOX state.

Since the data did not conform to the standard Q-TWiST health states, alternative progressive health states had to be considered which could be derived from the data and were also clinically meaningful. MAL, expressed as a two-level variable, no malaise (MAL = 0) and malaise (MAL = 1, 2 or 3), was used to define ‘good’ and ‘poor’ states of quality of life, respectively. Various sequences of good and poor quality-of-life periods were experienced by patients within the 18 weeks from study entry and these were explored to help define the progressive sequence of health states (see Table 23).

### Table 23 Sequences of good \((G)\) and poor \((P)\) quality of life in the MIC data

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>19</td>
</tr>
<tr>
<td>GP</td>
<td>19</td>
</tr>
<tr>
<td>GPG</td>
<td>2</td>
</tr>
<tr>
<td>GGP</td>
<td>3</td>
</tr>
<tr>
<td>P</td>
<td>36</td>
</tr>
<tr>
<td>PG</td>
<td>18</td>
</tr>
<tr>
<td>PGP</td>
<td>10</td>
</tr>
<tr>
<td>PGPG</td>
<td>2</td>
</tr>
</tbody>
</table>

To incorporate the sequences experienced by all patients (i.e. to include both GPGP and PGPG), where \(G = \) good and \(P = \) poor, the definition of progressive health states needed to include five progressive health states (i.e. GPGPG or PGPGP). This was considered to be beyond both the limit of intelligibility and the limited amount of data. Hence, although a few patients would need to be excluded from the analysis, a definition consisting of four progressive health states was considered preferable.
Having decided on a definition consisting of four progressive health states, there were two possible sequences to consider; PGPG or GPGP. The first option would exclude three patients with a GPGP sequence and the second option would exclude two patients with a PGPG pattern. In general, given several options for a definition, the final choice should be based on what is most clinically meaningful. In this case, neither option seemed clinically preferable to the other and there was no clear clinical explanation for the potential continued fluctuation between the two health states. The decision, therefore, to use the second option, GPGP, was arbitrary, although it did minimise the number of patient exclusions \((n=2)\). To determine if using a different definition affects the conclusions of the analysis a sensitivity analysis should be performed.

The progressive health states were defined as GOOD1, POOR1, GOOD2 and POOR2, where the numbers indicate the first and second visits to the ‘same’ health state. With this sort of definition of progressive health states, where a patient returns to a health state previously visited, the degree of similarity between the first and second visit to a state needs to be considered. For example, a return to a good quality-of-life state after a patient has been in a poor state (i.e. GOOD2) might describe a very different experience to the first good quality-of-life state (i.e. GOOD1).

Having defined the four progressive health states, the date of exit from each formed successive endpoints for analysis. The time from entry to study to each endpoint was calculated for each patient, with entry and exit times set equal if a state was skipped. For example, if a patient was in a poor state on entry to the study, then they were assumed to have skipped the first state, and their exit time from GOOD1 was set at zero. Also, for example, if a patient exited POOR1 at 18 weeks and did not therefore experience the other two states within the 18-week analysis period, then the exit times from GOOD2 and POOR2 were both set at 18 weeks.

Kaplan–Meier survival curves for successive endpoints were calculated and overlaid to give a partitioned survival plot for each treatment group (see Figures 21 and 22). The area under each curve, with an upper time limit of 18 weeks, was obtained from SAS/STAT output \(^{27}\) (corresponding to equation 28) and gave the restricted mean survival times from date of entry to trial to each endpoint (see Table 24). The differences between successive means gave the mean time spent in each health state (see Table 25).

<table>
<thead>
<tr>
<th>Health state</th>
<th>CT arm: restricted mean survival time (weeks)</th>
<th>PAL arm: restricted mean survival time (weeks)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOOD1</td>
<td>3.37</td>
<td>4.22</td>
</tr>
<tr>
<td>POOR1</td>
<td>11.47</td>
<td>12.18</td>
</tr>
<tr>
<td>GOOD2</td>
<td>14.63</td>
<td>14.33</td>
</tr>
<tr>
<td>POOR2</td>
<td>16.29</td>
<td>14.52</td>
</tr>
</tbody>
</table>
The Q-TWiST model for the MIC data was defined as follows:

\[
\text{Q-TWiST} = t_{\text{GOOD1}} + u_1 t_{\text{POOR1}} + t_{\text{GOOD2}} + u_2 t_{\text{POOR2}}
\]

(29)

where \(t_i\) represents time spent in state \(i\) and the utility values, \(u_1\) and \(u_2\) (\(0 \leq u_1 \leq 1, 0 \leq u_2 \leq 1\)) are unknown and reflect the reduction in quality of life during the first and second visits, respectively, to the poor quality-of-life state. This model assumes that the quality of life in the good state, whether at first or second visit, was equivalent to perfect health. It also assumes that the quality of life experienced in the poor state differed depending on whether it was the first or second visit.

Utility values are not known and so a threshold utility analysis was undertaken (see Figure 23). Substituting the mean times spent in each state (see Table 25) into equation 29 gives the following equations for the two treatment groups:

\[
\text{Q-TWiST}_{\text{CT}} = 3.37 + 8.10 u_1 + 3.16 + 1.66 u_2
\]

(30)

\[
\text{Q-TWiST}_{\text{PAL}} = 4.22 + 7.96 u_1 + 2.15 + 0.19 u_2
\]

(31)

Equating equations 30 and 31 gives the threshold line (see Figure 23), indicating pairs of utility coefficients for which the treatments have equal Q-TWiST. The threshold line is given by

\[
u_1 = -1.14 - 10.50 u_2
\]

(32)

The threshold utility plane, i.e. the plane containing all possible pairs of values for \(u_1\) and \(u_2\) (\(0 \leq u_1 \leq 1\) and \(0 \leq u_2 \leq 1\)), lies completely above this line in the region where \(\text{Q-TWiST}_{\text{CT}} > \text{Q-TWiST}_{\text{PAL}}\) (see Figure 23). Thus, whatever values are chosen for \(u_1\) and \(u_2\), Q-TWiST within 18 weeks of entry to the study is always greater on the chemotherapy arm than on the palliative arm of the trial. A confidence region for the threshold line could not be calculated because bootstrap estimates of the standard error for Q-TWiST were not readily available.

**Investigating the effect of imposing an upper time limit**

Using an upper time limit for the analysis allows for the possibility that further follow-up could alter the conclusions of the analysis. There is a need to examine the potential effects of having further follow-up time, in terms of changing the conclusions based on the restricted means. Glasziou and colleagues\(^1\) suggest two possible analyses for this purpose.

One method is to examine the difference in Q-TWiST for a range of truncation times. Using a reasonable set of values for the utility weights, Q-TWiST for each treatment can be calculated for a range of upper time limits from 0 to \(L\), the chosen upper time limit. The difference between treatments can then be plotted against the restriction time. This analysis can be enhanced further by plotting not only the treatment difference for a particular set of utility coefficients as a curve but also the whole range of possible treatment effects for all possible utility coefficients between 0 and 1 as a shaded region. This is called the Q-TWiST gain function.\(^{25}\) Based on clinical knowledge, a subjective judgement can be made as to whether the graph is sensible and has reached a level of stability.

The second method suggested\(^1\) is to fit a parametric survival model either to the entire distribution or just to the tail of the distribution. In this way, quality-adjusted survival can be extrapolated beyond the time limit \(L\) and the conclusions regarding which treatment is superior can be examined. Gelber and colleagues\(^{30}\) discuss projecting survival estimates beyond
the follow-up period of a trial by fitting parametric survival models to the tails of Kaplan–Meier survival curves and using the estimated models to make projections. They suggest that Weibull and log-normal models are likely to be the most useful. For the method to be successful in practice, there must be sufficient data in the tail of the survival curve for adequate estimation of a parametric model.

**Extending Q-TWiST to incorporate covariates**

In the standard Q-TWiST method, covariates are included by stratifying the sample of patients by the covariate values and performing a separate analysis on each stratum. The problem with this approach is that the sample sizes in some strata may not be sufficient for estimating the mean quality-adjusted survival with reasonable accuracy. In order to avoid stratification and thus enable the entire sample of patients to be used, Cole and colleagues extended the Q-TWiST methodology to allow for covariates using a Cox proportional hazards model. Accelerated failure time regression could be used in a similar manner.

The survivor functions, for time from study entry to exiting each successive health state, are estimated using the results from Cox proportional hazards models fitted to each endpoint, instead of the Kaplan–Meier method. When comparing two treatments, the treatment can be included either as a covariate in a Cox regression or, if the proportional hazards assumption is not valid, a stratified regression can be used. The proportional hazards assumption can be checked for each health state endpoint by plotting log (–log S(t)) against log t for the different levels of a covariate. If the proportional hazard assumption is valid, the curves will be parallel.

The use of Cox regression models allows quality-adjusted survival, for specific sets of covariate values, to be estimated for each treatment, allowing treatment comparisons in specific subgroups of patients. In a study of zidovudine therapy in asymptomatic HIV-infected patients, Lenderking and colleagues used proportional hazards models to investigate differences between treatments in terms of quality-adjusted survival, for different baseline values of CD4+ cell counts. In a breast cancer application, Gelber and colleagues used covariate values to define two patient profiles corresponding to a good and a poor prognosis, and used Cox models to compare treatments in terms of quality-adjusted survival in both types of patient. A further application in melanoma enabled the effect of the initial stage of disease on quality-adjusted survival to be investigated.

**Extending Q-TWiST to perform meta-analysis**

Meta-analysis is an analytical technique used to summarise quantitatively the results from a number of different studies all addressing the same basic research question. Cole and colleagues developed a meta-analysis methodology for combining the Q-TWiST results from individual trials in which patient-level data is not required. The method is a modified version of the standard Q-TWiST analysis and is summarised by Gelber and colleagues. Regression models are used to combine trials, enabling an overall comparison of treatments in terms of quality-adjusted survival that accounts for the differing follow-up times.

The method was applied to data from eight clinical trials comparing adjuvant chemotherapy with no adjuvant systemic therapy in pre-menopausal women with breast cancer. Gelber and colleagues also used the technique to combine the results from nine trials in order to compare the effect of adjuvant chemotherapy plus tamoxifen with tamoxifen alone in post-menopausal node-positive breast cancer.

**Critical appraisal of quality-adjusted survival analysis**

Quality-adjusted survival analysis is based on the concept of combining quality-of-life and survival data in a QALY model, with TWiST and Q-TWiST being special forms of QALYs. The assumptions and criticisms of the QALY model (see pages 58 and 61) also therefore apply to quality-adjusted survival analysis and these assumptions have been discussed in relation to a Q-TWiST model. A parametric approach to quality-adjusted survival analysis, as discussed later in chapter 18, allows a more general form for the function combining quality-of-life and survival data and may overcome some of the assumptions of the QALY model.

The concept of creating a quality-adjusted survival time for each patient in a study to be used as an endpoint in treatment comparison instead of survival time seems sensible. However, subject-based approaches in general suffer from the problem of informative censoring. Thus, population-based approaches to quality-adjusted survival analysis, which overcome this problem, are preferable in that respect.
Partitioned survival analysis is a population-based approach to quality-adjusted survival analysis, usually used with a Q-TWiST model but applicable to any QALY model. It has the advantage of being able to include covariates and requires few assumptions, being based on either the non-parametric methods of Kaplan and Meier or the semi-parametric method of Cox regression. It has two main limitations:

- the need to use progressive health states
- the need to restrict the period of analysis to an upper time limit.

Further, it has the disadvantage of not being readily accessible to researchers, since calculation of CIs for quality-adjusted survival uses the bootstrap method.

Partitioned survival analysis is sometimes difficult to apply because of the need for progressive health states. It may be possible to overcome this by specifying different phases of the same state, as in the MIC study, but this can become clumsy and may lose clinical meaning. An alternative method of analysis, multistate survival analysis, discussed in the next chapter, does not require progressive states and is thus a more flexible way of modelling the data.

There is no unique way of dividing the survival time of patients into periods of differing quality of life and the accuracy depends on the frequency of quality-of-life assessments. Different divisions should be considered as part of a sensitivity analysis. Missing quality-of-life data will also cause difficulties since, although quality-adjusted survival analysis deals with informative drop-out caused by death, it does not deal with other reasons for drop-out. Values for the missing data can be imputed or it may be possible to incorporate into the model the time spent as a drop-out, with an appropriate weighting factor to reflect quality of life. Alternatively, methods that explicitly model the drop-out process, such as multistate survival analysis (see chapter 17) or simultaneous modelling (see chapter 19) should be considered.

A further limitation of partitioned survival analysis is the need to restrict the period of analysis to an upper time limit. In some situations, as with the MIC study, the quality-of-life data may only be collected for a limited time, and the period of analysis will automatically be restricted to an upper time limit. Otherwise, methods to investigate the effect of imposing an upper time limit should be applied. The parametric approach to quality-adjusted survival analysis, as discussed in chapter 18, overcomes this limitation.

Quality-adjusted survival is calculated using a weighted sum of time spent in each health state. The choice of values for the weights is based on the perceived quality of life experienced in each health state and can be problematic. Consideration of a range of possible values should form part of a sensitivity analysis and, in particular, threshold utility analysis allows all possible combinations of values to be investigated.

References


2. Beacon HJ. The statistical analysis of self assessed quality of life data in cancer clinical trials [PhD thesis]. London School of Hygiene and Tropical Medicine, University of London; 1996.


Chapter 17

Multistate survival analysis

Introduction

Multistate models have been advocated in a number of reviews and discussions regarding the analysis of quality of life data as a possible means of analysing quality of life and survival data simultaneously. They were first proposed in a medical context by Fix and Neyman in 1951. The approach is discussed in some detail by Olschewski and Schumacher in their paper on the analysis of quality-of-life data in cancer clinical trials.

The statistical background to multistate survival analysis is derived from the analysis of event-history data and stochastic processes. Many of the theoretical aspects of multistate models fall within a counting process framework. The study of events occurring in individuals over time generates event-history data. In studies such as these, individuals can be thought of as occupying one of a finite number of states at any point in time and the movement between states can be described by conditional probabilities or transition rates. This dynamic process is known as a stochastic process. Quality-of-life assessment in clinical trials generates event-history type data, with events being defined as entry and exit from pre-defined health states. The movement of individuals between quality-of-life states can then be considered as a stochastic process and modelled accordingly.

As with quality-adjusted survival analysis, multistate survival analysis starts by defining a finite number of health states, including death, that patients experience during the study. Defining these health states and the possible transitions between them describes the multistate model. The transition rates, which describe the movement between health states, can then be modelled, possibly using covariates. In this way, the time-dependent structure and dynamic nature of quality-of-life data can be incorporated into an analysis comparing treatments and the effect of explanatory variables on transition rates from one state to another can be investigated. Multistate models overcome some of the limitations of the Q-TWiST approach, such as the requirement for progressive health states.

Multistate models in survival analysis have been applied in a variety of clinical settings such as diabetes, liver transplantation, bone marrow transplantation, heart transplantation, breast cancer, prostate cancer and HIV. These applications, however, used clinical criteria rather than quality-of-life data to define health states. There are limited examples of applications of multistate models to quality-of-life data. In this study, the multistate survival analysis methodology is illustrated by application to quality-of-life data collected in the MIC study.

Defining the model

Health states

The set of health states chosen for the model should be clinically meaningful and fully describe the experiences of the patients. The states should be mutually exclusive and exhaustive. The number of states should be restricted so that the model does not become overcomplicated and also to ensure that the number of patients passing from one state to another is sufficient for adequate modelling of the data.

There are two main types of health states. A transient state is one that a patient can pass through during the course of their follow-up and an absorbing state is one that a patient cannot leave once it has been entered. The standard model for survival analysis corresponds to the simplest multistate model, in which the patient can be in one of two possible states, a transient ‘alive’ state or an absorbing ‘death’ state. The competing risks model is an extension of this two-state survival model and forms a multistate model, with one transient alive state and several absorbing death states corresponding to different causes of death. In terms of modelling quality-of-life data, it is more relevant to extend the simple two-state survival model so that there are several transient alive states and one absorbing death state.

The simplest version of this sort of multistate model is the three-state disability model (also called the illness-death model), where there are two transient alive states: ‘alive without disability’
and ‘alive with disability’, and one absorbing death state. Andersen\textsuperscript{15} modelled diabetes data using a model with this structure (see Figure 24).

**Quality-of-life states for the MIC study**

The MIC quality of life data were modelled using a three-state illness-death model (see Figure 25). The patients’ health states were determined from the malaise question, which asked if they had been feeling generally ill. Patients were categorised as being in an ‘alive and well’ state if the patient had no malaise (MAL = 0) and were categorised as ‘alive and ill’ if the patient had malaise at any level (MAL = 1, 2 or 3).

The health states used in the model could have been defined using different cut-offs for responses to the malaise question or could have been based on other questions asked in the study or on a cut-off value for the MQS. A more complex model with more than three states was considered but such a formulation would have meant that there were too few transitions for adequate modelling of the data.

**Health state transitions**

The movement between states can be described using either a transition probability or a transition rate. A transition probability is the likelihood of an individual moving from one state to another within a specified period; a transition rate is the instantaneous potential of transition at any point in time.\textsuperscript{26} While the transition probability can take values between 0 and 1, the transition rate, sometimes called a transition intensity, has no upper bound. In the simple two-state survival model, the transition rate from a transient alive state to an absorbing death state is the standard hazard rate function for the survival time distribution (see page 41).

The two measures are related\textsuperscript{26} and, when the instantaneous transition rate $r$ remains constant during the period, $t_1$ to $t_2$, a transition probability $p$ can be estimated using

$$p = 1 - \exp (-r (t_2 - t_1))$$  \hspace{1cm} (33)

In some situations, patients in the trial may experience ‘recovery’ during the follow-up period and in these circumstances the model may need to include reverse transitions, allowing the ability to return to a state previously occupied. Andersen and colleagues\textsuperscript{27} model data from a liver cirrhosis trial as a three-state illness-death model but they allow patients with low prothrombin index to recover and return to the ‘alive with normal prothrombin index’ state, thus incorporating reverse transitions in their model. If reverse transitions are possible then repeat transitions may occur, that is, a patient may experience a particular transition more than once during follow-up. In some circumstances, it may be preferable to model repeat and initial transitions as separate events.\textsuperscript{28} This would only be possible if the number of patients experiencing repeat transitions was large enough for adequate modelling.

In the MIC study, patients moved between health states, often several times, until finally moving to the absorbing state of death. The model, therefore, included reverse and, hence, repeat transitions (shown in Figure 25 by arrows in both directions between the alive states).

**Assumptions of the underlying stochastic process**

In the most general model, the transition rates would depend on the whole history of the patient. However, it is often reasonable to assume that at any time point the state currently occupied by a patient contains all the information relevant to that patient’s future course. Under this assumption, the model represents a Markov process. If the transition rates from each state are conditional on the duration of time spent in the state, often called the sojourn time, then the model represents a semi-Markov process.\textsuperscript{9}

It may be necessary to define the health states of the model so that the Markov assumption is
valid. For example, in the MIC study, if the transition of patients between the alive—well and death states depended on whether the patient had originally been in the alive—ill state, then it would be necessary to create two alive—well states: an ‘always been alive and well’ state and a ‘was ill but now alive and well’ state. In this way the validity of the Markov assumption could be retained.

If the transition rates remain constant over time (i.e. are independent of time) then the process is time-homogeneous; otherwise, if they vary over time (i.e. are functions of time), the process is time-inhomogeneous. In some situations the transition rates may be ‘piecewise’ constant, that is, constant over defined subdivisions of the follow-up time.

**Modelling the transition rates**

In a standard survival model, the transition rate from the alive state to the dead state is commonly represented by a Cox regression model.29 The application of Cox regression models is limited to the more general multistate framework, which allows several transient disease states between entry to study and death, has been discussed.19,30–33 The exact dates of transient disease states between entry to study and general multistate framework, which allows several application of Cox regression models to the more

**Semi-parametric or fully-parametric**

The transition rates can either be modelled semi-parametrically using a Cox regression model where the underlying baseline transition rate is left unspecified19 or parametrically by assuming the transition times follow a specific distribution, thus giving a parametric form to the baseline transition rate.35,36 The most commonly used distributions are the exponential and Weibull distribution, the exponential being just a special form of a Weibull distribution (see page 41). If an exponential distribution is assumed for the transition times from state i to state j, then the underlying baseline transition rate is constant:

\[
\lambda_{0ij}(t) = \lambda_{ij}
\]  

The type of process is related to the timescale on which time in the model for a transition rate is measured, that is, it relates to the time at which the clock starts. If, in modelling the transition rate from state i to state j, time \( t \) is measured from the time of entry to the study (i.e., \( t = 0 \)), then the model represents a Markov process, since duration of time in state i is not included. Otherwise, if time is measured in the model from time of entry to state i, say \( w \) in relation to study entry time, then time is included in the model as \( t = w \) rather than \( t \) and, with sojourn time now included, the model represents a semi-Markov process.20

In a semi-Markov process, the clock is effectively ‘reset to zero’ every time a state is entered.9

Information regarding the history of the process prior to entering state i may be included as covariates. In particular, when individuals can experience the same transition more than once, the transition rate from state i to state j may depend on such aspects as whether state i has been occupied before, the number of times state i has been visited before and the total time previously spent in state i. If the covariates in the model do not include information regarding states prior to the current one, then the model implicitly assumes that the changes of state form a Markov process.19 The time from study entry to entry to state i could also be included as a covariate.20 When the time origin is taken as date of entry to study, then the covariates containing information on history before entering state i will be time-dependent, while if the origin is taken as date of entry to state i, then the history is already determined at the origin and the information will be included as fixed covariates.

The transition rate from state i to state j, \( \lambda_{ij}(t) \) can be modelled using a Cox regression model as follows:

\[
\lambda_{ij}(t) = \lambda_{0ij}(t) \exp (\beta_j^T x_{ij})
\]

where \( \lambda_{0ij}(t) \) is a baseline transition rate for the transition from i to j, \( x_{ij} \) is a vector of covariates specific to that transition and \( \beta_j \) is a vector of unknown regression coefficients specific to that transition. The model could be generalised to include time-dependent covariates by replacing \( x_{ij} \) with \( x_{ij}(t) \).

**Markov or semi-Markov**

When modelling the transition rates, consideration needs to be given to whether the process is Markov or semi-Markov. Semi-Markov processes have been discussed in relation to multistate models.30–32,34 In a Markov process, as discussed above, the transition rate to another state depends only on the present state occupied, while in a semi-Markov process the transition rate is also dependent upon the duration of time spent in the present state.13

**Health Technology Assessment** 1999; Vol. 3: No. 10
and the model is, thus, time-homogeneous, while for a Weibull distribution the underlying baseline transition rate takes the form

\[ \lambda_{ij}(t) = \lambda_{ij} y_{ij}^{\gamma_{ij}-1} \]  

(36)

and the model is time-inhomogeneous. If an exponential distribution is assumed for the transition times then, since the underlying transition rate is not dependent on time, the models under the assumption of either a Markov or a semi-Markov process will be equivalent. This, however, is not the case for other distributions, such as Weibull, since the baseline transition rate changes over time.

**Estimating parameters**

Kay\(^{39}\) showed that the partial likelihood for a transition from state \(i\) to state \(j\) is identical to the partial likelihood for the standard Cox regression model,\(^{37}\) with transitions from \(i\) to states other than \(j\) treated as censored data. In situations where individuals may experience the same transition more than once, the partial likelihood is still valid.\(^{39}\) Thus, for each transition in the model, estimates of the \(\beta\) regression coefficients are obtained by maximising the relevant partial likelihood.

Significance testing and CIs for the \(\beta\) parameters can be based either on the asymptotic normality of the distribution of the estimators or on the large sample likelihood ratio test for nested models,\(^{19}\) where changes in \(-2\log(\text{likelihood})\) are compared to a chi-square distribution. As with a standard regression analysis, the selection of covariates can be made using backward elimination, forward elimination or a stepwise approach.

**Model checking**

The assumption of proportional hazards for covariates in the model should be checked using standard graphical techniques and the fit of model can be assessed using residual analysis.\(^{38}\) These standard techniques require independent observations and so, in the case of repeat transitions, assessment of the model is problematic.

**Computing issues**

When modelling the transition rate from state \(i\) to state \(j\) as a semi-Markov process, that is, with the time origin taken as date of entry to state \(i\), the PHREG procedure (SAS Institute)\(^{39}\) can be used to fit semi-parametric models and the LIFEREG procedure (SAS Institute)\(^{39}\) can be used to fit parametric models.

In the parametric modelling of transition times \(T\), the SAS software fits an accelerated failure time model, which has the general form

\[ \log (T_i) = \mu + \alpha^T x_i + \sigma \epsilon_i, \quad i = 1, \ldots, n \]  

(37)

where \(x_i\) is a vector of covariates, \(\alpha\) is a vector of unknown regression coefficients, \(\sigma\) is an unknown scale parameter, \(\mu\) is an unknown intercept parameter, and \(\epsilon\) is an error term where errors are assumed to come from a known distribution. If transition times are assumed to follow a Weibull distribution then there is a direct correspondence between the parameters under an accelerated failure time model (\(\mu\), \(\sigma\) and \(\alpha\) in equation 37) and those under a proportional hazards model (\(\lambda\), \(\gamma\) and \(\beta\) in equations 34 and 36). The relationships are given by the following formulae:\(^{38}\)

\[ \lambda = \exp \left( -\mu / \sigma \right) \]

\[ \gamma = 1 / \sigma \]

\[ \beta = -\alpha / \sigma \]  

(38)

When modelling the transition rate from state \(i\) (the base state) to state \(j\) as a Markov process, that is the time origin is taken as date of entry to study, SAS software cannot be used since it will not allow for times of entry to the base state other than zero. Other statistical software, in particular the \texttt{agreg} function in S-PLUS (Statistical Sciences),\(^{11}\) allows entry times to the base state other than zero in semi-parametric models and thus can be used for Markov modelling. Specially written functions in S-PLUS\(^{23}\) needed to be written to fit Weibull–Markov models.

**Modelling transition rates in the MIC data**

**Introduction**

A three-state illness–death model was used to describe the MIC quality-of-life and survival data (see Figure 25). For reasons of data availability, the analysis was restricted to the 18-week period from study entry (see page 14). Before any modelling of the data could be undertaken, various assumptions relating to the quality-of-life data needed to be made.

Quality of life was assessed at distinct time points and assumptions were necessary to infer values over continuous time. These assumptions were the same as those made to calculate QALYs (see page 58). In particular, changes in quality of life were assumed to occur at the midpoint between assessments. This enabled the exact dates of
transition between the two alive states to be estimated. Dates of death for patients in the study were known and so exact dates for transitions to death were available. Sensitivity analysis should be used to assess the impact of the assumptions on the conclusions of the analysis.

Drop-outs in the data were dealt with by imputing values using a 'last value carried forward' approach. As well as other methods of imputation (see Table 7), other strategies for dealing with drop-outs in the context of multistate survival analysis were considered. Patients who dropped out of the quality-of-life study because of death were not a problem since their transition to death was included in the model. Transition times for patients who dropped out for reasons other than death could be censored at date of drop-out, with date of drop-out possibly defined as the next planned assessment date, although this is likely to result in informative censoring. Another option considered was to include 'drop-out' as a state in the model but this made the model too complex for the limited amount of data available.

Models fitted to the data
During the 18-week period from study entry, 113 transitions in total were experienced and these data are used to model the four transition rates (see Table 26). The transition rates were modelled using a Cox regression model with only a covariate for treatment (trt) included. In all cases the comparison is chemotherapy (trt = 1) against standard palliative treatment (trt = 0). Although treatment was the only covariate in the model, Kaplan–Meier survival estimates and log-rank tests could not be used for treatment comparison since the MIC data includes repeat transitions and therefore observations would not be independent.

For illustrative purposes a variety of models were fitted, all adaptations of the basic Cox regression model according to whether the process was assumed to be Markov or semi-Markov, and whether a parametric form for the underlying transition rate was specified (see Table 27).

The transition rates were analysed one at a time. When modelling the transition rate from state \( i \) to state \( j \), only individuals who occupied state \( i \) at some point during the analysis period contributed to the model, although individuals who passed out of state \( i \) at some point and then returned back to it later contributed twice to the model. If an individual was in state \( i \), then they were 'at risk' for the state \( i \) to state \( j \) transition. If they moved to state \( k \) rather than state \( j \), then the time for the \( i \)--\( j \) transition was censored at the time of passing to state \( k \). If they did not move from state \( i \) before the 18-week time limit, the \( i \)--\( j \) transition time was censored at 18 weeks.

The set-up of the data for analysis was different depending on whether the model being fitted was Markov (Models 1 and 4) or semi-Markov (Models 2 and 5). The data set-up for the time-homogeneous model (Model 3) was the same as that for a semi-Markov model.

- **Markov model**: The data for the transition from state \( i \) to state \( j \) consisted of one line per patient per visit to state \( i \). Each line consisted of the time of entry to and exit from state \( i \) in relation to trial time and an indicator variable which was 1 for an exit to state \( j \), representing an actual event, and 0 for either an exit to state \( k \) or an exit time censored at 18 weeks.

- **Semi-Markov model**: The data for the transition from state \( i \) to state \( j \) consisted of one line per patient per visit to state \( i \). Each line consisted of the duration of time spent in state \( i \) on that visit and an indicator variable which was 1 for an exit to state \( j \), representing an actual event, and 0 for either an exit to state \( k \) or an exit time censored at 18 weeks.

S-PLUS was used to fit Markov models, with the agrég function used for semi-parametric models (Model 1) and specially written functions used for Weibull models (Model 4). SAS/STAT software was used to fit semi-Markov models, with the PHREG procedure used for semi-parametric models.

**TABLE 26** Frequency of each transition in the MIC data

<table>
<thead>
<tr>
<th>Transition</th>
<th>Number at risk</th>
<th>Total</th>
<th>CT arm</th>
<th>PAL arm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alive and well → alive and ill</td>
<td>79</td>
<td>39</td>
<td>28</td>
<td>11</td>
</tr>
<tr>
<td>Alive and ill → alive and well</td>
<td>106</td>
<td>37</td>
<td>27</td>
<td>10</td>
</tr>
<tr>
<td>Alive and well → dead</td>
<td>79</td>
<td>10</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Alive and ill → dead</td>
<td>106</td>
<td>27</td>
<td>13</td>
<td>14</td>
</tr>
</tbody>
</table>
of the underlying distributions for the transition times. The five different models were only fitted for illustrative purposes. The effect of treatment on each transition rate is the main interest and thus the estimates of the treatment regression parameter given by each model are given here (see Table 28). Determination of which model best fits the data was not relevant here, hence,

---

**TABLE 27** Models fitted to transition rates in the MIC study

<table>
<thead>
<tr>
<th>Model</th>
<th>Distribution for transition times</th>
<th>Process</th>
<th>Model for transition rate from state ( i ) to state ( j )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Unspecified Markov</td>
<td>Markov</td>
<td>( \lambda_{ij}(t</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>where: ( \lambda_{0ij}(t) ) is baseline transition rate ( \lambda_{ij}(t) ) is constant over time ( \lambda_{ij}(t</td>
</tr>
<tr>
<td>2</td>
<td>Unspecified Semi-Markov</td>
<td>Semi-Markov</td>
<td>( \lambda_{ij}(t</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>where: ( \lambda_{0ij}(t) ) is the baseline transition rate ( \lambda_{ij}(t - t_i) ) is the baseline transition rate ( \beta_{ij} ) is the regression parameter for treatment ( t_i ) represents time from study entry to entry to state ( i )</td>
</tr>
<tr>
<td>3</td>
<td>Exponential, i.e. transition rates assumed to be constant over time</td>
<td>Not applicable</td>
<td>( \lambda_{ij}(t</td>
</tr>
<tr>
<td></td>
<td>(Markov equivalent to semi-Markov)</td>
<td></td>
<td>where: ( \lambda_{ij} ) is the constant baseline transition rate ( \beta_{ij} ) is the regression parameter for treatment ( t ) can be either time from study entry or time from entry to state ( i )</td>
</tr>
<tr>
<td>4</td>
<td>Weibull, i.e. transition rates assumed to vary over time</td>
<td>Markov</td>
<td>( \lambda_{ij}(t</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>where: ( \lambda_{ij} ) and ( \gamma_{ij} ) are the scale and shape parameters for the Weibull distribution ( \beta_{ij} ) is the regression parameter for treatment ( Y_{ij}(t) ) is the ‘at-risk’ process described in Model 1 ( t ) represents time from study entry</td>
</tr>
<tr>
<td>5</td>
<td>Weibull, i.e. transition rates assumed to vary over time</td>
<td>Semi-Markov</td>
<td>( \lambda_{ij}(t</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>where: ( \lambda_{ij} ) and ( \gamma_{ij} ) are the scale and shape parameters for the Weibull distribution ( \beta_{ij} ) is the regression parameter for treatment ( t_i ) represents time from study entry to state ( i )</td>
</tr>
</tbody>
</table>

---

**TABLE 28** Comparison of treatment regression parameters in all five multistate models

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Well → ill</td>
<td>0.48 (−0.22, 1.18)</td>
<td>0.47 (−0.23, 1.17)</td>
<td>0.46 (−0.24, 1.15)</td>
<td>0.47 (0.10, 0.84)</td>
<td>0.45 (−0.25, 1.15)</td>
</tr>
<tr>
<td>Ill → well</td>
<td>0.40 (−0.32, 1.13)</td>
<td>0.31 (−0.41, 1.04)</td>
<td>0.34 (−0.39, 1.07)</td>
<td>0.36 (−0.02, 0.74)</td>
<td>0.34 (−0.39, 1.06)</td>
</tr>
<tr>
<td>Well → dead</td>
<td>−1.39 (−2.74, −0.04)</td>
<td>−1.26 (−2.61, 0.09)</td>
<td>−1.33 (−2.68, 0.03)</td>
<td>−1.40 (−2.53, −0.27)</td>
<td>−1.27 (−2.63, 0.08)</td>
</tr>
<tr>
<td>Ill → dead</td>
<td>−0.77 (−1.53, −0.01)</td>
<td>−0.71 (−1.46, 0.05)</td>
<td>−0.73 (−1.48, 0.03)</td>
<td>−0.78 (−1.32, −0.23)</td>
<td>−0.70 (−1.45, 0.06)</td>
</tr>
</tbody>
</table>

---

models (Model 2) and the LIFEREG procedure used for parametric models (Models 3 and 5).

**Results for treatment comparison**

The semi-parametric models (Models 1 and 2) give estimates only for the treatment regression parameter, while the parametric models (Models 3, 4 and 5) also give estimates for the parameters of the underlying distributions for the transition times. The five different models were only fitted for illustrative purposes. The effect of treatment on each transition rate is the main interest and thus the estimates of the treatment regression parameter given by each model are given here (see Table 28). Determination of which model best fits the data was not relevant here, hence,
parameter estimates for the exponential and Weibull distributions have not been given.

The treatment parameter estimates for the five models are reasonably comparable, although there is some discrepancy regarding the width of the CIs, which warrants further investigation. The results for the semi-parametric Markov model (Model 1) suggest that, although the effects were not statistically significant, chemotherapy increased the risk of transition between the two different quality-of-life states. There was also statistically significant evidence to suggest that chemotherapy reduced the risk of transition to death from both the well and ill quality-of-life states. Although the significance of the treatment effects in the equivalent semi-Markov model (Model 2) differed from the Markov model, the point estimates and also, to a great extent, the CIs did not greatly differ. Thus, in this study, taking account of duration of time in a state has relatively little effect on the transition rates out of the state. The significance of the treatment effects for the parametric models (Models 3–5) also differ but, again, there is little to discriminate between the point estimates and CIs, indicating that assuming such underlying distributions for the transition times between states does not greatly influence the transition rates.

**Alternative modelling approaches when exact transition times are not known**

**Introduction**

In many prospective studies, although the exact time of death is known, it is not possible to observe the actual transitions of patients from one health state to another, all that is known for a patient is the health state occupied at certain follow-up times. The transition times in this situation are interval-censored data and the transition rates cannot be modelled directly using the methods discussed above. One approach, used in the analysis of the MIC data above, is to estimate the exact transition dates from the data. The most widely used approximations to exact transition dates are the actual follow-up dates or the mid-point between follow-up dates. Estimates of transition rates obtained from approximated data may be incorrect. Alternative approaches to modelling transition rates, that account for the interval censored data, are discussed here. The application of these methods is difficult in practice because standard software cannot be used, although special FORTRAN computer programs have been developed.42–44

**Models in which time is continuous**

In situations when exact transition times are not known, Kay42 proposed modelling transition rates using a general continuous-time Markov chain. Transition rates are assumed to be constant over time, although piecewise constant transition rates could be accommodated, by splitting time at a number of pre-defined points and estimating separate transition rates for each period. This method of multistate modelling has been applied to the data in a number of studies.16,22,27,45–47

Estimates of transition rates are obtained using maximum likelihood methods, where the likelihood function is formulated from the probability of each individual’s passage through the health states. The standard relationship between the matrix of transition rates and the matrix of transition probabilities, as described by Kolmogorov equations,10,11 enables the maximum likelihood estimation of transition rates. The maximisation process is an iterative procedure and computer routines in FORTRAN are available to compute parameter estimates together with their standard errors.42

Covariates such as treatment can be included in the modelling process as a proportional factor over the constant baseline transition rates, as in a Cox model,42 so that the transition rate from state $i$ to state $j$ can be represented by

$$\lambda_{ij} x = \lambda_{0ij} \exp \left( \beta^T \mathbf{x} \right)$$  \hspace{1cm} (39)

where $\lambda_{0ij}$ is the constant baseline transition rate, $x$ is the vector of covariates and $\beta_{ij}$ is the vector of associated regression parameters for that particular transition. A computer program written in FORTRAN called MARKOV43 has been designed to allow the inclusion of covariates in the model. The number of model parameters may be reduced either by assuming the effect of the covariates is either the same for all progressive and for all regressive transitions or is the same for all transitions.16

Hypothesis testing can be carried out using either likelihood ratio tests or Wald tests but opinions regarding which is optimal for various circumstances lack consistency. In the comparison of transition rates (i.e. $H_0: \lambda = \lambda_0$), the Wald test, which calculates a $\chi^2$ test statistic based on estimated transition rates and associated covariances, has been advocated.42 Gentleman and colleagues,22 however, have sometimes found the normal approximation to be unreliable and therefore recommend the use of likelihood ratio tests for
such comparisons. Other hypothesis tests of interest might be regarding the comparison of regression parameters for different transition rates (i.e. $H_0: \beta_j = \beta_m$). In this case, Marshall and Jones\textsuperscript{16} recommend the Wald test, whereas Kay\textsuperscript{42} recommends likelihood ratio tests. Marshall and Jones\textsuperscript{16} recommend likelihood ratio tests for covariate selection within the modelling of each transition rate.

Modelling transition times using a general continuous-time Markov chain is only valid when the pattern of follow-up times, called the examination scheme, is noninformative for the disease process.\textsuperscript{48} This condition is equivalent to noninformative censoring in survival analysis. In situations where examinations take place because of patient self-selection, the examination scheme may be informative and estimates of transition rates may be biased.

Related methods have been described for studies in which individuals move from state to state, according to a continuous-time Markov process, but examination times are fixed. Methods have been proposed for situations in which the data is in aggregate form, that is, it consists of the number of individuals in each state at the fixed time points.\textsuperscript{49} Alternatively, if individual patient data is known, a more flexible method that assumes transition rates are proportional over intervals of 90 days where time point 0 was at study entry and time point 1 was 90 days later, and so on. They then determined a patient’s state at each time point from the last available measurement in the preceding 90-day interval.

Various considerations are relevant when deciding on the length of the time interval. The model assumes that each patient may make only one transition during each cycle; hence the time interval needs to be kept short enough for this assumption to be tenable but needs to be long enough to allow a reasonable number of patients to experience transitions. The tenability of the Markov property will depend on the length of the time interval employed. For a time-homogeneous Markov process, the length of the cycles must be uniform, while for time-inhomogeneous processes the cycles may vary in length.

**Time-inhomogeneous process**

If the Markov process is time-inhomogeneous then each cycle, which may vary in length, will have its own transition probability matrix. Each matrix can be estimated from the data and can be used to estimate, for each fixed time point $t$ defining the cycles, the probabilities of being in each state.\textsuperscript{51}

Suppose the process consists of $s$ states and $P_{t-1}$ is the $(s \times s)$ transition matrix for the time interval $(t-1, t)$, then the maximum likelihood estimate for the $(i, j)$th element of the matrix is given by

$$P_{t-1}(i, j) = n_{t-1, i}(i, j) / n_{t-1}(i)$$

where $n_{t-1, i}(i, j)$ is the number of transitions from state $i$ at time $t-1$ to state $j$ at time $t$, and $n_{t-1}(i)$ is the number of individuals in state $i$ at time $t-1$ that are not censored in the interval $(t-1, t)$.

Given these estimated transition matrices, the vector $p_i$ containing the unconditional probabilities of being in each state $i$ ($i = 1$ to $s$) at time $t$ can be estimated using

$$p_i = p_0 \prod_{k=1}^{t} P_{k-1, k} = p_{t-1} P_{t-1}$$

where $p_0$ is the initial probability vector estimated from the data using the proportion of patients in each state at time 0.

This methodology is analogous to a standard life-table analysis, in that the probabilities are estimated using observed frequencies and the cumulative survival probability for any time $t$ is the product of all the preceding interval survivals.\textsuperscript{51} The method is based on the assumption that incomplete follow-up is not related to patient outcome.

**Time-homogeneous process**

If the Markov process is assumed to be time-homogeneous, then the probability of changing from state $i$ to state $j$ between times $t-1$ and $t$ is equal to that between times $t$ and $t+1$. Based on this assumption, estimates of transition probabilities can be calculated from the data. The probability of transition from state $i$ to state $j$ can be estimated by calculating

$$p_{ij} = \sum_{t} n_{t-1, i}(i, j) / \sum_{t} n_{t-1}(i)$$

\textsuperscript{42}}
where \( n_{t-1}, (i, j) \) and \( n_{t-1} (i) \) are as specified above in equation 40.

These can be presented in the form of a transition probability matrix, \( P \), which provides a useful summary description of the movement between states. The transition probabilities in the matrix can be modelled using log-linear models or the matrix can be calculated to the \( n \)th power to indicate, for each starting state, the probabilities of being in each state after \( n \) cycles. These can be plotted on a graph over time.

The expected duration in each state can be calculated from the transition probability matrix. A submatrix, \( P^* \), of the transition probability matrix can be formed from the probabilities of transition between the non-absorbing states. The matrix \( N \), known as the fundamental matrix of an absorbing chain, can then be calculated as

\[
N = (I - P^*)^{-1}
\]  

(43)

where \( I \) is the identity matrix. This matrix gives the expected number of cycles patients would spend in each health state given their initial state. The matrix \( V \) containing the variances of the expected number of cycles can be calculated as

\[
V = N (2N^* - I) - N^2
\]  

(44)

where \( N^* \) is a copy of the \( N \) matrix with only the diagonal elements preserved and zeros elsewhere.

**Critical appraisal of multistate survival analysis**

Multistate survival analysis can be problematic since the method requires extensive data and is based on quite strong simplifying assumptions. However, as quality of life increases in importance in clinical trials and becomes more routinely collected, the problem of lack of data should be less of an issue. Furthermore, if multistate models are proposed at the design stage of a clinical trial, then collection of data can be planned so that it yields adequate and appropriate data.

The data requirements for multistate modelling are strict, with not only dates of entry, death and censoring needed but also ‘exact’ dates of transition between health states. Transition dates can be estimated and the accuracy is determined by the frequency of the quality-of-life assessments. Alternative methods that do not require exact transition dates are available but are not readily accessible to researchers because specialised computer software is required.

Defining health states for a multistate model may be problematic, despite the fact that progressive health states are not required. The definitions based on quality-of-life data are subjective and different definitions need to be considered as part of a sensitivity analysis. The investigator has to make decisions on which quality-of-life variable to use, the number of health states to be included and the cut-off values used to discriminate between health states. At one extreme, the model needs to be complex enough to be clinically meaningful and to ensure that information from the data is utilised to a maximum. At the other extreme, a simple model is needed to allow an adequate number of transitions between health states, both to enable transition rates to be estimated with sufficient precision and to ease interpretation of the analysis.

Definition of the health states may be such that clinically important information is lost. It may not be possible to include the most clinically important transition in the model because of the small number of participants in the study experiencing it. For example, in the MIC study the moderately ill to severely ill transition could not be modelled because of too few patients but this could be a very important transition from a clinical viewpoint. Also the transition between health states may result from very small changes in quality of life while large changes in quality of life are not reflected in the model. For example, if the MQS had been used in the MIC study to define health states and a value of 1 had been used to discriminate between the alive–well and alive–ill states, then a small change in MQS from 0.9 to 1.1 would result in a transition while a large change from, say, 1.1 to 3 would not be reflected in the model.

A multistate model, despite its complexity, may be preferable to an overall survival model since a greater biological insight may be gained by analysing the steps in a disease process. Different covariates may affect different transitions and the effects of important covariates may be lost when considering just overall survival. Examples of applications of multistate models to quality-of-life data are limited. Such applications, including the possibility of the inclusion of drop-out states in the model to overcome problems of informative
drop-out, require further investigation. Methods discussed in this chapter have been based on a classical approach and, although there has been some work on Bayesian approaches to multistate modelling, further work in this field is needed.

References


Chapter 18
Incorporating results from a multistate model into a Q-TWiST analysis

Introduction
The quality-adjusted survival analysis discussed previously (see chapter 16) is, in general, a non-parametric technique for comparing health technologies in terms of both quality of life and survival simultaneously. Cole and colleagues\(^1\) devised a parametric approach to quality-adjusted survival analysis that enabled many of the limitations of the Q-TWiST model to be overcome. Essentially the method uses the results from a multistate survival analysis to carry out a more refined Q-TWiST analysis. This chapter is based purely on this paper,\(^1\) which is a seminal one on this topic.

The method was illustrated by Cole and colleagues\(^1\) by data from a breast cancer clinical trial. The health states used in the analysis were the usual Q-TWiST states (i.e. TOX, TWiST and REL), which were determined using clinical criteria rather than quality-of-life data. However, quality-of-life data rather than clinical criteria could be used to define the health states in the model.

Methodology
As with both Q-TWiST and multistate survival analysis, the analysis starts by defining the health states that patients occupy during the study. Possible transitions between health states are also defined. A multistate survival analysis approach is then used to model the transition intensities. Each successive transition is represented by a cause-specific hazard function which is conditional upon previous transitions. The model is equivalent to a semi-Markov model with hazard functions being conditional on the current state, the transitions to the current state and the duration of the current state. Covariates can be included in the model. Cole and colleagues\(^1\) used conditional log-normal distributions and conditional Weibull distributions to model their hazard functions and included covariates using accelerated failure-time models. Parameter estimates are obtained by maximum likelihood.

Using the models fitted to the health state transitions, the expected amount of time spent in each health state can be estimated using simulation for specific sets of covariate values. Standard errors for these estimates can be obtained using the bootstrap method. Cole and colleagues\(^1\) estimated the average time spent in each health state for two sets of covariate values, one relating to a good prognosis and the other to a poor prognosis. The estimates were restricted to a 10-year follow-up period but, in general, the estimates can be unrestricted.

The next stage is to combine the average time spent in each health state with utility measures representing the quality of life experienced in each health state. Cole and colleagues\(^1\) referred to the function which combines quality and quantity of life into a composite measure as a quality function. The non-parametric Q-TWiST model is a quality function but more general functions can be defined to allow the utility coefficients for a health state to depend on the entry and exit times and the duration of time spent in the state. Quality functions can also be defined to incorporate discounting. Cole and colleagues\(^1\) used the standard non-parametric Q-TWiST quality function in their analysis and compared treatments in terms of quality-adjusted survival using various arbitrary utility values. The significance of the difference between treatments can be established from the CIs for the difference in quality-adjusted survival.

Conclusions
The parametric approach to quality-adjusted survival analysis overcomes many of the limitations associated with a non-parametric Q-TWiST analysis. It does not restrict estimates to the follow-up period of the data and allows covariate values to be easily included. In theory it does not require progressive health states, although there may be strong assumptions in relation to repeat transitions. The method also allows quality of life and survival to be combined in a more general way than the Q-TWiST model with, for example, utility coefficients being functions
of time rather than just constants and discount rates being included where appropriate.

The method requires the number of observations and the number of states to be such that there are sufficient data for estimating each of the conditional cause-specific hazard functions. If estimation is based on few data, the variability of the estimates will be large giving perhaps excessive variability in the quality-adjusted survival.

In conclusion, when a particular parametric model is appropriate and a reasonably large amount of data is available, this method will give more efficient estimation than a non-parametric approach.

Reference

Chapter 19

Modelling quality of life and survival as two simultaneous processes

In studies where quality of life and survival are both important endpoints, the change in quality of life over time and the time to death can be considered as two simultaneous processes occurring in patients, and can be modelled as such. Models for analysing longitudinal quality-of-life data and survival as two separate processes were discussed earlier in this review (see chapters 5–9 and chapters 10–13, respectively). Simultaneous modelling allows survival data to be incorporated into the model for quality-of-life data, thus adjusting for informative missing data caused by drop-out due to death. Conversely, the analysis of survival data with quality of life as a time-dependent covariate is enhanced in simultaneous modelling, since covariate values are estimated from the model for quality-of-life data over time fitted to all subjects.

The simultaneous modelling of repeatedly measured covariates and survival data has been described for covariate data in the form of normally distributed clinical markers, but not for quality-of-life data, which may be non-normally distributed continuous measures or even ordinal measures. Models are set up for each process and Gibbs sampling is used to simultaneously fit the models in a single analysis, 1,2 Faucett and Thomas 1 refer to the two parts of the model as the ‘covariate tracking model’ for the repeated measures process and the ‘disease risk model’ for the survival process. In their simultaneous modelling they assume a random effects model for the covariate tracking part (see equations 45 and 46 below) and a proportional hazards model for the disease risk part (see equation 47).

Thus, \( z_{ij} \), the single continuous time-dependent covariate for the \( j \)th measurement of subject \( i \) at time \( t_{ij} \) is modelled using:

\[
z_{ij} = x_i(t_{ij}) + \varepsilon_{ij} \tag{45}
\]

where \( x_i(t_{ij}) \) is the value of the true unobserved covariate at time \( t_{ij} \) and \( \varepsilon_{ij} \) are independent, normally distributed errors, and

\[
x_i(t) = \alpha_i + \beta_i t \tag{46}
\]

where the random effects \( \alpha_i \) and \( \beta_i \) have a bivariate normal distribution.

At the same time, the hazard of death for subject \( i \) is modelled using

\[
\lambda_i(t) = \lambda_0(t) \exp (\gamma x_i(t)) \tag{47}
\]

where \( \lambda_0(t) \) represents the underlying baseline hazard function and \( \gamma \) is the regression coefficient for estimation.

The method is flexible in that it allows for unequally spaced and missing repeated measures data, with varying numbers of observations per subject, and it allows for censored survival times. Simulation was used to compare results from modelling each process separately with results from the combined model and showed that the separate models underestimated parameters while the combined model virtually eliminated the bias. 1

Instead of modelling the survival process simultaneously with a longitudinal measurements process, a more general approach would be to model the drop-out process, which includes other reasons for drop-out apart from death. Lindsey 3 mentions that any appropriate standard longitudinal repeated measures model can be used to analyse, conditionally, all observed responses up to drop-out, and his paper is focused instead on modelling the drop-out process. A survival model in a log-linear form is used to model the drop-out process, with subjects not dropping out being uninformatively censored. The longitudinal measurements can be included as covariates in the drop-out model. If the study is such that drop-outs occur for different reasons, modelling them as distinct risk processes should be considered. Lindsey 3 considers that the drop-out process should be modelled together with the longitudinal measurements process, since it is an integral part of the phenomenon under study.

References


In studies in which quality of life and survival are both important endpoints, it may be advantageous to assess health technologies in terms of these endpoints simultaneously. Three different approaches can be used to achieve this. One approach is to combine quality and quantity of life into a single endpoint and use quality-adjusted survival analysis to compare treatments. Another approach is to use multistate models to describe the movement of patients between various health states, defined by levels of quality of life and death, and to explore how treatments differ in terms of these movements. Finally, a simultaneous modelling approach considers quality of life and survival as two simultaneous processes and describes the data in terms of two inter-linked models.

There is some controversy regarding the simultaneous analysis of quality-of-life and survival data. One view held is that the time-dependent structure of the individual quality-of-life process can best be accounted for when quality and quantity of survival are analysed simultaneously, while another is that methods which attempt to combine quality of life and survival into a single measure are generally inappropriate. It may be preferable to analyse and report quality-of-life and survival outcomes separately so that any conflict between them in terms of treatment differences is apparent.

If simultaneous analysis is deemed to be appropriate, the choice of method depends on the aims of the study, the nature of the disease and treatments, and the quality-of-life data collected. The strengths and limitations of each method should be considered carefully (see Table 29). Quality-adjusted survival analysis is the most straightforward approach but may be difficult to apply because of the need for progressive health states. Multistate modelling is more flexible but more complex and has stringent data requirements. In

### TABLE 29 Comparison of quality-adjusted survival analysis and multistate survival analysis

<table>
<thead>
<tr>
<th>Quality-adjusted survival analysis</th>
<th>Multistate survival analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aim</strong></td>
<td></td>
</tr>
<tr>
<td>• Compares treatments in terms of a composite measure of quality and quantity of life</td>
<td>• Compares treatments in terms of movements between different health states</td>
</tr>
<tr>
<td><strong>Limitations</strong></td>
<td></td>
</tr>
<tr>
<td>• Need to be able to define a finite, mutually exclusive and exhaustive set of health states that are clinically meaningful and fully describe the experiences of patients</td>
<td>• Need to be able to define a finite, mutually exclusive and exhaustive set of health states that are clinically meaningful and fully describe the experiences of patients</td>
</tr>
<tr>
<td>• For partitioned survival analysis, the health states must be progressive</td>
<td>• QoL assessments need to be frequent enough to be able to estimate exact transition times adequately</td>
</tr>
<tr>
<td>• QoL assessments need to be frequent enough to be able to estimate transition times adequately</td>
<td>• The assumptions underlying the QALY model may not be valid</td>
</tr>
<tr>
<td>• The assumptions underlying the QALY model may not be valid</td>
<td>• For a subject-based approach, a cut-off time for the analysis needs to be chosen to minimise the number of censored survival times, otherwise a population-based approach should be used</td>
</tr>
<tr>
<td>• For a subject-based approach, a cut-off time for the analysis needs to be chosen to minimise the number of censored survival times, otherwise a population-based approach should be used</td>
<td>• Partitioned survival analysis requires the period of analysis to be restricted to an upper time limit</td>
</tr>
<tr>
<td>• Calculated of CIs in a partitioned survival analysis requires specialised methods</td>
<td>• Calculation of CIs in a partitioned survival analysis requires specialised methods</td>
</tr>
<tr>
<td><strong>Strengths</strong></td>
<td></td>
</tr>
<tr>
<td>• Enables covariates to be included</td>
<td>• Enables different covariates to be considered for different transitions</td>
</tr>
</tbody>
</table>
both cases, defining health states can be problematic and different definitions should be considered as part of a sensitivity analysis. The accuracy of the results from both methods depends on the frequency of the quality-of-life assessments. Simultaneous modelling could potentially provide the most flexible and powerful approach but further research of the method is required before it can be fully evaluated.

Methods that simultaneously analyse quality-of-life and survival data are based on a single measure of quality of life. This may cause problems in situations with several distinct quality-of-life measures, when the analysis would need to be repeated for each, and large numbers of quality-of-life endpoints could lead to problems of interpretation and multiple testing. All methods, however, not only have the advantage of allowing for informative drop-out due to death but also have the potential to be extended to deal with informative drop-out for reasons other than death. Further research into the application of these methods to quality-of-life data is required.

References
Overall summary and discussion

In health technology assessment with both quality of life and survival as endpoints, the appropriate method of analysis depends on which endpoint is most important. In some situations, the primary focus of the analysis will be quality of life, with consideration also given to survival but only in terms of the problematical impact the latter may have on the quality-of-life data. In other situations, the focus will be primarily on survival with the analysis adjusting for the effects of quality of life and, in further scenarios, both endpoints will be important.

Quality-of-life data are usually longitudinal in nature and, as with most longitudinal studies, the data suffer from the problem of missing values, especially as a result of patient drop-out. The missing values do not only cause the data to be unbalanced but also cause problems for the analysis because the missing data mechanism is likely to be non-ignorable. This is because, in quality-of-life studies in which survival is also an issue, patients are generally severely ill and drop-out caused by illness and death is a common occurrence. This drop-out process may be informative and needs to be accounted for in any analysis of the data.

Another problem with quality-of-life data is in the multivariate nature of the endpoint. Quality-of-life data may comprise a single global measure but often consist of a set of item or dimension scores. The analysis may require each quality-of-life endpoint to be considered separately but, if the number of endpoints is large, multiple testing and estimation may be a problem. Methods that handle the multivariate nature of the quality-of-life endpoint have been proposed but they either suffer from the disadvantage of providing results purely in terms of hypothesis testing and do not give an estimate of treatment effect, or they do not deal with the problem of informative drop-out. The analysis should either focus on a few key quality-of-life measures or should attempt to combine measures into an aggregate score on which the analysis can then be based.

Any analysis of longitudinal quality-of-life data should begin descriptively, using plots of individual patient profiles and group profiles, to give insight into the data before any formal testing or modelling is carried out. Group profiles should be interpreted with caution since, if informative drop-out is present, group summary measures may well be biased. A descriptive analysis can be used to explore the drop-out problem in the data, although establishing if drop-out is informative is not possible.

The application of standard methods of longitudinal data analysis to quality-of-life data is generally problematic because of the likely presence of informative drop-out. The simple approach of using summary measures does not fully capture the dynamic nature of the data and may be problematic in the presence of informative drop-out; although the more complex modelling techniques model the change in quality of life over time, they assume the missing data mechanism is ignorable. These methods could therefore give biased results and invalid conclusions when informative drop-out is present in the data. Modelling techniques that deal with informative drop-out have been developed and their application to quality-of-life data needs investigation.

If the treatment comparison is in terms of survival, then the analysis can adjust for the effects of quality of life by including it as a covariate in a standard survival model. This also enables the prognostic value of quality of life to be explored. Quality of life can be included either in terms of baseline values as a fixed covariate or in terms of changing values over time as a time-dependent covariate. If assessments of quality of life are infrequent or data are missing for reasons other than death, then it may be difficult to adjust for changing quality of life with any degree of accuracy. The analysis may be improved by modelling quality of life and survival as two simultaneous processes. In this approach, values of quality of life incorporated into a survival analysis as a covariate are estimated from the model for quality-of-life data over time fitted to all subjects.

In studies where quality of life and survival are both important endpoints, it may be
Conclusions and recommendations

advantageous to assess health technologies in terms of these endpoints simultaneously. Three different approaches can be used to achieve this. One approach is to combine quality and quantity of life into a single endpoint and use quality-adjusted survival analysis to compare treatments; another is to use multistate models to describe the movement of patients between various health states, defined by levels of quality of life and death, and to explore how treatments differ in terms of these movements (see Table 29 for a comparison of these two approaches). Finally, in a simultaneous modelling approach, quality of life and survival are considered as two simultaneous processes and the data are described in terms of two interlinked models.

The method of quality-adjusted survival analysis is technically straightforward, except for the calculation of CIs, but may be difficult to apply because it requires survival time to be partitioned into progressive health states. These may not always be easy to define in a clinically relevant way. A further limitation is the need to restrict the period of analysis to an upper time limit, although this can be overcome using a parametric approach. Quality-adjusted survival analysis deals with the problem of informative drop-out due to death and has the potential to be extended to deal with drop-out for reasons other than death.

Multistate survival analysis does not necessarily require progressive health states and provides a more flexible means for modelling quality-of-life and survival data. The inclusion of death as a health state in the model enables the analysis to deal with informative drop-out due to death, and it may be possible to deal with drop-out for other reasons by including a ‘drop-out’ health state. Depending on the multistate model, methods are reasonably accessible to researchers, except in those situations where exact transition times are not known. In general, multistate survival analysis can be problematic since the method requires extensive data for adequate modelling of the data.

Modelling quality of life and survival as two simultaneous processes has already been suggested as a method that may enhance the analysis of survival data with quality of life as a time-dependent covariate. Further, this approach allows both the interrelationship between the two processes to be assessed and, in particular, for the changes in quality of life over time to be explored while adjusting for informative drop-out due to death. Other reasons for drop-out can be included by modelling the drop-out rather than the survival process explicitly. The benefits of simultaneous modelling in a quality-of-life context need further investigation.

Recommendations

Implications for the design and conduct of HTA research

This review of the methods proposed in the scientific and medical literature for the analysis of quality-of-life and survival data, together with their application to data from a previously conducted study, has given rise to a series of recommendations for practitioners and researchers. These recommendations should complement those given by other groups reviewing quality-of-life assessment in clinical trials.

Obtaining appropriate data

Consideration of the proposed analysis of quality-of-life and survival data at the design stage of a study could result in better quality and more appropriate data for analysis. The method of analysis may influence decisions regarding the frequency and timing of quality-of-life assessments and also the instrument used to measure quality of life. Some methods, such as those using exact dates of health state transitions, may require frequent collection of data to ensure accurate assessment of changes in quality of life. The choice of fixed or varying assessment times may be partly influenced by the proposed analytical method.

In choosing the instrument with which to measure quality of life, consideration should be given to the number of quality-of-life endpoints that will be measured. Some instruments result in a single global measure of quality of life, while others produce a number of dimension-specific scores. In some cases, large numbers of quality-of-life measures will be produced. Most methods of analysis are based on a single measure of quality of life and the analysis will need to be repeated for major quality-of-life endpoints. If a large number of quality-of-life measures are produced then both the analysis and the interpretation may become problematic. Use of a single global measure of quality of life simplifies statistical analysis and should be aimed at when sensible and justifiable. Otherwise, to overcome problems of multiple testing, the most important quality-of-life endpoints on which hypothesis testing will be based should be specified at the design stage, with all other outcomes assumed to be of secondary importance.
Missing data cause the most serious problems for the analysis of quality-of-life data and thus the study design, in terms of quality-of-life assessment, should be chosen to minimize the chance of non-compliance. Patient drop-out is still likely to occur and recording reasons for drop-out during data collection will maximize insight into the drop-out process, thus enabling modelling of separate drop-out processes when necessary. The number of patients required for a study will also depend on the method of analysis. Multistate survival analysis, for example, needs a relatively large overall sample size, so that there are sufficient patients in each transition for adequate modelling of the data.

- The method of analysis needs to be decided upon at the design stage of a study so that appropriate quality-of-life data can be collected. Issues to consider are:
  - the quality-of-life instrument to be used
  - the frequency and timing of quality-of-life assessments
  - the need to minimize non-compliance
  - the collection of additional information, such as reason for drop-out
  - the sample size required.

Choosing the appropriate method

There are several factors affecting the choice of method for analysing quality-of-life and survival data. The research question that the study aims to answer is a major factor and, if the choice is being made after data collection rather than at the design stage, the nature of the quality-of-life data that has been collected will also have an effect. The analysis needs to produce results that are relevant and accessible to health service professionals and patients. Some methods of analysis are complex and interpretation of results may be difficult.

In broad terms, it may be argued that, in view of the problems of data definition and collection, too much sophistication in analysis would be misguided and the methods used for analysis should broadly match the quality of the data.4 In general, when choosing the appropriate method of analysis the advantages and disadvantages of each method should be weighed against each other, with consideration given to the quality of the data.

- The choice of method should be based on the research question that the study aims to answer. The advantages and disadvantages of each method should be considered carefully, together with the relevance and interpretability of the results to clinicians and patients.

Analysis of longitudinal quality-of-life data must consider the problem of informative drop-out. Standard methods for longitudinal data analysis assume missing data mechanisms are ignorable and should be avoided because they may produce biased results and invalid conclusions. Methods that deal with informative drop-out should be used. The appropriate method will depend on whether drop-out is primarily caused by death or occurs for multiple reasons.

- Methods used to analyse longitudinal quality-of-life data must allow for informative drop-out.

Reporting the analysis

The method used for analysis of quality-of-life and survival data should be described clearly when reporting the results. Some justification for the choice could also be included. To analyze the data, various assumptions usually need to be made, whatever method is used. Assumptions such as these should be reported clearly. In the MIC study, for example, most methods of analysis were based on the major assumption that changes in quality of life occurred midway between assessments rather than on the actual assessment date. Sensitivity analysis should be carried out to assess the effect of the assumptions on the conclusions of the analysis.

- Methods used should be reported clearly, with details of definitions and assumptions used in the analysis.
- Sensitivity analysis should be carried out to assess the robustness of conclusions to any critical assumptions made in the analysis.

Recommendations for further research

Some of the methods described in this report have, to date, been applied to only a few examples of quality-of-life and survival data. Although used fairly extensively, the application of quality-adjusted survival analysis has been based mainly on clinical data rather than on patient-assessed quality-of-life data; hence, further work evaluating the use of such a technique in a quality-of-life context needs to be addressed. There has been limited use of other techniques, such as hierarchical modelling, multistate modelling and simultaneous modelling, in the quality-of-life field and further research is needed. Most methods are based on a classical approach and the application of Bayesian methods requires further investigation. Although the available computer software has not been reviewed in this project, it has become apparent through worked examples that development of software may be required; a review of this would form a useful supplement to this review.
The number of applications of the methodology discussed in this review to quality-of-life data should continue to expand; the methodological areas most likely to develop rapidly are multistate survival analysis and simultaneous modelling. The problems of analysing longitudinal data with informative drop-outs are also being addressed in other parallel areas, such as AIDS research; developments within this context should also be monitored.

- Further experience in the application of quality-adjusted survival analysis techniques to quality-of-life data is needed to enable a proper evaluation of such methods.
- Further research is needed to develop hierarchical models, multistate models and simultaneous modelling methods in their practical application to quality-of-life and survival data using both classical and Bayesian approaches. Consideration should be given as to how such methods could deal with the multivariate nature of the quality-of-life endpoint.
- A full review of available computer software for methods that simultaneously analyse quality-of-life and survival data is needed to highlight areas requiring further development.
- Progress in the most rapidly developing areas, multistate survival analysis and simultaneous modelling, should be monitored, together with parallel areas of methodological development, such as in AIDS research.

References


Acknowledgements

We would like to thank the HTA programme for their continuing financial support for this project.

We would like to thank the Cancer Research Campaign Trials Unit at Birmingham University for allowing Cindy Billingham to be seconded to Leicester University for the duration of the project. Particular thanks go to Janet Dunn who kindly instigated and negotiated the arrangement.

We are indebted to Dr Michael Cullen, Consultant and Honorary Reader in Medical Oncology at the Queen Elizabeth Hospital in Birmingham, for allowing us to use the MIC data as the illustrative example in this report. We also gratefully acknowledge the assistance of Sue Spriggs from the Clinical Sciences Library at Leicester University and Andrew Booth from the Sheffield Centre for Health and Related Research (SCHARR) for their advice on searching the literature.

We would also like to thank all our colleagues at Leicester and Birmingham for their invaluable support and encouragement.

Our thanks are also due to the referees for their perseverance in reading the report and the quality of their comments.
All 1127 references found by the search strategy discussed in chapter 2 are listed here, including those which are not referenced in the report.

The following publications were found too late for inclusion in the study but are likely to be of particular interest.


Report bibliography


Beacon HJ, 1996. The statistical analysis of self assessed quality of life data in cancer clinical trials [PhD thesis]. London School of Hygiene and Tropical Medicine, University of London.


Campbell M, Gibbard N. Frequency of use of quality-of-life instruments. To be published.


Bibliography


Danese MD, Powe NR, Sawin CT, Ladenson PW, 1996. Screening for mild thyroid failure at the periodic health examination: a decision and cost-effectiveness analysis. JAMA;276:859–64.


Murray JR, 1995. Do you CEA the need for CMAs or have a CUA about extra QALY ratios? J Dermatol Treat;6:3–4.


Weinstein MC, 1988. A QALY is a QALY is a QALY – or is it? *J Health Econ*;7:289–90.


Appendix 1

Details of search terms used in electronic database searching

Notes on searching in BIDS
• Putting an asterisk at the end of a search term allows any ending to the word to be searched for on the database.
• Search terms consisting of hyphenated words need to be enclosed in double quotes otherwise the hyphen will be interpreted as an exclusion sign.
• Using a comma between two search terms allows the database to be searched for articles containing either term.

Selecting type of journal
The following criteria were used in BIDS to restrict the search to just the statistical, biometry, clinical trial and epidemiology journals.

journal = statistic*, biometri*, clinical trial*, epidemiol*

Quality-of-life search terms
The search terms used in BIDS to select articles containing some information of interest on quality of life are shown in Table 30.

Survival search terms
The search terms used in BIDS to select articles with some interest in survival are shown in Table 31.

TABLE 30 Search terms: quality of life

<table>
<thead>
<tr>
<th>Search terms</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>quality of life, “quality-of-life”</td>
<td>Also captures more specific phrases such as health-related quality-of-life.</td>
</tr>
<tr>
<td>life quality, “life-quality”</td>
<td></td>
</tr>
<tr>
<td>quality life</td>
<td>Also captures articles where a non-standard hyphenated version has been used, e.g. quality-of-life, quality-of-life, quality of-life.</td>
</tr>
<tr>
<td>“well-being”, wellbeing, well being</td>
<td></td>
</tr>
</tbody>
</table>

TABLE 31 Search terms: survival

<table>
<thead>
<tr>
<th>Search terms</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>survival</td>
<td>Captures a whole range of survival-type phrases, e.g. survival time, survival analysis, survival date, survival endpoints, survival rate, survival curve, survival probability, disease-free survival, event-free survival.</td>
</tr>
<tr>
<td>length of life</td>
<td></td>
</tr>
<tr>
<td>relapse</td>
<td>Also captures more specific terms such as time to relapse.</td>
</tr>
<tr>
<td>failure time*, “failure-time”*</td>
<td>* captures both time and times and also hyphenated phrases such as failure-time-data.</td>
</tr>
<tr>
<td>time to death, time to recurrence, time to progression, time to disease, time to failure</td>
<td>Also captures more specific phrases such as time to disease progression, time to disease recurrence.</td>
</tr>
<tr>
<td>life table*, “life-table”*</td>
<td>* captures both table and tables and also hyphenated phrases such as life-table-analysis.</td>
</tr>
<tr>
<td>Kaplan Meier, “Kaplan-Meier”</td>
<td></td>
</tr>
<tr>
<td>log rank, “log-rank”</td>
<td></td>
</tr>
<tr>
<td>proportional hazard*, “proportional-hazard”*</td>
<td>* captures both hazard and hazards and also hyphenated phrases such as proportional-hazards-analysis.</td>
</tr>
<tr>
<td>Cox* + (model*, Cox* allows for various versions such as Cox, Coxs or Cox's. Cox* can not be used as a search term on its own because it lacks precision. The specification captures phrases such as Cox multivariate regression or multiple regression model of Cox.</td>
<td></td>
</tr>
<tr>
<td>event histor*, “event-histor”*</td>
<td>* captures both both history and histories and also hyphenated phrases such as event-history-analysis.</td>
</tr>
<tr>
<td>censor*</td>
<td>* captures censor, censored and censoring and thus captures more specific phrases such as informative censoring.</td>
</tr>
</tbody>
</table>
Methodology search terms
The search terms used in BIDS to select articles which used a known methodology for the simultaneous analysis of quality-of-life and survival data are shown in Table 32.

<table>
<thead>
<tr>
<th>Search terms</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>quality adjusted*, “quality-adjus”</td>
<td>captures hyphenated phrases such as...-life, ...-years, ...-years. Captures other phrases such as... life, ... life years, ... years of life, ... life expectancy, ... loss in life expectancy, ... survival, ... days, ... time, ... tooth years.</td>
</tr>
<tr>
<td>quality survival time, “quality-survival time”, “quality survival-time”, “quality-survival-time”</td>
<td></td>
</tr>
<tr>
<td>healthy years equivalent*, “healthy-years equivalent***”, “healthy year-equivalent***”, “healthy-years-equivalent***”</td>
<td>captures equivalent and equivalence.</td>
</tr>
<tr>
<td>healthy year equivalent*, “healthy-year equivalent***”, “healthy year-equivalent***”, “healthy-year-equivalent***”</td>
<td>captures equivalent and equivalence.</td>
</tr>
<tr>
<td>(“multi-state”, multistate) + (censor*)</td>
<td>Captures articles using multistate survival analysis.</td>
</tr>
</tbody>
</table>

Longitudinal search terms
The search terms used in BIDS to select articles describing longitudinal studies (which were more relevant to this review) are shown in Table 33.

<table>
<thead>
<tr>
<th>Search terms</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>longitudinal</td>
<td></td>
</tr>
<tr>
<td>repeated measure*, “repeated-measure***”, repeated assessment*, “repeated-assessment***”</td>
<td>Measure* captures measure, measures, measurement or measurements.</td>
</tr>
<tr>
<td>measure* over time, assess* over time</td>
<td>Measure* captures measure, measures, measurement or measurements. Assess* captures assessment or assessments.</td>
</tr>
<tr>
<td>profile, profiles</td>
<td>Profile* could not be used because it had a lack of precision.</td>
</tr>
<tr>
<td>dynamic</td>
<td></td>
</tr>
</tbody>
</table>
# Health Technology Assessment panel membership

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

## Acute Sector Panel

<table>
<thead>
<tr>
<th><strong>Current members</strong></th>
<th><strong>Past members</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chair:</strong> Professor Francis H Creed, University of Manchester</td>
<td>Professor Clifford Bailey, University of Leeds</td>
</tr>
<tr>
<td>Professor Senga Bond, University of Newcastle-upon-Tyne</td>
<td>Ms Tracy Bury, Chartered Society of Physiotherapy</td>
</tr>
<tr>
<td>Professor Ian Cameron, Southeast Thames Regional Health Authority</td>
<td>Professor Collette Clifford, University of Birmingham</td>
</tr>
<tr>
<td>Ms Lynne Clemence, Mid-Kent Health Care Trust</td>
<td>Professor Michael Sheppard, Queen Elizabeth Hospital, Birmingham</td>
</tr>
</tbody>
</table>

## Diagnostics and Imaging Panel

<table>
<thead>
<tr>
<th><strong>Current members</strong></th>
<th><strong>Past members</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chair:</strong> Professor Mike Smith, University of Leeds</td>
<td>Professor John Farndon, University of Bristol</td>
</tr>
<tr>
<td>Dr Philip J Ayres, Leeds Teaching Hospitals NHS Trust</td>
<td>Professor Senga Bond, University of Newcastle-upon-Tyne</td>
</tr>
<tr>
<td>Dr Paul Collinson, Maiday University Hospital, Thornton Heath</td>
<td>Professor Ian Cameron, Southeast Thames Regional Health Authority</td>
</tr>
</tbody>
</table>

*Previous Chair continued*
### Health Technology Assessment panel membership

#### continued

<table>
<thead>
<tr>
<th>Methodology Panel</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Current members</strong></td>
</tr>
</tbody>
</table>
| **Chair:**  
Professor Martin Buxton,  
Brunel University |
| Professor Ann Bowling,  
University College London Medical School |
| Professor Jeremy Grimshaw,  
University of Aberdeen |
| Dr Nick Payne,  
University of Sheffield |
| Dr Stephen Harrison,  
University of Leeds |
| Professor Margaret Pearson,  
NHS Executive North West |
| Professor Doug Altman,  
Institute of Health Sciences, Oxford |
| Professor Michael Drummond,  
University of York |
| Professor Richard Lilford,  
Regional Director, R&D,  
West Midlands |
| Professor David Sackett,  
Centre for Evidence Based Medicine, Oxford |
| Dr Mike Clarke,  
University of Oxford |
| Mr John Henderson,  
Department of Health |
| Dr PAG Sanderson,  
University of Edinburgh |
| Professor Michael Baum,  
Royal Marsden Hospital |
| Professor Marta Kuklinski,  
University of Oxford |
| Professor Teresa Marteau,  
Guy’s, King’s & St Thomas’s School of Medicine & Dentistry, London |
| Mr Philip Hewitson,  
Leeds FHSA |
| Professor Charles Warlow,  
Western General Hospital, Edinburgh |
| Dr Vikki Entwistle,  
University of Aberdeen |
| Professor Ewan Ferlie,  
Imperial College, London |
| Dr Henry McQuay,  
University of Oxford |
| Professor Nick Black,  
London School of Hygiene & Tropical Medicine |
| Professor Ray Fitzpatrick,  
University of Oxford |
| Professor Joy Townsend,  
University of Hertfordshire |
| Past members |
| Professor Anthony Culyer,  
University of York |
| Professor George Davey-Smith,  
University of Bristol |
| Mr Nick Mays,  
King’s Fund, London |
| Professor Charles Warlow,  
Western General Hospital, Edinburgh |
| Dr Doug Altman,  
Department of Health Sciences, Oxford |
| Professor Stephen Frankel,  
University of Bristol |
| Professor Ian Russell,  
University of York |
| Dr Maurice Slevin,  
St Bartholomew’s Hospital, London |
| Dr David Armstrong,  
Guy’s, King’s & St Thomas’s School of Medicine & Dentistry, London |
| Dr Vikki Entwistle,  
University of Aberdeen |
| Professor Ewan Ferlie,  
Imperial College, London |
| Dr Henry McQuay,  
University of Oxford |
| Professor Nick Black,  
London School of Hygiene & Tropical Medicine |
| Professor Ray Fitzpatrick,  
University of Oxford |
| Professor Joy Townsend,  
University of Hertfordshire |

<table>
<thead>
<tr>
<th>Pharmaceutical Panel</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Current members</strong></td>
</tr>
</tbody>
</table>
| **Chair:**  
Professor Tom Walley,  
University of Liverpool |
| Professor Rod Griffiths,  
West Midlands |
| Dr Andrew Mortimore,  
Southampton & SW Hants Health Authority |
| Dr Frances Rotblat,  
Medicines Control Agency |
| Dr Felicity Gabbay,  
Transcrip Ltd |
| Mrs Jeanette Howe,  
Department of Health |
| Mr Nigel Offen,  
Essex Rivers Healthcare, Colchester |
| Dr Eamonn Sheridan,  
St James’s University Hospital, Leeds |
| Mr Peter Golightly,  
Leicester Royal Infirmary |
| Professor Trevor Jones,  
ABPI, London |
| Mrs Marianne Rigge,  
The College of Health, London |
| Mrs Katrina Simister,  
Liverpool Health Authority |
| Dr Alastair Gray,  
Health Economics Research Unit, University of Oxford |
| Ms Sally Knight,  
Lister Hospital, Stevenage |
| Mr Simon Robbins,  
Camden & Islington Health Authority, London |
| Dr Ross Taylor,  
University of Aberdeen |
| Past members |
| Professor Michael Rawlins,  
University of Newcastle-upon-Tyne |
| Dr Tim Elliott,  
Department of Health |
| Dr John Posnett,  
University of York |
| Dr Colin Bradley,  
University of Birmingham |
| Mrs Julie Dent,  
Ealing, Hammersmith &  
Hounslow Health Authority, London |
| Dr Desmond Fitzgerald,  
Mere, Bucklow Hill, Cheshire |
| Dr Tim van Zwanenberg,  
Northern Regional Health Authority |
| Professor Alasdair Breckenridge, RDRD,  
Northwest Regional Health Authority |
| Mr Barrie Dowdeswell,  
Royal Victoria Infirmary, Newcastle-upon-Tyne |
| Professor Keith Gull,  
University of Manchester |
| Dr Kent Woods,  
RDRD, Trent RO, Sheffield |
| Ms Christine Clark,  
Hospice, Salford |
| Dr Keith Jones,  
Medicines Control Agency |
| *Previous Chair* |
### Population Screening Panel

**Current members**

<table>
<thead>
<tr>
<th>Chair:</th>
<th>Professor Sir John Grimley Evans, Radcliffe Infirmary, Oxford</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ms Stella Burnsdie, Altnagelvin Hospitals Trust, Londonderry</td>
</tr>
<tr>
<td></td>
<td>Mr John Cairns, University of Aberdeen</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dr Sheila Adam, Department of Health*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professor George Freeman, Charing Cross &amp; Westminster Medical School, London</td>
</tr>
<tr>
<td>Dr Mike Gill, Brent &amp; Harrow Health Authority</td>
</tr>
<tr>
<td>Dr Anne Ludbrook, University of Aberdeen</td>
</tr>
<tr>
<td>Professor Theresa Marteau, Guy's, King's &amp; St Thomas's School of Medicine &amp; Dentistry, London</td>
</tr>
<tr>
<td>Professor Catherine Peckham, Institute of Child Health, London</td>
</tr>
<tr>
<td>Dr Connie Smith, Parkside NHS Trust, London</td>
</tr>
<tr>
<td>Ms Polly Toynbee, Journalist</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Past members</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr Sheila Adam, Department of Health*</td>
</tr>
<tr>
<td>Professor George Freeman, Charing Cross &amp; Westminster Medical School, London</td>
</tr>
<tr>
<td>Dr Mike Gill, Brent &amp; Harrow Health Authority</td>
</tr>
</tbody>
</table>

| Dr Anne Ludbrook, University of Aberdeen |
| Professor Theresa Marteau, Guy's, King's & St Thomas's School of Medicine & Dentistry, London |
| Dr Connie Smith, Parkside NHS Trust, London |
| Ms Polly Toynbee, Journalist |

---

### Primary and Community Care Panel

**Current members**

<table>
<thead>
<tr>
<th>Chair:</th>
<th>Dr John Tripp, Royal Devon &amp; Exeter Healthcare NHS Trust</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mr Kevin Barton, East London &amp; City Health Authority</td>
</tr>
<tr>
<td></td>
<td>Professor John Bond, University of Newcastle-upon-Tyne</td>
</tr>
<tr>
<td></td>
<td>Dr John Brazier, University of Sheffield</td>
</tr>
</tbody>
</table>

| Ms Judith Brodie, Age Concern, London |
| Mr Shaun Brogan, Daventry & South Northants Primary Care Alliance |
| Mr Joe Corkill, National Association for Patient Participation |
| Dr Nicky Cullum, University of York |
| Professor Pam Enderby, University of Sheffield |
| Mr Andrew Farmer, Institute of Health Sciences, Oxford |
| Professor Richard Hobbs, University of Birmingham |
| Professor Allen Hutchinson, University of Sheffield |
| Dr Phillip Leech, Department of Health |
| Dr Aidan Macfarlane, Oxfordshire Health Authority |
| 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 |
| Ms Hilary Scott, Tower Hamlets Healthcare NHS Trust, London |
| 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 |
| 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 |

* Previous Chair continued
## National Coordinating Centre for Health Technology Assessment, Advisory Group

### Current members

| Chair: Professor John Gabbay, Wessex Institute for Health Research & Development |
| Ms Lynn Kerridge, Wessex Institute for Health Research & Development |
| 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 |

### Past member

Dr Paul Roderick, Wessex Institute for Health Research & Development
## Current members

<table>
<thead>
<tr>
<th>Chair:</th>
<th>Professor Charles Florey, Department of Epidemiology &amp; Public Health, Ninewells Hospital &amp; Medical School, University of Dundee</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Professor Doug Altman, Director of ICRF/NHS Centre for Statistics in Medicine, Oxford</td>
</tr>
<tr>
<td></td>
<td>Professor John Bond, Professor of Health Services Research, University of Newcastle-upon-Tyne</td>
</tr>
<tr>
<td>Mr Peter Bower, Independent Health Advisor, Newcastle-upon-Tyne</td>
<td></td>
</tr>
<tr>
<td>Ms Christine Clark, Honorary Research Pharmacist, Hope Hospital, Salford</td>
<td></td>
</tr>
<tr>
<td>Professor Shah Ebrahim, Professor of Epidemiology of Ageing, University of Bristol</td>
<td></td>
</tr>
<tr>
<td>Professor Martin Eccles, Professor of Clinical Effectiveness, University of Newcastle-upon-Tyne</td>
<td></td>
</tr>
<tr>
<td>Dr Mike Gill, Director of Public Health &amp; Health Policy, Brent &amp; Harrow Health Authority</td>
<td></td>
</tr>
<tr>
<td>Dr Alastair Gray, Director, Health Economics Research Centre, University of Oxford</td>
<td></td>
</tr>
<tr>
<td>Professor Mark Haggard, MRC Institute of Hearing Research</td>
<td></td>
</tr>
<tr>
<td>Dr Jenny Hewison, Senior Lecturer, Department of Psychology, University of Leeds</td>
<td></td>
</tr>
<tr>
<td>Professor Sir Miles Irving (Programme Director), Professor of Surgery, University of Manchester, Hope Hospital, Salford</td>
<td></td>
</tr>
<tr>
<td>Professor Alison Kitson, Director, Royal College of Nursing Institute</td>
<td></td>
</tr>
<tr>
<td>Dr Donna Lamping, Senior Lecturer, Department of Public Health, London School of Hygiene &amp; Tropical Medicine</td>
<td></td>
</tr>
<tr>
<td>Professor Alan Maynard, Professor of Economics, University of York</td>
<td></td>
</tr>
<tr>
<td>Professor Jon Nicholl, Director, Medical Care Research Unit, University of Sheffield</td>
<td></td>
</tr>
<tr>
<td>Professor Gillian Parker, Nuffield Professor of Community Care, University of Leicester</td>
<td></td>
</tr>
<tr>
<td>Dr Tim Peters, Reader in Medical Statistics, Department of Social Medicine, University of Bristol</td>
<td></td>
</tr>
<tr>
<td>Professor Martin Severs, Professor in Elderly Health Care, Portsmouth University</td>
<td></td>
</tr>
<tr>
<td>Dr Sarah Stewart-Brown, Director, Institute of Health Sciences, University of Oxford</td>
<td></td>
</tr>
<tr>
<td>Professor Ala Szczepura, Director, Centre for Health Services Studies, University of Warwick</td>
<td></td>
</tr>
<tr>
<td>Dr Gillian Vivian, Consultant, Royal Cornwall Hospitals Trust</td>
<td></td>
</tr>
<tr>
<td>Professor Graham Watt, Department of General Practice, Woodside Health Centre, Glasgow</td>
<td></td>
</tr>
<tr>
<td>Professor Kent Woods, Regional Director of R&amp;D NHS Executive, Trent</td>
<td></td>
</tr>
<tr>
<td>Dr Jeremy Wyatt, Senior Fellow, Health &amp; Public Policy, School of Public Policy, University College, London</td>
<td></td>
</tr>
</tbody>
</table>

## Past members

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