Identification of Randomized Trials for Inclusion in Meta-Analyses of Treatments for Childhood Acute Lymphoblastic Leukaemia, and Investigation of Factors Leading to Publication Bias

Thesis

IDENTIFICATION OF RANDOMIZED TRIALS FOR INCLUSION IN META-ANALYSES OF TREATMENTS FOR CHILDHOOD ACUTE LYMPHOBLASTIC LEUKAEMIA, AND INVESTIGATION OF FACTORS LEADING TO PUBLICATION BIAS

Julie Ann BURRETT

Degrees currently held:

Bachelor of Science with Honours in Mathematics, University of Durham

Diploma in Advanced Study with Distinction in Statistics, Oxford Brookes University

Degree for which this thesis is submitted:

Doctor of Philosophy of the Open University

Date of submission: 31 March 2003

Date degree awarded: 28 August 2003

Disciplines: Statistics, Clinical Trials

IDENTIFICATION OF RANDOMIZED TRIALS FOR INCLUSION IN META-ANALYSES OF TREATMENTS FOR CHILDHOOD ACUTE LYMPHOBLASTIC LEUKAEMIA, AND INVESTIGATION OF FACTORS LEADING TO PUBLICATION BIAS

J A BURRETT

CONTENTS ii

(i) ABSTRACT

Purpose: Some randomized trials are reported widely, while others remain unpublished. It is essential to systematic reviewers and meta-analysts that factors leading to publication bias in the form of delayed or non-publication of an eligible study are identified. This thesis is an attempt to do this.

Data: The set of randomized trials identified by the Childhood Acute Lymphoblastic Leukaemia (ALL) Collaborative Group was used. This consists of 149 trials comprising 243 randomized comparisons (randomizations), starting prior to 1 January 1988, reported in 257 articles, published prior to 1 January 2000. Each mention of a randomization in an article (irrespective of whether results are given) generates a publication record, of which there are 610.

Methods: The main focus is on identifying which trial characteristics lead to a delay in publication of a randomization. Time to the first mention of a randomization in an article (irrespective of whether any results are given) and to the first reporting of its results are both modelled using ordinary linear regression (the independence model). However, when these analyses are extended to include all mentions and all reportings of results respectively, non-independence necessitates the use of techniques for dealing with repeated measures. In such cases the independence model is the starting point, the residuals from which are used to form the covariance matrix, which in turn is used to suggest plausible correlation structures for repeated measures models. Generalised estimating equation (GEE) analysis is used to select an appropriate correlation structure, and a linear mixed effects model serves to confirm this. The conclusions are then discussed in the context of other studies identified. Finally logistic regression is used to identify trial characteristics associated with a randomization remaining unpublished, and Poisson and negative binomial models to identify those affecting frequency of reporting.

Results: Evidence was found of 'pipeline bias' in the reporting of first results since, although direction of effect was not found to be significant, highly statistically significant results are published faster than others. However this is not so for first mentions. Negative results (i.e. those in favour of the standard/control) arm were submitted for first publication faster than all others, although this did not effect time to publication. In addition, geographic location is an important predictor of whether a randomization is ever mentioned in an article, frequency of mentions and of time to first publication and results from single-centre trials are published more frequently than those with multi-centre participation.

Conclusions: Although 'pipeline bias' was identified in the analysis of time first reporting of results, it was not present in the analysis of time to first mention, and so not a problem for those wishing only to identify randomized trials for inclusion in meta-analyses. The importance of geographic location suggests that the practice of contacting known trialists is worthwhile in addition to the computerised literature searches and should be continued.

(ii) ACKNOWLEDGEMENTS

I would like to thank the Clinical Trial Service Unit for giving me the opportunity me to do this PhD and members of the Childhood ALL Collaborative Group[†] who provided information for the childhood leukaemia overview. I would also like to thank Dan Lunn* and Mike Clarke[†] for their supervision, Paul Garthwaite[#], Ruth Ripley* and Jill Boreham[†] for their helpful comments and Sue Richards[†] for support and encouragement throughout. Thanks also to the many work colleagues[†] including Helen Schoepf and Alison Palmer, family members and friends who have given practical help and moral support over the last five and a half years.

* Department of Statistics, University of Oxford

[†] Clinical Trial Service Unit and Epidemiological Studies Unit, Oxford

[#] Open University

(iii) LIST OF FIGURES AND TABLES

(iv) PUBLICATION ARISING FROM THIS RESEARCH

Some parts of this thesis were published as:

Burrett JA, Clarke MJ (2002) A descriptive study of randomized trials of treatments for childhood acute lymphoblastic leukaemia *British Journal of Haematology* **118:** 986-990

Permissions given by John Wiley & Sons Ltd..

SUMMARY

Chapter 1 provides an introduction to this thesis. It includes a brief history of the use of randomized trials in childhood acute lymphoblastic leukaemia and a description of some trial characteristics.

Chapter 2 describes the main aim of the thesis, which is to discover which trial characteristics are likely to affect three time periods: from the end of the accrual period to submission for publication, from receipt of the article by the publisher to actual publication, and the sum of these, the time from close of trial to publication. A literature search was conducted on this topic, and this is also mentioned in this chapter.

Chapter 3 describes the data collected; information on the set of clinical trials used, and the publications related to these trials. This chapter also deals with the data management process, inclusion/exclusion criteria and checking procedures. Chapter 4 explains how more meaningful variables were derived from collectable variables. It also describes some of the problems relating to data collection and coding and how these were overcome.

Chapter 5 looks at how the preliminary analyses were undertaken. Since the data are reportings of randomizations from clinical trials, there may be several records relating to a single randomization, which are not independent of each other. The problem of non-independence, and how it is overcome, using repeated measures analysis, is dealt with in Chapters 9-11. However, models that assume independence are used for the preliminary analyses in order to identify subsets of the explanatory variables which may be significant in the final analysis. The independence models are valid without further work when dealing with the first mention, or the first reporting of results, of each trial. Also covered in Chapter 5 are the exploratory graphical analyses and diagnostic plots. There are also sections on problems relating to the preliminary analyses, the incorporation of interaction terms, the use of indicator variables to represent classes of a categorical variable, the range of values over which the analyses are valid and outlying observations.

The six analyses for which the independence model can be used without adjustment are for the three response variables; time from close of trial to submission of article, time from receipt of article to publication, and time from close of trial to publication, each measured for the first mention of each trial and for the first reporting of results for each trial. The findings are given in Chapter 6.

Chapter 7 reports the results of the six preliminary analyses, which will need to be adjusted to take into account repeated measures. These are for the three response variables; time from close of trial to submission of article, time from receipt of article to publication, and time from close of trial to publication, each analysed for all mentions of each trial, and for all reportings of results for each trial.

Chapter 8 describes the method for obtaining the covariance matrix for each of the six analyses that involve repeated measures. This is used to suggest plausible correlation structures for repeated measures models. For each analysis this involves taking the residuals from the independence model, finding the covariance between the first and second publication, the second and third, the first and third and so on and developing an empirical correlation structure. Several possible correlation structures and preliminary investigations into selecting the most appropriate are discussed.

Chapter 9 discusses the application of generalised estimating equations (GEE) to the largest of the six datasets in order to select the most appropriate correlation structure to impose on all six datasets requiring repeated measures. A linear mixed effects model is also used in order to confirm the findings from the GEE analyses.

The two correlation structures judged most appropriate are then imposed on all six analyses and the final choice is made in Chapter 10. Chapter 11 reports the final results for the six analyses incorporating repeated measures analyses using GEE.

Chapter 12 describes an investigation into how the group of reportings of randomizations published prior to closure differ from the rest, and whether the results of the analyses change when these are excluded.

Chapter 13 presents the results of the twelve analyses in tabular form and discusses possible implications for the identification of randomized trials for inclusion in meta-analyses. The findings are also compared with those from other studies.

Chapter 14 begins with an introduction to other questions which can be answered from the data collected. The bulk of the chapter deals with one of the two main topics, that of investigating which trial characteristics affect whether or not a randomization is ever mentioned in an article, and whether or not its results are ever reported. The second main topic, trial characteristics affecting the number of articles in which a randomization is mentioned and the number of articles in which its results are published, is covered in Chapter 15.

An update of the data since the cut-off date for analysis is given in Chapter 16, and the main findings are summarized in the concluding chapter, 17.

1 INTRODUCTION

1.1 Introduction

This chapter provides a brief history of the use of randomized trials of treatments for childhood acute lymphoblastic leukaemia, describes some of the more interesting trial characteristics, and introduces the current project undertaken [Burrett and Clarke (2002)].

1.2 The use of randomized trials of treatments for childhood acute lymphoblastic leukaemia

During the 1950s randomized trials became increasingly common as a way of assessing treatments for many health conditions. However, leukaemia was not one of these. It was, at that time, a fatal disease. The general policy was to use any treatment that might prolong survival.

However, attitudes were changing. In 1963 the Medical Research Council Working Party on the Evaluation of Different Methods of Therapy in Leukaemia, in their paper comparing high and low dose steroids in the treatment of leukaemia in adults, stated that *'current forms of treatment* (for leukaemia) *were in need of critical appraisal'* and that *'the* (1957 Steering) *Committee* (had) *recommended that six Working parties should be formed, and that one of them should examine the possibilities of carrying out therapeutic trials in leukaemia'* [MRC (1963)].

Through the 1960s, survival rates in childhood acute lymphoblastic leukaemia gradually improved. By 1971 investigators at the St Jude hospital in Memphis concluded '*Childhood lymphocytic leukaemia can no longer be considered an incurable disease. Palliation is no longer a justifiable approach to its initial treatment'* [Pinkel *et al* (1971)]. Randomized trials were needed to distinguish between treatments with moderate differences in their effects on survival or disease-free survival.

In the early years of the 21st century, children with acute lymphoblastic leukaemia have a good chance of cure. Randomized trials are now used not only to investigate survival and disease-free

survival differences between treatments but also to investigate differences in their late effects and toxicity, and need to be larger than before, in order to detect smaller differences in outcomes between increasingly complex treatment regimens.

In the following examples the typical survival rates for the early 1970s and late 1980s were obtained from unpublished graphs using data from the series of Medical Research Council UKALL trials [personal communication with Dr S Richards] and the estimates of sample sizes number of events required are taken from Machin and Campbell (1987). Details of the trials are given in Appendix I.

- For patients recruited in the early 1970s, when the survival rate at ten years was typically 40%, in order to detect a 5% difference between two randomized treatments at the 5% level with a two-sided test with 80% power, 832 patients are required for each treatment arm, i.e. for a two-arm trial 1664 patients must be recruited. (Larger differences require a far smaller number of patients, for example to detect a 15% difference with the same significance level and power, only 89 patients are needed in each arm, i.e. a total of 178 for a two-arm randomization.)

- In contrast, the ten year survival rate for patients recruited in the late 1980s was typically 70%. This requires far smaller numbers of patients to be recruited. Using the same significance level and power, 343 per treatment arm are required (a total of 686 for a two-arm randomization) in order to detect a 5% difference between the two treatments and 29 per treatment arm (a total of 58) in order to detect a 15% difference. At first glance it may seem as if the targets for more recent trials should be easily achievable. However, the event rate for patients in these more recent trials is very low and the differences in outcome to be detected are very small. The number of events needed to detect a 5% difference is 343 in each treatment arm (a total of 686 in total for a two-arm randomization). In order to obtain a sufficient number of events far greater numbers of patients are needed.

The data on number of randomized patients in Section 1.4 indicates how rarely these targets were, and are, achieved. Therefore individual randomized trials are still rarely big enough to provide reliable evidence on some important outcomes on their own. Sufficiently large-scale randomized evidence might only be possible through systematic reviews. For these to be valid, as high a proportion of relevant trials as possible must be included [Clarke and Stewart (1994)], and this thesis is based on the trials that have been identified for a systematic review of treatments for childhood acute lymphoblastic leukaemia.

1.3 Current project

Trial identification can be the most difficult and time-consuming part of systematic reviews. It needs to overcome recognised biases that lead to some trials being published quickly and/or more than once while others never reach publication.

The Cancer Overviews Group at the Clinical Trial Service Unit, University of Oxford has considerable experience in the identification of randomized trials [EBCTCG (1990)]. In the early 1990s it began a collaborative overview of individual patient data from randomized trials of any treatment of childhood leukaemia [Childhood ALL Collaborative Group (1996)].

The methods used for trial identification included computer-aided literature searches (on-line and using CD-ROM); hand-searching of journals and of abstract books from major meetings; searching trial registers; and contacting known trialists who might have conducted, or know of, further trials.

A register was compiled of randomized trials, begun before 1 January 1988 and therefore eligible for the collaborative overview, and of reports of these trials published before 1 January 2000.

1.4 A description of some of the main variables

A total of 149 randomized trials and 257 reports of these have been identified (some of which reported more than one trial or randomization). The 149 trials include a total of 243 separate randomizations, since some trials involved more than one randomization (for example, for

induction and maintenance treatments). Ninety trials (60%) contained one randomization, 35 trials contained two, 19 trials three, two trials contained 4 and 5 and one contained ten randomizations. Of the 257 reports, 195 were journal articles, 51 meeting abstracts and 11 book chapters.

In many of the analyses presented here it will be the **randomization** rather than the **trial** which is the object of interest. For example, a trial may consist of two randomizations, each addressing a completely separate treatment question; the first between two induction treatments and the second between two different lengths of maintenance treatment. Throughout this report the terms 'trial' and 'randomization' will be used in this sense.

Unfortunately for some randomizations even the most basic information is missing. For example the Cancer Overviews Group may have identified a randomization but not know either when it took place or the number of patients it accrued. Although individual patient data is requested from the trialists, which could provide such information, there is no guarantee that this will be provided in all cases.

Number of patients

Data on the number of randomized patients were available for 212 of the 243 randomizations (Figure 1.1). The median number of patients accrued was 126. Less than 200 patients were accrued by 134 of the 212 randomizations (63%). Of these, 41 (19% of the 212) accrued fewer than 50 patients. Five randomizations included more than 1000 patients. The randomization with the most patients was the CCG-105 trial, which randomized 1606 children into four arms: intensified induction and consolidation, delayed intensification, both or neither. (The trial also contained a two-way randomization between cranial irradiation and methotrexate.) [Tubergen *et al* (1993b)]

Geographic location

Data are available on whether a trial is single- or multi-centre for 147 of the 149 trials. Of these 123 (84%) are multi-centre and 24 single-centre.

Information is available on whether a trial took place in one country or was international

for 129 of the 149 trials. Of these 129, 82 (64%) were single-country trials and 40 (31%) involved a few adjacent countries. For example those run by the BFM Children's Group (based in Germany) involved West Germany and Austria and trials coordinated by GATLA included Argentina, Brazil, Cuba and Uruguay. Only 7 trials (5%) recruited patients in different continents. For example, the USA Cancer and Leukemia Group B (CALGB) have run trials collaborating with Switzerland, Finland and South Africa.

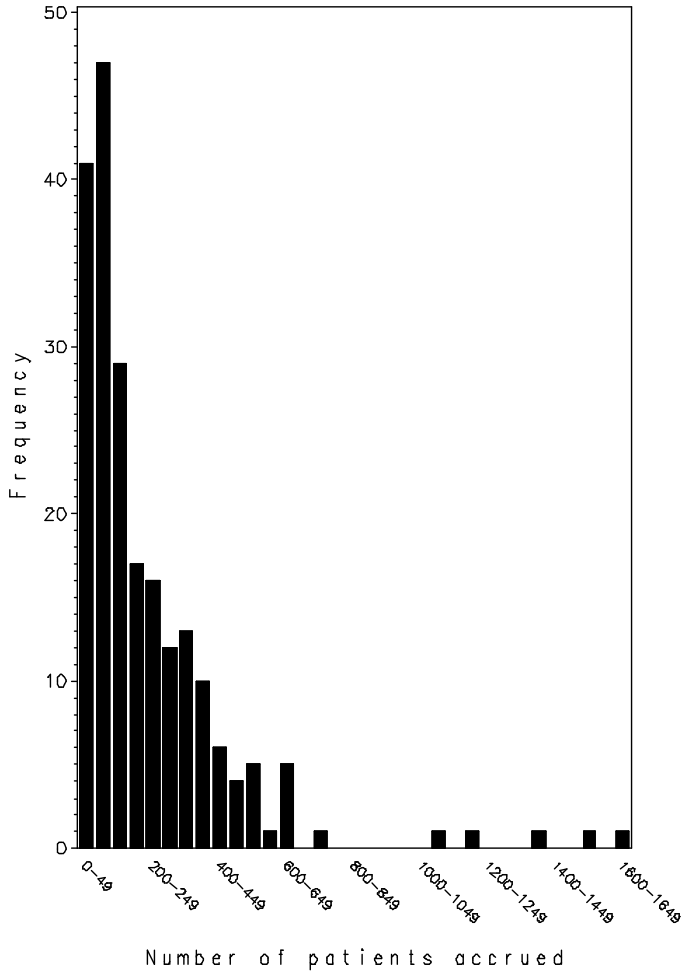


Figure1.1 Randomizations: number of patients accrued
31 randomizations do not contribute to the figure because the number of patients accrued is currently not available.

Date of start of trial

As mentioned above, trials beginning in or after 1988 were excluded from the list that forms the basis of this report. Data on the date the trial started are available for 139 of the 149 trials (Figure 1.2). The earliest identified randomized trial in childhood acute lymphoblastic leukaemia began in 1962. This was an immunotherapy trial run by the French Institut de Cancerologie et d'Immunogenetique (INSERM) [Mathe *et al* (1977a)]. There has been a steady increase in the number of new trials starting since then. The year 1979 appears to have been particularly productive, with 18 of the 139 (13%) trials with known start date beginning then. The five randomizations with the largest number of patients began in 1978, 1981, 1983 (2 trials) and 1985.

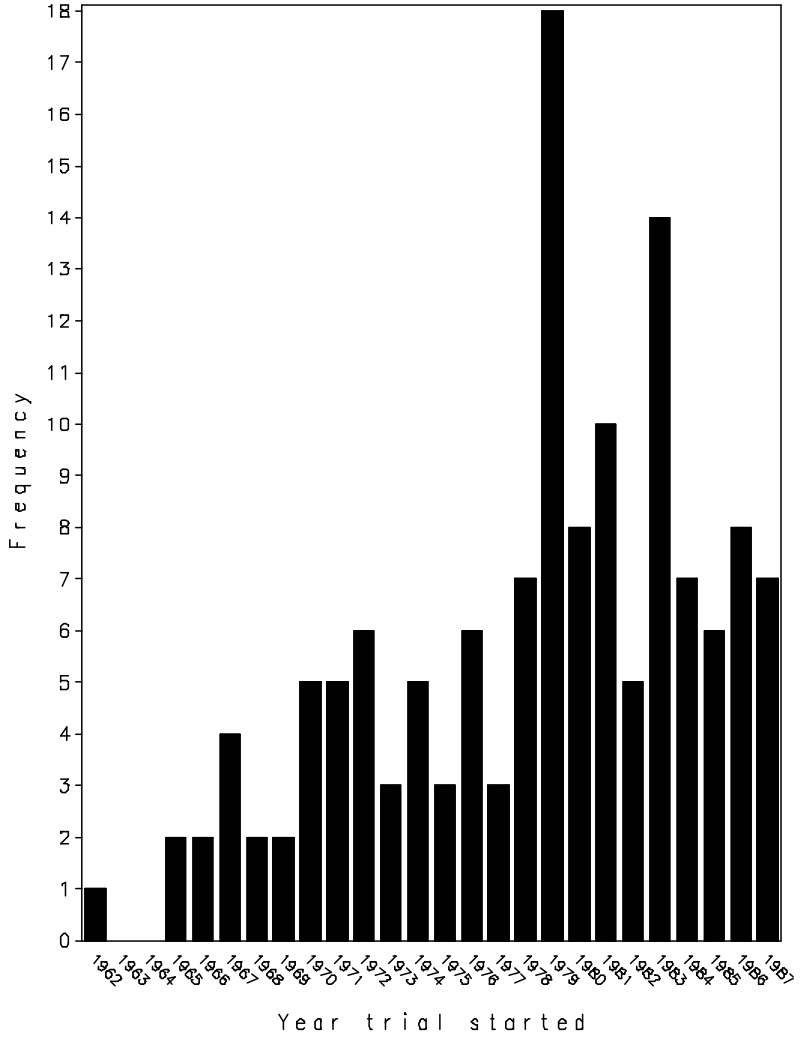


Figure 1.2 Trials: year started
10 randomized trials do not contribute to the figure because the year they started is currently not available.

Duration of accrual period

Data are available on the length of time that the randomization was open for 219 of the 243 randomizations (Figure 1.3). The median length of this period was 2 years and 7 months. 18 of the 219 randomizations (8%) were open for more than 5 years.

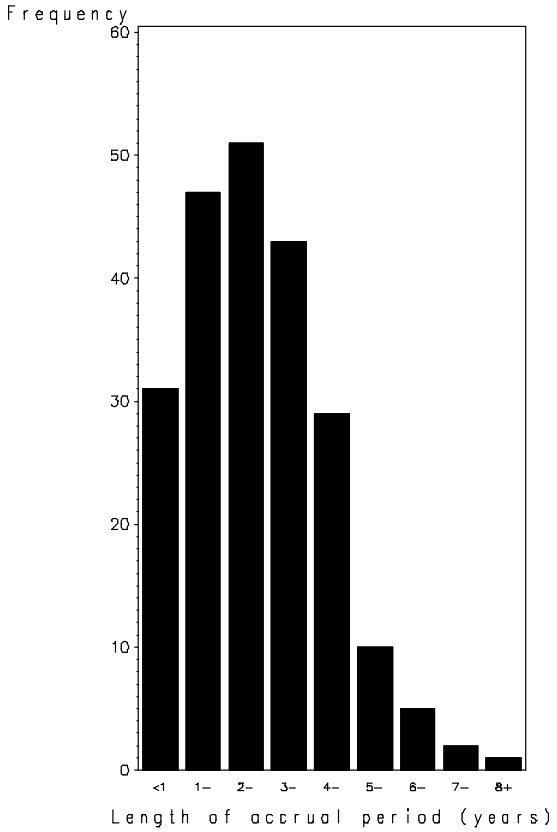


Figure 1.3 Randomizations: length of accrual period
24 randomizations do not contribute to the figure because one or both of their start and close date are currently not available.

Method of randomization

This was rarely reported in these trials in childhood leukaemia, as has been found in studies in other areas of health care [Juni *et al* (2001)]. Out of these 243 randomizations, information on the method of randomization was available in the reports for only 47. Of these, six randomizations were done via a central computer, 24 via notification to a central office and 17 using sealed envelopes in the individual centres.

Frequency of publication

Figure 1.4 shows the distribution of frequency of publication for the randomizations. Publications have not been found for 26 (11%), and these are assumed to be unpublished. Most randomizations have been published more than once, although the most frequent number of publications per randomization is one [84 (35%) of 243 randomizations]. The five randomizations with the largest number of patients were published three, four, five and six (2 trials) times.

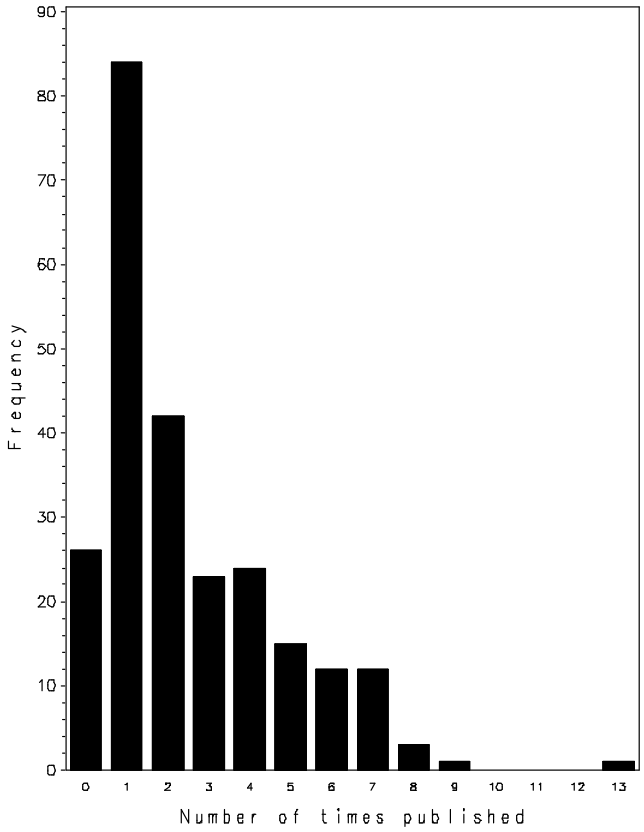


Figure 1.4 Randomizations: number of publications

The most frequently published randomization (13 reportings) was SJCRH X, a single randomization equivalence trial, conducted by the St Jude Children's Research Hospital in Memphis, USA [Abromowitch *et al* (1988a, b), Bowman *et al* (1984), Mulhern *et al* (1991), Ochs *et al* (1983), Ochs *et al* (1986), Ochs *et al* (1989), Ochs *et al* (1991a), Pui *et al* (1985), Pui *et al* (1989), Pui *et al* (1991), Pui *et al* (1992), Williams *et al* (1991)]. This was a randomization between two first-line consolidation and maintenance treatments. The second treatment group included cranial irradiation during consolidation and a more complex maintenance regimen than

the first. The trial was open between May 1979 and January 1984, and accrued 309 patients. It was funded with both Government and charity money. The randomization procedure was done using a central computer and a minimisation of imbalance design. The trial was open to standard risk children and was a single-centre study.

The first report for the trial was published in December 1983 [Ochs *et al* (1983)], and the first report giving results appeared in March 1984 [Bowman *et al* (1984)]. The thirteenth reporting was in February 1992 [Pui *et al* (1992)]. All reports were published in the English language. Eight of the articles reported results. Six of these reported on the main questions in the study, by comparing the relapse rates and in particular the CNS (central nervous system) relapse rates for the two randomized treatment groups. The other two articles dealt with other outcomes only: one compared the testicular relapse rates for the two groups, and the other the incidence of acute myeloid leukaemia as a secondary disease.

1.5 Conclusion

This introductory chapter gives a historical and descriptive account of randomized trials in childhood leukaemia since the earliest such studies in the 1960s. It focuses on trials that began before 1988 in order to make use of the register of trials that was developed for a systematic review of treatments for acute lymphoblastic leukaemia in children. Many randomized trials will have started since then, [Childhood ALL Collaborative Group (2001)], but the comprehensive searching done for this existing register has not yet been done for more recent trials. The number of randomized trials starting each year has increased from one or two in the 1960s to an annual average of five or six in the 1980s. However trials remained relatively small, with more than half of all randomizations accruing less than 200 patients, and only five having more than 1000. Most trials were published more than once, typically in journals. Unfortunately, as has been found with other series of trials, the reporting of the methods of these randomized trials in leukaemia could have been better.

Since many of the trials of treatments for childhood leukaemia are small, systematic reviews are essential in order to obtain sufficient numbers of patients to produce reliable results. In order to be unbiased, these should include patients from all eligible randomizations if possible. The primary aims of the project investigating this series of randomized trials are to assess the extent of searching needed to identify trials and to discover which factors are leading to publication bias in the form of delayed publication or to a randomization remaining unpublished, since this could, in turn, bias the findings of the systematic review.

This brief introduction highlights some important factors relating to this: the longevity of randomized trial research in this disease, the fact that most of these trials have been relatively small and conducted within single countries, and that many of them have been published once only.

2 AIMS

2.1 Introduction

This chapter begins by stating the main aims of this thesis. Various types of bias in the reporting of clinical trials are then described. It then sets out the main questions asked in this study. Finally a literature search undertaken to identify similar research in other areas is described.

2.2 Main aim of this thesis

As was stated in Section 1.5, since many of the trials of treatments for childhood leukaemia are small, systematic reviews and meta-analyses are essential in order to obtain sufficient numbers of patients to produce reliable results. Ideally all eligible randomizations should be identified and included. However some are more quickly and more widely published, and hence more easily identified than others. Publication bias in the form of non-publication or delayed publication could therefore lead to the results of a systematic review being based on a biased sample of randomized trials.

"It is of the utmost importance that as high a proportion as possible of all relevant trials are identified, regardless of their results or publication status. Any trials that are missing should not be too numerous or unrepresentative to affect the results of the meta-analysis in any important way."
[Stewart and Clarke (1995)]

The main aim of the thesis is therefore to identify, and where possible, to quantify potential biases in types of trial identified and included in meta-analyses and to get some idea of the effect of this bias.

2.3 Bias in the reporting of clinical trials

In the introduction to the paper 'Modelling publication bias in meta-analysis: a review', Sutton et al (2000) describe various types of bias that can cause problems when conducting meta-analyses. Apart from publication bias, where research with statistically significant results is more likely to be

submitted and published than that with 'null' or non-significant results, four other types of bias are outlined:

- Pipeline bias – results with smaller *p*-values associated with them are likely to be published faster than others
- Subject reporting bias – only results of a trial with the smallest *p*-values are reported
- Duplication reporting bias – authors deliberately report the same results more than once
- Language bias – for example where only papers written in the English language are searched for trials for inclusion.

The first of these, 'pipeline bias', is the primary focus of this thesis but the importance of trial characteristics other than the *p*-values associated with results is also investigated.

2.4 Questions to be asked

Methods currently used for trial identification are as follows:

- Hand-searching abstract books from conferences worldwide for relevant trials
- Computer-aided literature searching, on-line and using CD-ROM
- Searching trial registers
- Contacting trialists and asking whether they know of any other randomized trials apart from those already identified. This is especially useful for identifying unpublished trials

Which factors affect how long the searching process should continue in order to identify all trials through publication?

The sooner a trial is reported, the sooner it can be identified by those conducting meta-analyses, and the sooner its results can be incorporated into the meta-analysis. To discover why some trials take longer to get published than others, it is necessary to investigate which factors (or trial characteristics) affect the following three time periods:

- Time from close of randomization to date of submission/receipt for publication (an indication of how important the trialists consider their findings)

- Time from date of receipt/submission to date of publication (an indication of how important the journal editors consider the trial)
- Time from close of trial to date of publication (the sum of the above two time periods, and for which there are much more data available than for either of the above alone)

Only a small proportion of the journals (and no meeting papers or books) state date of submission/receipt. Some also give date of acceptance of the article. However the date of close of randomization and date of publication are known in almost all cases. Time from date of submission/receipt to date of publication is more useful than time from date of acceptance to date of publication because the time taken for the editor's decision to be made is included in the former.

Analyses of how various trial characteristics affect the above three time periods provide the basis of this thesis.

How wide does the search need to be?

The more widely a trial is published the more likely it is to be identified by those conducting meta-analyses, and its results included. Therefore the other major aspect to investigate is why some trials are published many times, some once only and others not at all. Also how do trials published in the most prestigious journals differ from the other trials? The latter is important since trials reported in such journals are more easily located and hence more likely to be included in systematic reviews.

This means looking at the following questions:

- Which trial characteristics affect the number of articles in which a randomization is mentioned, or whether a randomization is published more than the median number of times?
- Which trial characteristics affect the type of journal in which a randomization appears?
- How do the unpublished randomizations differ from those published?

Some of these questions will be discussed in Chapters 15 and 16 and it is intended that they should be investigated fully at a later date.

2.5 Literature search

A brief literature search was undertaken. This deals only with the main topic of this thesis; factors affecting length of time to publication for a randomization, as opposed to those affecting either the number of times published or causing it to remain unpublished.

The Cochrane Methodology Register (in Issue 3, year 2002 of the Cochrane Library) was searched on 7 September 2002. It contained a total of 3866 references at that time. The keywords searched for were 'publication' and 'delay', such that both were required. This produced a list of 14 references with abstracts. All 14 were read on screen and of these seven were found to be relevant. Each of these seven abstracts mentions a time period to do with publication as a response variable and various trial characteristics, typically the *p*-values associated with results, which were investigated to see if they had an effect on the time period. Full papers were obtained where possible. This summary of other similar research is not an attempt at a systematic review of the literature, merely an attempt to put the findings from this project in the context of other work in the field.

Chapter 13 contains a comparison of these studies with the work in this thesis and Appendix XIX summarises the studies identified in tabular form. A brief description is given for each study, including the time period response variable, the characteristics of the trials investigated and a summary of the findings.

3 DATA AND DATA MANAGEMENT

3.1 Introduction

This chapter begins by describing the data used; the set of randomized clinical trials and the publications relating to them. This is followed by sections on how the data were collected and the structure of the data. The data management process is described next and the algorithm for the main data management program is given. A list of the trial characteristics collected is given, as are the inclusion/exclusion criteria, and the checking procedure used. There are also sections focusing on new trials found and on why some records were deleted from the database.

3.2 Data

The closed set of trials used is as follows: all properly randomized trials beginning before 1 January 1988[1] and all articles referring to these trials published before 1 January 2000, as identified by the Cancer Overviews Group at the Clinical Trial Service Unit, University of Oxford. Appendix I summarises the trials, randomizations and articles used to compile the data. For some randomizations additional information was obtained from the trial protocol. These are listed in Appendix II. Trials still open to randomization would have been eligible for inclusion, as well as those that have closed. However, although date of closure is missing for 20 of the 243 randomizations, it is believed that all had closed by the cut-off date for analysis, 28/11/00.

All properly randomized trials in childhood acute lymphoblastic leukaemia (ALL) were eligible for inclusion in this thesis. Randomizations, for example, between treatments for relapsed patients, and between antibiotic treatments were not included in the 1992 Collaborative Overview, but are included here.

[1] This was the set of trials included in the Second International Collaborative Workshop on Childhood ALL Studies at the end of 1992.

3.3 Starting point for data collection

The starting point for data collection is the trial summary. For each randomization they identify, the Cancer Overviews Group produce a trial summary, which provides the following information:

- name and reference code of the group running the trial
- name and reference code of the randomization
- name and reference code of the trial that the randomization is linked to (In some cases this will be the same as the randomization reference code.)
- type of trial (for example whether the randomization is between different lengths of treatment or between different types of therapy)
- entry dates (start and close date of the randomization period)
- total number of patients randomized
- status code ('ok' indicates that the randomization procedure used is valid, the other three codes 'nr', 'bin', or 'dup' indicate that it is not. Further explanation will be given in Section 3.6.)
- whether the results have been published and, if so, whether as an abstract or in full
- eligibility criteria relating to age, white blood count, other disease details, risk group, prior treatment etc.
- treatment arms (brief outline)
- which are the main questions (if there are several treatment arms, which comparisons the trialists consider most important)
- list of publications reporting this randomization
- contact details for the trialists
- treatment arms (in more detail)

As mentioned above, the trial summaries list the publications for each randomization. Each article is important, as the goal is not to discover in which article a randomization was first identified, but rather all possible articles in which it could have been found. All the listed articles were read, all

details on trial listings (except the last two) were checked and additional information was filled in on a coding sheet. It was especially important to check the entry dates for the randomizations, since the start date determined whether a randomization was eligible for inclusion. A single article might report several trials, each consisting of several randomizations. Coding sheets were completed for all reportings of all eligible randomizations.

The data from the coding sheets were then entered into an ACCESS database using codes rather than descriptive character fields wherever possible, so that they were in a suitable format for analysis. For each categorical variable, a new code was added whenever necessary without too much consideration at that stage. The categories were combined more meaningfully, once there was a feel for the data.

3.4 Structure of the data

The ACCESS database consists of four tables. Three contain roughly the same information as each of the coding sheets; a table for variables to do with the article itself, another for variables specific to the trial and a third for those to do with the randomization. The fourth table links the article with the trials it reports.

In the publications table, each article was assigned a unique reference number in the order the data were entered into the database from the coding sheet.

In the trials table, each trial also has a unique identifier – this is the number assigned to the trial by the Cancer Overviews Group. The first three digits indicate the group of trialists who conducted the trial, and the fourth and fifth specify the particular trial.

The randomizations table differs importantly from the tables for publications and trials. Since each trial may contain more than one randomization, and each publication may report more than one trial, there is a record for each **reporting** of each randomization, i.e. multiple records for each randomization. Again each record has its own unique identifier, assigned by the computer in order

of entry to the database. The randomization code (assigned by the Cancer Overviews Group) and the code of the trial to which it belongs are also stored. From here onwards these randomization records, each obtained from a different article, are referred to as **'publication records'**.

In addition a **'definitive record'** for each randomization was created. This contains the 'best' data for each field taken from looking at all reportings together with information from other sources where available such as raw data, the Cancer Overviews Group trial listings, trial protocols and *Clinprot* trial summaries. *Clinprot* is an electronic protocol database, a register of trials. The trials for which there are further data either from *Clinprot* or from the full trial protocol are listed in Appendix II. For randomizations published several times or for which a protocol is available, the definitive record is likely to be complete. However, for others, particularly those remaining unpublished, important information may be missing.

By definition the definitive record only contains data unchanging over time, for example size of trial, start and close dates for entry. In contrast, details such as results reported are specific to a publication record. For some analyses it is sufficient to use just the definitive records. For others it is necessary to look at all reportings, i.e. the set of publication records. Also, the number of reportings of each randomization must be attached to the definitive record for that randomization.

An ACCESS 'query' was compiled to pull together from the three tables all fields that might be useful in the analyses. At this stage descriptive fields such as name of journal and authors' details were omitted. The data in the query were then read in as a SAS dataset. SAS was used for the majority of the data management as well as for the analysis.

Each record in these data is a randomization record, either a publication record relating to an individual randomization from a particular article or a randomization's definitive record. Information about its trial has been appended, as has information about the article, if it is a publication record. If the record is a definitive record, this is indicated.

The algorithm for the main data management program, JMAIN.SAS, is given as Appendix III. This performs the following tasks:

- Reads in the data from the ASCII file produced from the ACCESS query.
- Incorporates special records from a small dataset. These were created to replace individual records, which have publication dates that tie. The reasons for this will be discussed in Section 5.3.1.
- Excludes records where the randomization and/or the article are ineligible. See Section 3.6.
- Creates new variables for the purpose of combining categories of some variables more meaningfully. This will be discussed further in Sections 3.14 and 4.2.
- Date fields are character fields of the form DD/MM/YYYY. For each date field, creates a numerical SAS date variable for use in arithmetic operations, and also a discrete 'year' variable for use in some graphs.
- Converts numerical unknowns such as '999' into missing values.
- For dichotomous yes/no variables where blank and 'no' are essentially the same, converts blanks to 'no'.
- Splits the records into 'publication records' and 'definitive records' and creates a new permanent dataset for each.
- Totals the number of publication records for each randomization and attaches this information to the definitive record for that randomization.
- Removes dates too inaccurate to be useful in the analyses, and positively misleading. These have either year or year and decade unknown and should be replaced by missing values. They cannot be removed earlier in the process since estimated start of randomization and date of publication are used in deciding whether the randomization/publication satisfies the date inclusion criteria.
- Calculates new variables from existing ones such as duration of randomization period = close date – start date. This will be discussed further in Section 4.2.

- Creates a new ‘merged’ dataset using data from the ‘definitive records’ dataset for unchanging variables such as size of trial, start date etc. and data from the ‘publications’ dataset for those that vary between reportings such as publication date or details of the results reported. This is the dataset used for the majority of the analyses.
- Attaches an ‘order of publication number’ to each record. For each randomization this specifies the order in which the articles were published. Records which contain results have an additional order number assigned, so that it is also possible to know, for each randomization, the first article to publish the results, the second and so on. The merged datasets with publication ordering are kept as two permanent datasets, for all records and all records containing results respectively.

3.5 Trial characteristics

Appendix IV contains two listings of the trial characteristics collected. The first is in alphabetical order and gives the variable’s (SAS) name and meaning, its primary use being as a look-up table. The second lists the variables in logical order, with the original categories tidied up and made meaningful in the data management program. As well as the description of the variable, its SAS name as used in the analyses is given, as are its type (dichotomous/categorical/continuous), whether it is in the ‘publication records’ dataset, the ‘definitive records’ dataset or both, and the number and percentage of missing values. For continuous explanatory variables the mean, standard deviation and range are given as Appendix V.

3.6 Exclusion criteria for randomizations

- Randomization opened on or after 1 January 1988.
- Those believed by the Cancer Overviews Group to be duplicates. These are marked status ‘dup’ on the trial summaries.
- Those allocated status ’bin’ by Cancer Overviews Group. For example studies identified which did not accrue any patients.

- Those believed by the Cancer Overviews Group not to have been properly randomized, i.e. where the treatment allocation could have been predicted. These have status 'nr' and include cases where date of birth has been used to allocate treatment. This is not a valid randomization technique since the allocated treatment can be predicted in advance.
- Where the randomization is known to have been intended to be exclusively for adult patients. Publications relating to these have not been sought systematically by the Cancer Overviews Group. However some were found by chance and include older paediatric patients whose data were used in the 1992 overview. Different groups use different cut-off ages for inclusion in paediatric trials, for example Medical Research Council trials in the UK use 15 years but many US trials use 21.

3.7 New randomizations identified

Identification of new randomizations was not an aim of this project. It was assumed that the Cancer Overviews Group's thorough procedures would have already done this. However during the process of data collection, 23 (what were thought to be) new randomizations were identified. These were checked with the Cancer Overviews Group. Subsequently six were discovered to be duplicates of randomizations already found and seven had already been found but were rejected because they were either not randomized or not properly randomized. This left nine genuine new randomizations. Both publication records and definitive records for these were added to the database. The randomization IDs assigned are for the purpose of this project only and numbered in reverse, starting at 99 for each trial group. For example there are four new randomizations for trial group 14, namely 1499, 1498, 1497 and 1496. Details of the new randomizations have been passed to the Cancer Overviews Group for processing and inclusion in future meta-analyses.

Brief descriptions of the genuine new randomizations identified now follow. Four of the new randomizations found are from trials conducted by Cancer and Leukaemia Group B (CALGB), USA (group 14), two from the Children's Cancer Group (CCG), USA (group 16) and three from GATLA, Argentina (group 21).

- 1499 is one of three randomizations belonging to trial 1402 (ALGB 6601). The other two had already been identified in the usual way. This randomization began in 1966 and was open to patients on a particular arm of another randomization (1403). It was between continuing maintenance with twice weekly methotrexate after 8 months versus stopping treatment. There are two publications, a journal article, publication 52 [Holland and Glidewell (1972a)] which mentions 8 other randomizations, and a book chapter, publication 53 [Holland (1978)]. The former states that 8 out of the 43 patients on Regimen C (Treatment 3) were allocated to continue the maintenance treatment and that these numbers were too small to form any conclusions.

- 1498 is also mentioned in publication 53 [Holland (1978)]. This is one of four randomizations from trial 1404 (ALGB 6801). Again the other randomizations had been identified already. The entry period began in 1973. This was a maintenance duration randomization, open to patients on the best arm from randomization 1406 and still in remission at 5 years.

- 1497 is one of a pair of randomizations from trial 1428 (ALGB). The other randomization had already been identified. The entry period was during the 1960s and 19 patients, who had relapsed but were still in second remission, were randomized between two maintenance treatments: subcutaneous cytosine arabinoside + intramuscular methyl both weekly versus oral 6-mercaptopurine + oral methotrexate single dose. This was reported as a journal article, publication 71 [Jones et al (1972)], which states that the maintenance treatment received was not a determinant of length of remission. The stronger treatment fared worse than the less intensive.

- 1496 is one of a pair of randomizations from trial 1429 (CALGB 6611). The other had already been identified. The entry period of this 3-arm maintenance randomization began in 1966. The arms were daunarubicin weekly versus oral mercaptopurine 90 mg/m^2/day +

intramuscular methotrexate 15 mg/m^2/wk versus subcutaneous arabinoside 30 mg/m^2/wk + intramuscular methylglyoxal bis guanylhydrazone 350 mg/m^2/wk. This is mentioned in the journal article, publication 288 [Jones et al (1971)].

- 1699 is one of two randomizations from trial 1645 (ALBG 6601). Again the other had already been found. It was open between 1963 and 1967 and accrued 165 patients. This was a maintenance randomization of cyclic versus sequential chemotherapy. Its only mention is as a meeting abstract, publication 107 [Nesbit et al (1973)]. It is worth noting that this was the main randomization of the trial. The randomization that was found in the routine way was for maintenance duration, to stop or continue after 2½ years' treatment, and it accrued only 15 patients.

- 1698 is one of three randomizations from trial 1612 (CCG-141). It opened in 1975 and randomized between two CNS treatments, 24Gy cranial irradiation + intrathecal methotrexate versus intrathecal methotrexate. It is mentioned in a journal article, publication 97 [Bleyer et al (1983a)].

One journal article, publication 284 [Sackmann-Muriel et al (1998)] provided the means to identify the following three randomizations. The new randomizations described so far are all linked to known trials. However the following had not been found previously:

- 2199 (Protocol 11-ALL-67) is a single-randomization trial which was open between November 1967 and September 1970. It compared two maintenance treatments; methotrexate 30mg/m^2 versus 15mg/m^2 twice weekly, both arms also receiving 6-mercaptopurine + vincristine + prednisolone pulses every 6 months with a total duration of treatment of 5 years.

- The other two randomizations 2198 and 2197, identified in this article belong to the same trial, Protocol 1-ALL-76. Both were open between January 1976 and December 1978 and accrued 336 patients. 2198 was a maintenance randomization between

vincristine/prednisolone alternating with cytarabine-cyclophosphamide and vincristine/prednisolone alone. 2197 randomized patients between immunoestimation with levamisole versus not, during continuation treatment.

3.8 Comparing trial characteristics between randomized and non-randomized trials

Data on non-randomized studies and their reportings have not been collected, although some non-randomized studies were found unintentionally by the Cancer Overviews Group while searching for randomized studies. These are not included in this thesis. However in some studies, non-randomized patients were treated alongside randomized patients, and mentioned in papers, which describe the randomized group, for example, the non-randomized patients in UKALL I who were treated at centres giving CNS prophylaxis to all or none of their patients, as opposed to randomizing between the two treatments. See publications 5 [Medical Research Council (1973)] and 1 [Medical Research Council (1975)]. A code to indicate this is attached to the randomization record so that this information could be retrieved easily. The data from the non-randomized contingent are not used for this thesis, but have been typed separately into the 'notes' section of the record for the randomization to which it is linked. Therefore it would be possible to pull out randomizations with a non-randomized contingent and compare, for example, the benefits of CNS treatment for those from the randomized group who received it with those who received it but where randomization did not take place. Also it would be possible to compare CNS treatment versus control in the randomized patients with CNS treatment versus the control group of the non-randomized patients. Using only the non-randomized patients treated alongside those who were randomized into the study provides a closed and well-defined set.

3.9 Exclusion criteria for publications

- Date of publication on or after 1 January 2000. This date was chosen to create a closed set of publications. It is likely that most articles relating to trials that opened before 1 January 1988 would have been published by this date. For some articles the date of publication is

incomplete, for example frequently the journal reference includes the year of publication but gives no mention of the month or day. If, however, the date of publication is known to be during the 1990s or known to be post millennium, it is clear whether or not the article should be included in data used for this thesis.

- Review articles summarising several trials. To be excluded the publication has to satisfy all the following: (i) reports trials conducted by more than one group of trialists (ii) not written by a Working Party member (iii) reports summaries of past trials only, with no new results. The reason for this is that the aim is to include new information. Therefore a report of results from analysis of updated data from a randomization previously reported is eligible, but re-reporting of unchanged results is not.

- Reports of meta-analyses, whether or not new information is reported, for both published and unpublished trials.

- Draft papers, since these were only available for one trial group, the Medical Research Council, UK.

- Unpublished articles, since these have not been collected systematically.

- *Clinprot* is an online electronic database of records for clinical trial protocols, maintained by the National Cancer Institute in the USA. It was searched for randomized trials of treatments for leukaemia in 1989. This, alongside the full trial protocol, where available, can be used to check trial characteristics. However, neither *Clinprot*, nor the full trial protocol, count as publications in themselves.

- Articles which do not include any of the following: that a randomized allocation was used, any details of the treatment arms, information on outcome by treatment. An example of an article excluded for this reason is publication 208, the conference abstract Rivera et al (1989).

3.10 Duplicate publications

Attempts to publish a paper are generally sequential in that it is submitted to a journal and, if rejected, then submitted to a second. An example of where this procedure appears not to have been followed, and a pair of duplicate publications resulted, is the article *Comparison of Intermittent or Continuous Methotrexate Plus 6-Mercaptopurine in Regimens for Standard-Risk Acute Lymphoblastic Leukemia in Childhood (JCCLSG-S811)* by The Japanese Children's Cancer and Leukemia Study Group, which was published in *Cancer*, publication 161 [Koizumi et al (1988a)] and also in *Medical and Pediatric Oncology*, publication 163 [Koizumi et al (1988b)]. Both are included in the data since both are a possible source from which a randomization can be identified.

3.11 What counts as a 'publication' of a randomization?

An article may report a randomization in detail but mention several others briefly. The following are examples of situations where a very brief mention warrants representation as a publication record, provided that it is clear that randomization took place:

- This is the first mention of a randomization. For example, if the article reports the results of a first randomization and mentions that the trial also contains a second, and the latter is not cited at the end of the paper.
- The design of the next trial is mentioned, typically in the discussion.
- Previous trials, which have never been reported, are mentioned.

Some examples

An interesting example is publication 37 [Riehm et al (1984)]. In this article the discussion between several collaborators was reported verbatim. The following exchange is sufficient to establish the existence of a randomization:

> *SIMONE: How long should you give maintenance therapy?*
>
> *RIEHM: As Dr Pinkel pointed out, there may be differences regarding high risk patients. We have a randomized study designed to answer that question, but at this point I would speculate that 1½ years is enough. For low-risk patients, I'm not sure.*

This was the first mention of the maintenance duration randomization (1204) of trial ALL-BFM-81 (1204).

Publication 54 [Rowland et al (1984)] is an example of where it is easy to decide which randomizations should have a publication record. It is entitled *'Effects on Different Forms of Central Nervous System Prophylaxis on Neuropsychologic Function in Childhood Leukaemia'*. In the '*patients and methods'* section it states clearly the four CNS prophylaxis randomizations reported in the paper, and that there were also randomized comparisons between induction and between maintenance therapies in the trials, which have already been published. Therefore publication records were created for the four CNS randomizations only. These are 1405 from the ALGB 6801 trial (1404), 1408 from the CLB 7111 trial (1407), 1414 from the CLB7411 trial (1414) and 1416 the sole randomization of the CLB7611 trial (1416).

Not all articles are as clear. Publication 46 [Riehm et al (1987)], discusses the use of risk factors for prognosis in five BFM studies since 1970. The following sentence suggests three randomizations. In fact there are two only, 1204 for maintenance duration, and 1205 for CNS prophylaxis treatments, the latter for standard risk patients only. *"Randomizations for preventative CNS irradiation in SR patients, duration of therapy, and the introduction of a more intensive therapy in HR patients have been the rationals and questions in study ALL-BFM 81."*

Publication 69 [Rausen et al (1979)] reports the three randomizations (1423, 1424 and 1425) of the Group 14 trial: CLB 6911 4-way (1423). In the discussion, reference is made to the induction randomization (1407) of the CLB 7111 trial. However a publication record was not created, since this is not the first mention, and the word 'randomization' is not used in connection with it. Instead it is stated that: *'An extension of these observations led to the* ***evaluation*** *of L-asparaginase therapy in combination in newly diagnosed children with acute lymphocytic leukemia by the Cancer and Leukemia Group B (Protocol 7111).'* There is also a reference to a relapse trial, ALGB, trial name unknown, (1428), but again this had been previously reported and so no publication record was created.

It can be difficult to know whether previous trials have been reported or not. Publications 18 [Peto et al (1986)] and 19 [Eden et al (1987)] are very similar. Both compare the Group 27 trial UKALL VIII (2727) with previous trials UKALL II –VII. However, the former states that UKALL IV, V and VI were previously unpublished. Therefore a publication record was created for each randomization mentioned in the article. However, the latter, a book publication, does not mention this and, so, publication records for the previous randomizations have not been created from this second article.

However, enough information must be given to decide that there are at least two treatment arms, and that a randomization process is used. An example of where insufficient information is given is publication 55 [Holland and Glidewell (1972b)], which reports three of the randomizations (1404, 1405 and 1406) from the Group 14 ALGB 6801 trial (1404). Publication records were created for these. It also mentions a later trial CLB 7111 (1407), stating that *'Protocol 7111 ... contains therapeutic programs incorporating the best regimens shown in Figures … and a modification of 6601 with intensive parenteral courses of methotrexate and of 6-mercaptopurine together with vincristine and prednisolone reinforcement dosing. The early results are superior to our prior regimens.'* Since no indication is given that trial CLB 7111 is randomized and neither the number of arms nor what they consist of is clear, no publication records were created for any of the three randomizations belonging to trial CLB 7111.

Some examples of publication records and articles deleted from the ACCESS tables will be given in Section 3.12.

Since one criterion for trial inclusion is that the entry period must start before 1 January 1988, some of the later papers report both eligible and ineligible trials. An example of this is publication 283 [Richards et al (1998)]. This describes combined results from two Medical Research Council childhood ALL randomized trials, UKALL X which is eligible and UKALL XI which is not. A publication record was made for UKALL X (randomization 2730) only and not for UKALL XI

(randomizations 2932, 2733 and 2734). Currently there are ineligible randomizations in the ACCESS database – usually references to future trials. These are not included in any analyses.

3.12 Examples and notes on which randomizations are included from an article and have a publication record in the randomizations table in ACCESS and which do not

Why have some records been deleted from the randomizations table in ACCESS?

This occurs with Group 27 (Medical Research Council, UK), the first group dealt with. Some randomizations were mentioned very briefly in an article, but not reported as such. If the reference was to a previous trial, a publication record has not been created, but if to a future trial, and therefore probably a first mention, then one has.

An example of this is publication 10 [Chessells et al (1992a)]. Primarily this paper deals with trial UKALL X (2730), which has a single randomization between four intensification treatments. However a previous trial, UKALL VIII (2727), consisting of an induction randomization (2727) and a maintenance duration randomization (2728) is mentioned also. Initially these were entered on computer as records 383 and 384 in the randomizations table but they have subsequently been deleted.

Publication 1 [Medical Research council (1975)] reports the CNS prophylaxis randomization (2701) of the UKALL I trial (2701), record 363 in the randomizations table. It also briefly mentions a second randomization belonging to that trial, for maintenance duration/immunotherapy (randomization 2702). Since this is a later randomization, probably not previously mentioned, it should have a record in the randomization table. Due to initial indecisiveness it was first entered as record 364 in the randomization table, then deleted, and finally re-instated as record 405.

A similar situation arose with publication 4 [Campbell et al (1973)] which also primarily reports the CNS prophylaxis randomization in UKALL I (randomization 2701, record 370 in the randomizations table) but also briefly mentions the second randomization for maintenance

duration/immunotherapy (randomization 2702). This latter randomization was first entered as record 372, deleted, then re-instated as record 406.

This should serve as an explanation for the missing record numbers in the randomizations database. With an auto-numbering system, this was unavoidable.

Why are some records deleted from the articles table in ACCESS?

Again, Group 27 was the pilot for this work, being the first group dealt with. Publication 12 was from *Clinprot*, not strictly a publication. It was later decided that any information from this source should be added to the 'publication 0' (definitive) record for each randomization in the randomizations table in order to eliminate unnecessary records.

The 'publication 0' record for each randomization is an amalgamation of data from all publications found, the Cancer Overviews Group's trial list (available for all randomizations) and from *Clinprot* and/or any trial protocols obtained from the groups. There is a separate field in the database, which specifies the sources.

3.13 Checking procedure

A series of codes were used to denote queries and points of interest. These are numerical codes entered into the 'notes' field of randomization or article records in the ACCESS database. At the end of the data collection process the codes were evaluated and checked where considered necessary. Records with codes corresponding to the following were checked, and the data amended where appropriate:

- Studies thought to be newly found randomizations. These were added to the database and the Cancer Overviews Group were informed.
- Review articles and editorials summarising several (often past) trials. Some of these were excluded.
- Two or more publications for the same abstract or full paper. These were left in the database but investigated to see if there was anything unusual about the trial reported.

- Only combined results, for randomizations from more than one trial, were given. These were removed from all records.
- Only combined results for randomizations from the same trial were given. These were attached to all relevant records.
- Only combined results were given and it is uncertain whether they are for randomizations from the same or different trials. These were checked, re-classified and dealt with as described above.
- Possibly for exclusion (miscellaneous category)
- Unclear which data relate to adult patients (for exclusion) and which to children
- Number of patients and possibly other data seems to have been attached to the wrong randomization of a particular trial.

Notes codes relating to the following were not investigated, since the Cancer Overviews Group have investigated these thoroughly as part of their identification process:

- The trial – randomization structure appears wrong
- A proper method of randomization appears not to have been used

Records with other notes codes were not investigated unless outliers from the preliminary analyses had these attached to them. In this case other observations used in the analysis were checked for the code. If no others had the code, then the code might be significant. If, however, several non-outliers also had the code then it is unlikely that it is important.

After exclusions were removed, 257 articles reporting 149 trials, consisting of 243 randomizations remained. There were 610 publication records.

3.14 The trial characteristics (variables) considered

The trial characteristics collected are listed in Appendix IV. Here the variable name as used in the analysis is given along with a description of what it is, its type (categorical, continuous or

dichotomous) and whether it is found in publication records, definitive records or both. The latter was discussed in Section 3.4.

4 DERIVING MEANINGFUL VARIABLES FROM COLLECTABLE VARIABLES

4.1 Introduction

The information that is easiest to collect is not necessarily in the most useful form. This chapter begins by dealing with the process of calculating the latter from the former, describing the derivation of individual variables. Next some of the problems relating to data collection and coding, and their solution, are outlined. These include: (i) the need to re-assess definitions and make corrections, (ii) variables of interest that could not be collected and other variables used as surrogates, (iii) variables collected which could not be analysed, (iv) variables with missing values that could be estimated and (v) categorical variables with an uneven spread of data. The final section of this chapter deals with other points of interest concerning individual variables.

4.2 Calculated variables

For calculated variables, data from the definitive record were used where possible, since this is considered most accurate. For example when calculating the time interval from close of randomization to date of publication (TCLPUB) using the merged dataset, the date of close of randomization (NCLOSE) was taken from the definitive record. However date of publication (NDPUB) is different for each article and so was taken from the publication record. The following are examples of calculated variables:

Duration of randomization period

The duration of randomization period (DURRAN) is the difference between date randomization started (NSTART) and date randomization closed (NCLOSE).

Target accrual reached

The target accrual number is considered reached (TARGET) if actual number of patients (NOPAT) is equal to or exceeds planned size of randomization (PLSIZE).

Multi-centre participation

A trial is considered to be multi-centre (MULTIC) if the number of centres participating is greater than or equal to five; multi-centre (limited) if the number of centres is two, three or four; and single-centre if only one centre participates.

Target number of centres reached

The target number of centres is considered reached (CTARGET) if actual participation is as great as or greater than planned centre participation. This is calculated by comparing multi-centre code (MULTIC) with planned centre code (PLCENT), both of which have the same three categories; multi-centre, multi-centre (limited) and single-centre. A disadvantage of comparing MULTIC and PLCENT by category is that, for example, if the planned number of centres was stated to be 10 but only 6 actually participated, this would resolve as target number of centres reached. However, to use the actual numbers to calculate CTARGET would lead to a far larger proportion of missing values for this variable, since often both intended and actual participation are stated as simply 'multi-centre'.

Decline in interest

The decline in interest in a randomized comparison (WANE) can be assessed crudely by subtracting the accrual time to randomize the first half of the patients from that for the second half

i.e. (date last patient accrued – date middle patient accrued) –
(date middle patient accrued – date first patient accrued)

These accrual dates were obtained from raw data collected by the Cancer Overviews Group.

Number of questions

In a randomization, if the number of treatment arms is greater than two, number of questions (comparisons) (NOQ) depends on what the arms (ARMS) are, since some comparisons are more important than others. For most randomizations the number of (important) questions is known to the Cancer Overviews Group. For the rest a crude count, based on the number of treatment pairs

there are, can be used. For a randomization with N arms, number of questions = 1 + 2 + …+ (N-1) = $N(N$-1)/2 (Table 4.1) It was necessary to use this for 12 randomizations.

The number of arms and the number of questions are, of course, very highly correlated (Pearson Correlation coefficient $\rho = 0.94412$ using the definitive records for the 243 randomizations). In any regression analysis only one of ARMS or NOQ was used at a time. If it was eliminated, the other was tried in its place. If it remained in the reduced set of significant variables, the other was also tried to see if this improved the fit of the model.

Number of arms	Number of questions
2	1
3	3
4	6
5	10
6	15
7	21
8	28

Table 4.1 Number of arms and number of questions

Size of randomization

The scatter diagram of the time period in question against number of patients accrued (NOPAT) in Figure 4.1 shows bunching to the left, indicating that a logarithmic transformation of 'number of patients' is necessary. The order of magnitude, expressed by LOGSIZE = log_{10} (NOPAT), is of interest rather than the actual number of patients accrued. Under this transformation the graph shows a more even spread.

The three key time periods being investigated

- Time from close of randomization to date of submission/receipt of article (TCLREC). The date of submission is assumed to be the same as the date of receipt for analysis purposes.
- Time from submission/receipt of article to publication (TRECPUB)
- Time from close of randomization to publication (TCLPUB)

For brevity these time periods will be referred to as 'close to submission', 'receipt to publication' and 'close to publication'.

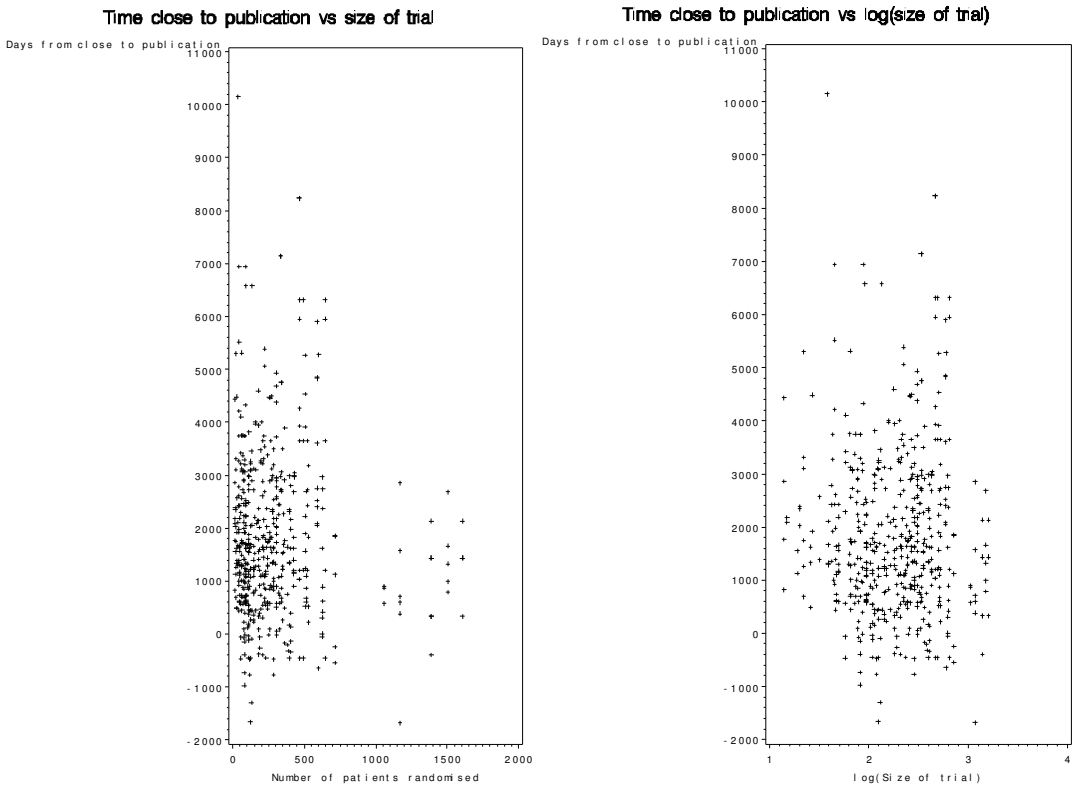


Figure 4.1 Use of scatter plots to determine whether the use of a transformation of a continuous variable is necessary. Example: time from close to publication vs. size of trial and vs. log_{10} (size of trial)

Length of follow-up period

When reporting results, some articles state the median or mean length of follow-up (FUPDAYS). If this is not given, but the cut-off date for analysis is given, it can be estimated roughly in the following way:

length of follow-up period =cut-off date for analysis – ½ (start date + close date)

Time from cut-off date for analysis to submission of article

Conversely, if the cut-off date for analysis is unknown, but the mean or median number of days on follow-up is known, then an estimate for the time from cut-off date for analysis to submission can be calculated using:

time from cut-off date for analysis to submission of article

= date article received – [½ (start date + close date) + number of days on follow-up]

4.2.1 How some of the results variables were obtained

From each reporting of a randomization, up to two main trial questions (e.g. survival, event-free survival etc.) and up to two main results were collected. The reason for allowing two main questions and results is that it is not always clear from the article which is the main question and main result. For each result the following information was sought:

- type (the same categories as for the main questions, for example; survival, event-free survival etc.)
- statistical method used
- length of follow-up of the study
- whether the article answered the main questions
- *p*-values associated with the results
- clinical significance
- direction (The direction of a result is considered positive if the experimental arm fared better than the standard arm, negative if it fared worse and null if there was no difference between the two.)

From the pairs of question and result fields an overall value was selected for use in the analyses. To make this selection hierarchical systems were used, as listed in the algorithm for the data management program given as Appendix III. However this method led to a complication which is described in Section 4.4 in the discussion of the relationships between the three variables used to represent the results of a randomization.

Statistical technique (TECH)

The broad categories used, ranked in order of merit, are:

- Cox regression (proportional hazards): a form of survival analysis but with the added advantage of allowing extra variables to be included in the analysis
- longitudinal (survival analysis) methods i.e. methods for analysing curves such as Kaplan Meier and log rank tests

- cross-sectional tests which analyse data at one point in time, such as the Mantel-Haenszel or chi-square test, generally not thought to be as good as either of the above
- other methods

If either result was obtained using Cox Regression, then the overall statistical technique for that publication record is 'Cox Regression'. Otherwise, if one or both used a longitudinal method, then the overall technique is 'longitudinal'. Otherwise, if cross-sectional methods were employed, then the overall technique is 'cross-sectional'. If another method was used then the category is 'other'.

Whether the article has answered the trial questions (ANSWER)

To decide whether the article has answered the main question(s) stated in that article the following strategy was used. If two main questions were stated and both were answered, or if only one main question was stated and it was answered then the paper has answered the main questions. If no main questions were described in the paper then it is not known whether they were answered. Otherwise the paper is considered not to have answered the main questions.

Statistical significance (LOGPEST)

An important variable is the measure of statistical significance. From each article two main results were coded and, of the two, that with the smallest *p*-value was used. Some articles reported actual *p*-values. Others stated that the results were significant or not at a particular level (0.05, 0.01 or 0.001). One way of representing statistical significance in the regression is to take the logarithm of the *p*-value in each case. However, since the *p*-value was not always reported, this would have led to many missing values. An alternative method is to treat statistical significance as a categorical variable with, for example, 0=not significant, 1=significant at the 0.05 level, 2=significant at the 0.01 level and 3=significant at the 0.001 level, which would have resulted in fewer missing values. However, information concerning the spacing between the significance levels would have been lost. Therefore, for each category of statistical significance a typical value of *p* was assigned and log_e of this taken, creating the new variable, LOGPEST. This is the distance of log_e (typical value of *p*) for each category from that for the not significant/not reported category, and can be treated as

a continuous variable, with the spacing between values for each category built in. The five categories are given in Table 4.2.

Statistical significance	**Typical value of *p***	***log*** $_{e}$ **(typical value of *p*)**	**Spacing (LOGPEST)**
Not significant/not reported	0.15	-1.9	0
Possibly significant	0.075	-2.6	0.7
Significant at 0.05 level	0.05	-3.0	1.1
Significant at 0.01 level	0.01	-4.6	2.7
Significant at 0.001 level	0.001	-6.9	5

Table 4.2 Levels of statistical significance and how these are represented in the analyses

An additional variable (LOGPNR) is set to 1 if statistical significance is not reported and to 0 if it is. By using the two variables LOGPEST and LOGPNR in conjunction it is possible to distinguish between results that are not significant and those that were not reported.

Initially the intention was to record whether the *p*-values reported were one- or two-sided. Since a two-sided *p*-value is more conservative, the reporting of a one-sided *p*-value could indicate that the result would not have been significant if a 2-sided test had been used. However this was abandoned since it was rarely reported.

Clinical significance (CLNSG)

Clinical significance can be thought of as a hard measurement (calculated), rather than merely an impression. McNeil (1996) gives the following definition in his comparison of statistical and clinical significance:

"It is important to understand the difference between 'statistical' and 'clinical' significance. According to scientific convention, a finding is statistically significant if the p-value associated with the null hypothesis of interest is smaller than 0.05. On the other hand, a result is clinically important if the confidence interval for the parameter of interest differs from the null value by a worthwhile amount."

However, for the purpose of this project, where articles rarely give concrete figures either for the 'worthwhile amount' or the confidence interval, it is necessary to use the impression given by the paper. A result was considered to be clinically significant if the wording of the text, irrespective of the *p*-value stated, suggested that it clearly favoured the experimental arm, or in the case of an equivalence trial the experimental arm was at least as good as the standard arm. A result was considered to be not clinically significant if it favoured the standard (control) treatment. If the impression was less clear then 'possibly' and 'not known' categories were used.

If one or both of the two main results were reported to be clinically significant then the overall clinical significance category is 'yes'. Otherwise, if one or both of the results might possibly be clinically significant then it is 'possibly'. Where the article implies that neither result is clinically significant or that it is too early to tell then the clinical significance category is 'no'. Where no information is given about clinical significance, the category is 'not reported'.

Direction of results (POSNG)

As has been mentioned previously, results can be **positive** (the new treatment was found to be better than the standard treatment), **negative** (the new treatment was found to be less effective than the standard treatment) or **null** (results are about the same for the two treatment arms).

If both main results (or one if only one was reported) are positive, or if one is positive and the other null, then the direction category for the publication record is 'positive'. Similarly, if both main results (or one if only one was reported) are negative, or if one is negative and the other null, then the direction category for the publication record is 'negative'. If both main results (or one if only one was reported) are null then the direction category is 'null'. If one main result is positive and the other negative, then the direction category is **'opposite'**. If no indication of direction of results is given then the category is 'not reported'.

4.3 Some problems relating to data collection and coding and how they were solved

4.3.1 The need to re-assess definitions and make corrections

For each of the following variables, the data collection process was well underway when it became necessary to re-assess definitions and make corrections.

Clinical significance (CLNSG)

It is important not to be influenced by the *p*-value associated with a result when deciding whether that result is clinically significant. Late realisation of this necessitated re-reading the results section of each article and re-classifying clinical significance where appropriate, whilst attempting to ignore the *p*-values quoted.

Whether data from a meeting abstract have necessarily been presented at a meeting (PRESENTD)

Articles can be from journals, books or meeting abstracts (PUBTYPE). Also there is a variable indicating whether or not data from the article have been presented at a meeting (PRESENTD). Initially the assumption had been made that data from a meeting abstract must necessarily have been presented at that meeting and the 'presented' flag set accordingly. Later, after discussion with colleagues, it was decided that this is incorrect and the 'presented' flag should only be set to 'yes' if the article actually stated that the data had been presented at a particular meeting or if a meeting abstract contained results. This has since been corrected. However, there remains the problem that results obtained between the submission of the abstract and the meeting presentation may have been presented even though the abstract does not contain them.

Equivalence trials (EQUIV)

The purpose of many trials is to discover whether a new treatment is better than the standard treatment. However, others aim to show that a less toxic treatment is equally as effective as the standard treatment. The latter are known as '**equivalence trials'**. There is a variable in the data to indicate whether a randomization is thought to be an equivalence trial. Special care must be taken when coding results of equivalence trials. For example; a null result (no statistical difference in the

results of two or more treatment arms) in an equivalence trial can be (and probably is) clinically significant. [Djulbegovic and Clarke (2001)]

Combined results

Some articles combine data from more than one randomization as a result. It was decided that if these randomizations belong to the same trial then the results should be attached to both/all records for those randomizations. However if the randomizations are from different trials, the results should not be attached to either/any. An example of attaching combined results to more than one randomization record is trial 2403, COALL 85/89, which contains two randomizations, 2403 and 2404. Randomization 2403 compared slow rotation of six drug combinations versus rapid rotation. Randomization 2404 also compared slow rotation versus rapid, but using only five drug combinations (dropping the high dose cytosine arabinoside and asparaginase block, [Janka-Schaub et al (1988), Janka-Schaub et al (1990), Janka et al (1991), Janka-Schaub et al (1996)] publications 158, 159, 156 and 157 respectively. Where combined results are reported, whether or not the results are attached, this is indicated by a code in the notes.

Start and close date for entry period (NSTART, NCLOSE)

If a trial contained more than one randomization, often the start and close dates for entry were only reported for the first randomization. If the second randomization was at a late stage in the treatment, for example for an extra year's maintenance duration, using the same entry dates for both randomizations would be inaccurate. For this reason raw individual patient data were used to check and refine these dates. The date of randomization was obtained for the first and last few patients in each randomization. These data had been collected, by the Cancer Overviews Group, from the trialists who had conducted the randomizations, for use in the overview.

Number of patients (NOPAT)

The number of patients, for a given randomization, is often reported differently between articles. Likely reasons for this variation are that an article may include some adult patients, or that some patients may be registered but relapse, die or become protocol deviants before the point of

randomization. The latter is a flaw in the trial design and makes the analysis more difficult, since all randomized patients must be included and analysis should be by intention to treat. In order to avoid this problem, the randomization process should take place as late as possible. Papers may also include non-randomized patients who had been treated alongside those who had been randomized. The individual patient data were used to help decide on the correct number of patients randomized.

4.3.2 Variables of interest that could not be collected and other variables that could be used as surrogates

Cost of the trial

This was not mentioned in any article. It may be possible to obtain the estimated cost (as opposed to the actual cost) by contacting all trial groups. However since the cost of follow-up of trial patients is unlikely to be available, this idea was abandoned. Instead, size of the randomization, the number of patients accrued (NOPAT), can be used as a surrogate for cost.

4.3.3 Variables collected that could not be used in the analyses

Decline in interest in a randomized comparison (WANE)

Unfortunately the data used to assess this were available for only 105 of the 243 randomizations (43%). For 85 of these 105 (81%), the accrual times for the first and second halves of the randomization were judged to be similar (i.e. less than 6 months difference). Six randomizations took considerably less time (i.e. more than 6 months quicker) to accrue the second half of their patients than the first half. This might be an indication that the interest in the trial increased over time. However, 14 randomizations took more than six months longer to accrue the second half of their patients than the first, which might indicate a decline in interest in randomizing into the trial. Due to the large number of missing values for this variable, it could not be incorporated into the analyses.

Details of the randomization procedure and trial design

Four variables to do with the randomization procedure and trial design were created using the sparse information reported in a few articles and trial protocols:

- Method of randomization used (RANDMETH) has the following categories: 'by a central computer' (the most reliable), 'notification to a central office' (less good), and 'use of sealed envelopes' (believed to be the least secure i.e. the easiest to predict). This is known for only 47 of the 243 randomizations (19%).

- Timing of late randomization (RANDTIME). As mentioned previously, the randomization procedure should be scheduled as late into the trial as possible, so that few randomized patients relapse, die or deviate from the protocol. This variable is only applicable to randomizations where the arms differ late on in the protocol. The categories are 'late randomization done at correct time' and 'late randomization done too early'. There is sufficient information on this for 67 of the 243 randomizations (28%).

- Randomization design (RDESIGN) has three categories: 'simple randomization', 'block randomization' and 'minimisation of imbalance'. Simple randomization is self-explanatory. No attempt is made to balance either the numbers randomly assigned to the different treatment arms, or to balance for any additional factor, for example by sex. An example of block randomization is as follows: if the randomization is between two treatment arms (A and B) and uses a block of size four, the number of patients randomized to each treatment will always be equal every fourth patient. For example if the first three patients of four are allocated A, B, B then the fourth must receive A. Minimisation of imbalance is a more complex way of balancing the numbers allocated to two or more treatment arms. The advantage of this is that it is possible to balance groups for several factors, for example, age group, sex, white blood count group, hospital entering the patient into the study. This is known for only 13 of the 243 randomizations (5%).

- Whether any attempt was made to balance or not (BALANCED) is the fourth variable, which is known for 94 of the 243 randomization (39%).

Due to the vast majority of observations having missing values for all four of the above, none of these could be used in any analyses.

Statistical techniques used (TECH)

Another variable, not initially collected, is the statistical method reported. It was decided, towards the end of the data collection process to go back through all articles for this. One intention was to investigate whether a conclusion obtained using a new technique is more or less likely to be published and how the use of different statistical methods has changed over time. Another was to look at whether non-standard tests are used in order to obtain a significant result, not achievable using standard methods. However, in order to do this properly it must be known whether or not the group conducting the trial routinely uses the non-standard test, and this information was not available. These questions were not investigated due to the large proportion of articles that fail to report methods used. Of the 394 publication records which contain results, the statistical technique is given for 201 (48%). The categories for this variable were described in Section 4.2.1.

Number of days on follow-up (FUPDAYS)

As was described in Section 4.2, if either the mean or median time on follow-up is reported, then this is used. Otherwise the following estimate is calculated:

$$\text{cut-off date for analysis} - \tfrac{1}{2}\,(\text{start date} + \text{close date})$$

If the response variable is 'time from close to submission', or 'to publication', then number of days on follow-up is necessarily significant. The controlled period from trial closure to cut-off date for analysis must not be included in the response variable. An alternative approach is to use as the response variable 'time from cut-off date for analysis to submission', rather than 'time from close of randomization to submission'. However, this idea was abandoned for two reasons:

- ‘Time from close of randomization to submission for publication’ is more meaningful than ‘time from cut-off date for analysis to submission for publication’.
- The cut-off date for analysis is dependent on the date of submission, and may also be dependent on explanatory variables, perhaps the best example being the statistical significance of the results. If a randomization is published early there will be fewer events and so the *p*-value associated with the result may remain greater than 0.05. If published late there will be more events, perhaps resulting a *p*-value of less than 0.05.

The only time period analysed, for which number of days on follow-up can used, is time from receipt of article to publication. When all records containing results are included there are 50% missing values for this variable. When first results only are included, 60% observations have missing values. So again the variable cannot be used.

Risk group (RISK)

Data were collected on the type of patients who were eligible for randomization. As well as age group, the risk group was often also specified. The following categories were used:

- any
- low
- low-standard
- standard
- standard-high
- high

Preliminary graphs did not indicate any association between risk group and any of the ‘time to’ response variables. There was still no pattern when the categories were compressed as follows:

- any
- low/standard
- standard/high
- high

Also for any regression there are approximately 40% observations with this variable missing. Therefore this variable was omitted from the analyses.

Target number of patients reached Yes/No (TARGET)

As described previously this variable is calculated using the actual number of patients accrued (NOPAT) and the intended number of patients (PLSIZE). Unfortunately the latter is rarely given resulting in TARGET being missing for 206 of the 243 randomizations (85%). For this reason TARGET is excluded from the analyses.

4.3.4 Variables with missing values that could be estimated

Dates

ACCESS and SAS were both used in this project. These require full dates to be used. Where partial dates only are known a qualifier is used to indicate the accuracy of that date:

- If month and year are known the date used is 15/MM/YY and qualifier is set to 1.
- If year only is known the date used is 30/06/YY and the qualifier is set to 2.
- If year is not known but the decade is then the date used is 30/12/Y4 where Y=6 for the 1960s for example, i.e. the middle of the decade is used, and the qualifier is set to 3.
- If the decade is unknown it is estimated, the date is set to 30/12/Y4 similarly, taking the decade most likely to be correct, and the qualifier is set to 4.

Dates with qualifiers 3 and 4 are used only for deciding whether randomizations and articles are eligible for inclusion (as described in Sections 3.6 and 3.9). For analysis purposes they are considered potentially misleading and have been removed.

International/single-country participation (INTERNL)

If it is known whether a group generally conducts international or single-country trials but this information is missing for a small proportion of their randomizations, it can be assumed that the missing information is the same as the typical policy for the group.

Multi-centre/single-centre participation (MULTIC)

Similarly, if a trial group usually runs multi-centre (or single-centre) trials, it can be assumed that the randomizations where this information is missing are also multi-centre (or single-centre). There

is also a variable to indicate whether the randomization was planned as a multi- or single-centre trial (PLMULTIC). Missing data for this can be imputed in the same way.

Country group of the centre conducting the trial (CGROUP)

The broad categories 'North America', 'Europe' and 'Other' were obtained using the first three digits of the trial or randomization identifier in conjunction with the list of centres provided by the Cancer Overviews Group.

Country group of publisher (JGROUP)

Again the three categories are 'North America', 'Europe' and 'Other'. The articles themselves, the Internet and informed colleagues provided this information. A third variable was then created. This indicates whether the country group of those conducting the trial is the same as that of the journal publishing it *(SGROUP).*

4.3.5 Categorical variables with a very uneven spread of data

The following meaningful categorical variables, with few enough missing values to be useful in analyses, have a very uneven spread of data. None were excluded from the analyses for this reason.

However it is important to be aware of this since missing values, either for the response or for any of the explanatory variables used, result in a reduced dataset being used in the analysis. This could mean that a particular category of a variable may be represented by only one or two observations. If membership of that category appears to have a significant effect on the response i.e. it remains in the model, this must be investigated, since a single observation (or two observations) should not be allowed to influence the choice of model so strongly.

This is usually taken care of when consulting preliminary graphs prior to the analysis. Categorical variables are graphed using box-plots, as will be described in Section 5.3.2. Often it is appropriate to pool a category with few observations with another category, and so the problem of a category with few observations having undue effect on the choice of model does not arise.

Table 4.3 lists variables where at least one category (not including the missing values category) contains less than 10% of the total. The figures quoted may be based on the total number of trials, randomizations, articles or publication records, whichever is most appropriate for that variable.

Funding (FUNDG)

Category	Number of trials (%)	
1=Government	53	(35.6%)
2=drug company	**1**	**(0.7%)**
3=charity	**1**	**(0.7%)**
4=Government + charity	28	(18.8%)
5=Government + charity + drug-company	**1**	**(0.7%)**
missing	65	(43.6%)
Total	149	(100%)

Treatment type (TXCHEMO)

Category	Number of randomizations (%)	
Chemotherapy	167	(68.7%)
Bone Marrow Transplant	**1**	**(0.4%)**
Radiotherapy	54	(22.2%)
Immunotherapy	**11**	**(4.5%)**
Antibiotic	**3**	**(1.2%)**
missing	7	(2.9%)
Total	243	(100%)

Age eligibility (AC)

Category	Number of randomizations (%)	
Children only	214	(88.1%)
Both adults and children	**23**	**(9.5%)**
missing	6	(2.5%)
Total	243	(100%)

Multi-centre/single-centre participation (MULTIC)

Category	Number of randomizations (%)	
Multi-centre (>=5)	199	(81.9%)
Limited multi-centre (2-4)	**9**	**(3.7%)**
Single-centre	32	(13.2%)
missing	3	(1.2%)
Total	243	(100%)

Target number of centres reached? (CTARGET)

Category	Number of randomizations (%)	
Yes	169	(69.6%)
No	**8**	**(3.3%)**
missing	66	(27.2%)
Total	243	(100%)

International/single-country participation (INTERNL)

Category	Number of randomizations (%)	
Single-country	126	(51.9%)
Limited international	68	(28.0%)
International	**19**	**(7.8%)**
missing	30	(12.4%)
Total	243	(100%)

Publication type (PUBTYPE)

Category	Number of articles (%)	
Journal	195	(75.9%)
Book	**11**	**(4.3%)**
Meeting abstract	51	(19.8%)
Total	257	(100%)

Country group of publisher (JGROUP)

Category	Number of articles (%)	
US	164	(63.8%)
Europe	74	(28.8%)
Other	**2**	**(0.8%)**
Missing	17	(6.6%)
Total	257	(100%)

Published in full in the English language? (ENGLISH)

Category	Number of articles (%)	
Yes	250	(97.3%)
No	**7**	**(2.7%)**
Total	257	(100%)

Clinical significance (CLNSG)

Category	Number of publication records (%)	
Yes	191	(31.3%)
Possibly	**17**	**(2.8%)**
No	121	(19.8%)
Not reported	281	(46.1%)
Total	610	(100%)

Table 4.3 Distributions of categorical variables with a very uneven spread of data

4.4 Other points of interest concerning individual variables

The majority of the explanatory variables are categorical. Some are graded or ordered categorical variables with more than two categories. For each categorical variable indicator (or dummy) variables are used, numbering one less than the number of categories. If an indicator variable is found to be the least significant at a stage in the backwards elimination process of the regression, care must be taken to combine categories in a meaningful way.

Trial category

Randomizations were categorised by the treatments being compared:

- induction
- central nervous system prophylaxis
- intensification
- maintenance
- a combination of induction, intensification and maintenance
- duration of treatment
- testicular radiotherapy

- bone marrow transplant
- immunotherapy
- treatment after relapse
- miscellaneous such as antibiotic or cardio protection

For analysis purposes, two new variables were created from the above:

- first-line or relapse therapy
- chemotherapy only or transplant or radiotherapy or antibiotic or other

Funding source (FUNDG)

The categories for funding source are Government, drug company, charity or any combination of these. This field is attached to the trial as opposed to the individual randomization, since it is unlikely to differ between randomizations belonging to the same trial. It may be interesting to look at changes in funding source over time, for example to investigate whether there may be a trend from Government funding to funding from other sources. This has not been addressed in this thesis.

Main questions, main results and whether the main questions have been answered: in publication records and in the definitive record

Data on two of the main questions, two of the main results, and whether the two main questions have been answered, have been collected from each publication. The main question asked in a study is often survival. The main result reported may also be what the trial was designed to look at e.g. survival. However it could be an additional finding, with more endpoints available for analysis and having a smaller *p*-value associated with it for example event-free survival, disease-free survival or relapse in a specific site. Often, in the paper, the main question is not specified.

It is a fairly straightforward process to deal with one article at a time, and to decide whether the main questions as reported in that paper have been answered in it. However there are other measures of interest to do with the main questions of the trial and whether these have been answered. One of these is the overall main question of the trial at its conception. This may not

necessarily be mentioned in a paper, particularly if the article is reporting the answer to a different question. It is likely to be stated in the trial protocol or *Clinprot*, neither of which are available for many of the randomizations.

An attempt was made to estimate the overall main question for each randomization, using all available sources. If the trial protocol or *Clinprot* are available, the main questions specified (CQUEST1 and CQUEST2) take precedence over those mentioned in articles (PQUEST1 and PQUEST2). If more than two questions were stated in the protocol, then two were selected using the hierarchy given below. These were ranked in order, with the exception of 3, 4 & 5, ranked equal.

1=survival

2=EFS (including DFS)[2]

3=treatment-related deaths

4=achieving CR

5=any relapse

6=specific relapse sites

7=toxicity

8=other

Otherwise the articles alone were used to deduce the main questions. Here the main questions stated in all articles for a randomization were pooled and the two with the highest ranks from the hierarchy were used. These were attached to the definitive record for that randomization. A new variable was then created to specify whether the two overall main questions were ever answered in any article, or if there was only one question, whether this was ever answered (ANSEVER). If so, it is set to 'yes'. If there was no indication anywhere what the main questions were, it is set to 'unknown'. Otherwise it is set to 'no'. This information is also stored in the definitive record.

[2] Note: Event Free Survival (EFS) is time from diagnosis until induction death or failure, death in remission or first relapse. Disease-free survival (DFS) is the time from achieving complete remission to relapse or death.

Unfortunately, the main questions for randomizations where *Clinprot* and/or the trial protocols were available differed so greatly from those for randomizations where articles were the only source that 'whether ever answered' (ANSEVER) was not used in the analyses. Its inclusion might have led to bias between the two groups. However it is valid to use 'whether the main questions stated in a paper were answered in that paper' (ANSWER) in the analyses, since this is available for the whole set of publication records.

Relationships between the three variables used to represent the results of a randomization

(i) Statistical significance and direction of results

Section 4.2.1 explained how three results variables, measures of statistical significance, clinical significance and direction of results, were collected for two of the main results for each randomization reported in an article and how a hierarchy was used to select the definitive value from the two, to be used in the analyses.

Frequency tables for level of statistical significance versus direction of results and for level of statistical significance versus clinical significance categories were produced. From the former it can be seen that the majority of statistically significant results ($p<0.05$) are in a positive direction, i.e. in favour of the experimental treatment arm. Of the 610 records, 129 have one or both results with statistical significance $p<0.05$. Of these 129, 80 (62%) are positive, 21 (16%) are negative, i.e. in favour of the standard treatment, 17(13%) are in opposite directions, i.e. one result is positive and the other negative, 10 (8%) have direction not reported and 1 (<1%) is a null result. This latter observation has been investigated, as described below.

If, for a particular record Result 1 has a *p*-value of <0.05 associated with it and Result 2 does not, but it is unclear whether Result 1 is in favour of the experimental arm or of the standard arm, then the values used in the analyses are as follows: statistical significance is taken from Result 1, since a *p*-value of less than 0.05 takes precedence over a *p*-value of greater than 0.05, whereas direction of result is taken from Result 2 since a null result takes precedence over 'not known'. Therefore the record contains a result both statistically significant at $p=0.05$ and null.

Record 697 contains the results of randomization 1902, an immunotherapy randomization from a two-randomization trial 1901 (EORTC 58741), open between 1972 and 1980 and accruing 123 patients. The record is from the sole publication for that randomization, publication 135 [Otten et al (1988b)]. The two results extracted are as follows:

> *Result 1:* Disease-free survival: The difference between the two arms is statistically significant with p=0.024, but it is unclear which was the experimental arm, so the direction is coded as 'unknown'.
>
> *Result 2:* Survival: The result is null with 65% survival in both arms. Therefore statistical significance is 'no'.

Therefore the data from this record used in the analysis are statistical significance 'p<0.05', from Result 1, and direction 'null', from Result 2.

(ii) Statistical significance and clinical significance

The frequency table for level of statistical significance versus clinical significance category shows that of the 129 records for which statistical significance of one or both results is p<0.05 or better, the majority of these, 93 (72%), are also clinically significant. Eleven (9%) were clearly stated to be not clinically significant, 4 (3%) can be interpreted as 'possibly' clinically significant, and for 21 (16%) clinical significance was not stated. Section 4.3.1 described how an attempt was made to prevent knowledge of the p-values associated with the results reported from influencing the coding of clinical significance.

The correlations between variables representing statistical significance, clinical significance and direction of results were obtained. These are generally low, possibly due to the fact that the results variables are obtained by using the 'best of two' results in each case, leading to a dilution of the correlation between each pair of results variables. The correlation coefficients of greatest magnitude are:

- Direction of result is null and result is not clinically significant $\rho = 0.44$

- Statistical significance (a continuous variable with a high value indicating a small *p*-value) and result is clinically significant $\rho = 0.44$
- Statistical significance is not reported and result is clinically significant (i.e. results clearly favour the experimental arm, or in the case of an equivalence trial the experimental arm is as good as the standard arm) $\rho = - 0.40$
- Statistical significance is not reported and result is clearly not clinically significant (i.e. results favour the standard treatment) $\rho = -0.40$

Subgroups reported

It is important to record whether an article reports the results for subgroups as well as or instead of for the whole group randomized. [Counsell et al (1994), Clarke and Halsey (2001)] The larger the number of subgroups used, the more likely it is that a significant result will be obtained, unless multiple comparisons procedures are used. By chance alone, from twenty subgroups, there is likely to be one with a *p*-value of less than < 0.05. It is important to note that if results are reported by subgroups separately for each variable, this is an additive process. For example if there are three age groups, five white blood count categories and the two sexes, then there is a total of ten subgroups. If, however, results are given for combinations of these, for example, for older males with a high white blood count, the process is multiplicative. In this case there are $3 \times 5 \times 2 = 30$ possible subgroups, and it is likely that at least one will have a result with a *p*-value <0.05 associated with it by chance. Variable SUBGRP has value 1 if one or more subgroup results are reported and 0 if no subgroup results are reported.

Journals/Books/Meeting papers (PUBTYPE)

Of the 257 articles, 195 are from journals, 11 from book chapters and 51 from meeting abstracts. Preliminary graphs confirm that, as expected, randomizations are published as meeting abstracts before being reported as full journal articles. Of the journal articles 82 (42%) give date of submission/receipt and 89 (46%) give date of acceptance. No articles from books or meeting papers

give either date of receipt or date of acceptance. Therefore variable PUBTYPE is not included in any of the 'time from close to submission' or 'time from receipt to publication' analyses.

Impact Factor (IMPACT)

The definition of impact factor is as follows:

"The journal impact factor is a measure of the frequency with which the "average article" in a journal has been cited in a particular year. The purpose of the impact factor is to help the reader to evaluate a journal's relative importance, especially when comparing it to others in the same field. It is calculated by dividing the number of current citations to articles published in the two previous years by the total number of articles published in the two previous years." [3]

The impact factor for any journal changes from year to year, and ideally the value attached to an article should be taken from the edition for the year in which the article was published. However since the year of publication for the articles included in this data ranges from 1965 to 1999, this would not be practical. Also it is likely that the earlier articles pre-date this measure of readership. The 1995 edition impact factor was used irrespective of the year of publication, since this lies within the range of year of publication for the articles included, and it was assumed that any changes in impact factor over time would make little difference to the models.

A summary of the distribution of the variable 'impact factor' for the 257 articles used in this thesis is given in Table 4.4. and an alphabetical list of journals, the number of articles used from each and the impact factor is given as Appendix VI. For journals which have not been assigned an impact factor, and for most meeting papers and all books, the impact factor has been set to zero for the purpose of this project. Where a journal is linked to particular scientific meeting, the abstracts from that meeting are assigned the same impact factor as the journal. There is an additional variable (NOIMPACT) which is set to 1 if there is no impact factor and to 0 if there is an impact factor. This overcomes the problem of missing values.

Impact Factor	Frequency	Cumulative Frequency	Percentage	Cumulative Percentage
No impact factor	79	79	30.7	30.7
0.071-0.754	12	91	4.7	35.4
1.073-1.582	37	128	14.4	49.8
2.095-2.864	48	176	18.7	68.5
3.106-4.549	14	190	5.4	73.9
6.922	24	214	9.3	83.2
8.206-8.569	21	235	8.2	91.4
17.49	12	247	4.7	96.1
22.412	10	257	3.9	100.0

Table 4.4 Distribution of values for journal impact factor
This includes book chapters and meeting abstracts

The articles used in this project are taken from 44 different journals, 8 books and 11 meeting abstract books. Considering journals alone, 17 have an impact factor greater than 2, of which 10 are American, 6 European and for one the country group of publisher is unknown. There are five journals with impact factor greater than 5, of which four are American; *New England Journal of Medicine* (with an impact factor of 22.412), *Blood* (8.569), *Cancer Research* (8.206) and *Journal of Clinical Oncology* (6.922), and one is European (British); *Lancet* (17.490).

Article published in the English language (ENGLISH)

97% of the articles used are in the English language. Of the remaining articles four are in German; publications 33 [Henze (1981)], 34 [Henze (1982)], 35 [Schrappe (1987)] and 256 [Zintl (1992)] and there is one in each of the following languages: Dutch; publication 133 [van der Does-van den Berg (1989)], Spanish; publication 257 [Ortega Aramburu (1985)] and Japanese; publication 164 [Children's Cancer and Leukemia Study Group (1989)]. All seven articles also give the abstract in English. In the absence of a translation, due to financial constraints, it was necessary to use the abstract alone to collect data. However this may be a biased approach. It would be useful to take a sample of abstracts from full papers in English and compare the data from these with those from the English abstracts from the papers in other languages. For example perhaps only striking results

[3] The source is the Institute for Scientific Information: Journal Citation Reports 1995 Science Edition. The 1995 edition was used rather than a more recent version because only articles published before 1 January 2000 have been included in this thesis.

are given in abstracts, so it may appear that foreign papers are more likely to report striking results than those in English. This is a possible topic for future investigation but the sample available in this study (i.e. seven non-English articles) is too small to provide a meaningful basis for such investigation.

5 PRELIMINARY ANALYSES

5.1 Introduction

This chapter starts by stating the twelve analyses to be performed. It then describes the problem of having only a partial date of publication for some records, and hence not knowing for some randomizations which article is the first mention or the first reporting of results. A solution to this is then described, and also the case of '**true ties**', where a randomization is reported in two papers, within the same issue of a journal, and how this is dealt with.

The second part of this chapter deals with the strategy for conducting the preliminary analyses: the use of preliminary graphs, the analysis itself - selecting the model that best fits the data, and the use of diagnostic plots to check that the assumptions of the model are satisfied.

The third part of the chapter deals with problems that arose relating to the preliminary analyses and their solutions, notably (i) choosing the best model when, due to missing values for some of the variables, there are several contenders, and (ii) the introduction of new variables, including interaction terms, at a late stage.

The next part of the chapter summarises the guidelines used in deciding whether a variable, or a particular class of a categorical variable, should be included in an analysis. It also explains how categorical variables are represented. A section on the range of values over which the regressions are valid follows. Finally there is a section about atypical observations, describing one particular observation in detail. This features in both some of the 'time to first' analyses and some of the 'all records' analyses and so is covered in this chapter.

5.2 The twelve analyses to be performed

In order to investigate which trial characteristics affect time to publication, twelve analyses are run. In each case the response variable is one of the following three time periods:

- Time from close to submission (TCLREC)
- Time from receipt to publication (TRECPUB)

- The sum of the above, time from close to publication (TCLPUB)

For each of the three time periods, four analyses are done using:

- All mentions of each randomization
- All reportings of results for each randomization
- The first mention of each randomization
- The first reporting of results for each randomization

For brevity, these analyses will be referred to as 'all mentions', 'all results', 'first mentions' and 'first results', respectively.

The amount of information in a record varies. Some contain detailed results whereas others mention only that a trial was randomized.

The purpose of looking at mentions is that this is of significance to meta-analysts, who need to identify all relevant trials, if possible, for inclusion. Detailed results are not essential for this process. On the other hand, analyses done using only those records reporting results are important since clinicians treating patients will be influenced by the results reported. Table 5.1 shows how much data are available to answer each of the questions.

	All	**TCLREC present**	**TRECPUB present**	**TCLPUB present**
All mentions	610	209	218	582
All results	394	129	137	372
1st mentions	217	63	72	195
1st results	188	52	60	170

Table 5.1 Number of observations available for inclusion in analyses

Initially the plan was to investigate which variables affected time from close to submission (TCLREC) and time from receipt to publication (TRECPUB) and then try only the significant variables from the two analyses for the combined time period 'close to publication' (TCLPUB). This would have cut down the number of full preliminary analyses from 12 to 8. However because

there are so many more data available for analysing 'close to publication' (TCLPUB), it was decided that all 12 regressions should be run from scratch.

5.3 The strategy used for conducting the preliminary analyses

5.3.1 Using the correct dataset

The relevant dataset must be used depending on whether the analysis to be done involves all mentions, all results, first mentions or first results. The main data management program, for which the algorithm is given as Appendix III, contains the various options. For analyses involving all mentions, and all reportings of results, the merged dataset (i.e. the set of publication records but using data from the definitive records where appropriate) can be used without any modification.

However for analyses of the first mention or the first reporting of results for each randomization, a modified dataset must be used. The reason for this is that in order to select the first mention or first result published for a randomization it is necessary to first put the records in chronological order of publication. This involved much more work than was initially thought. To begin with the merged dataset was sorted by randomization, and within randomization, by date of publication.

Incomplete dates

The date of publication is not always given in full. Sometimes only the month and year, or the year alone is known. For one article even the year of publication had to be estimated.

Dates where the year or decade are unknown are considered too inaccurate to be useful in regression analyses. However they can still be used to put articles in chronological order of publication. To have obtained full dates of publication for all 257 articles would have been impractical. Instead all cases where inaccurate dates led to the order of publication of a randomization being unknown were found and accurate dates sought for these only. Of the 610 publication records, there were forty cases where randomizations had an order of publication tying for two records, three cases where the order number tied for three records and one case where the order number tied for four records.

Methods used to obtain more accurate dates are as follows:

- The publishers of 16 journals and books were contacted and asked for an accurate date of publication for each of the articles in question. Alternatively, if the publisher explained the numbering system for journal issues, the publication dates were estimated. Addresses and fax numbers were obtained using the Internet. If both were given, both were used. The telephone was used as a last resort.

- If no reply was received from the publisher, MEDLINE and EMBASE on the Internet were used to find the article. This yielded only four of the 45 outstanding dates. In most cases the date of article given by MEDLINE was the same as that on the hard copy of the article, from which the information was taken originally.

'True ties'

Having obtained the previously missing dates of publication, one problem remained – that of 'true ties'. This is when one issue of a journal reports two or more articles describing the same randomization. The date of publication is necessarily the same for both articles, although the dates of submission and acceptance may differ. A solution to this was to combine the records for each randomization from that issue into a single joint record. There were nine genuine tying pairs and one genuine tying trio. To deal with these SAS was used to create a small dataset of ten records combining the data from the ties. When analysing data using only the first mention, or first results, for each randomization, this small dataset is merged with the full dataset and the 21 records that the 'tie' dataset replaces are deleted.

The 'true ties' are as follows:

- Randomizations 1205 and 1208, CNS prophylaxis randomizations from trials ALL-BFM-81 (1205) and ALL-BFM-83 (1206) are both mentioned in two articles, publications 32 [Riehm et al (1990)] and 47 [Buhrer et al (1990)], which are in the 1990 edition of *Haematology and Blood Transfusion* volume 33. 1206, a maintenance duration

randomization, also from ALL-BFM-83, is mentioned in both of these as well, and in publication 41 [Henze et al (1990b)], again in the same book. 1207, the intensification randomization from the same trial, is mentioned in publications 32 [Riehm et al (1990)] and 41 [Henze et al (1990b)].

- Randomizations 1603 (combination) and 1604 (CNS prophylaxis) from trial CCG-105 (1603) are mentioned in publications 82 [Tubergen et al (1993a)] and 84 [Tubergen et al (1993b)], a pair of consecutive articles, published in *Journal of Clinical Oncology* 1993; volume 11 no. 3 (March) pages 520-526 and 527-537.

- Randomization 1604 is also mentioned in two meeting abstracts for the 26th Annual Meeting of the American Society of Clinical Oncology; publications 78 [Tubergen et al (1990)] and 79 [Gilchrist et al (1990)].

- Randomization 2701 (CNS prophylaxis) from the MRC UKALL I trial (2701) is mentioned in publications 5 [Medical Research Council (1973)] and 4 [Campbell et al (1973)], a pair of consecutive articles in *British Medical Journal* 1973 volume 2, pages 381-384 and 385-388 respectively.

- Finally randomization 2708 (maintenance duration) from the MRC UKALL III trial (2706) is mentioned in publications 16 [Medical Research Council (1982a)] and 15 [Medical Research Council (1982b)], a pair of consecutive articles in *Medical and Pediatric Oncology* volume 10 (5) pages 501-510 and 511-520 respectively.

The only ties for first mention are those for randomizations 2701and 2708. The only tie for first reporting of results is that for 2701. The other ties are for subsequent reportings.

5.3.2 Preliminary graphs

Having selected the appropriate data, preliminary graphs were produced in order to help rule out some of the many explanatory variables that could be included in the analyses. Since many of the

variables have missing values, not all can be included in the analyses, as this would seriously reduce the amount of the available data that could be used. Therefore it is essential that those variables most likely to affect the response are identified.

- Firstly the response variable to be considered was chosen. This is one of the following: time from close to submission, time from receipt to publication, or time from close to publication.

- Then the appropriate dataset was selected from the following: 'all mentions', 'all results', 'first mentions' or 'first results'.

- For the chosen dataset, using only records where the chosen response variable is present, bar charts were produced for all meaningful variables, in order to see the spread of the data and the proportion of observations for which data are missing. If, for example, 98% of the records are one category of a variable, it is unlikely that that variable would have an effect on the response. Similarly, a variable with missing values for the majority of the observations would also be of little importance.

- For each continuous explanatory variable, a scatter plot was produced of the response versus that variable. The purpose of this is to see whether there is a linear relationship between the two, whether a transformation is needed or whether no relationship is indicated. An example of where graphs show that a transformation is indicated is given in Section 4.2.

- Similarly, for each categorical explanatory variable, a box-plot was produced of the response versus each class. The bottom and top edges of the box are the 25^{th} and 75^{th} percentiles and the horizontal line is the median. If the box for one category could fit inside the box for another, it was assumed that the two categories were similar and could be pooled. (Of course, ordered categories could only be pooled if they were consecutive.) This was useful for reducing the number of dummy (indicator) variables needed in the analysis,

and could also be used to rule out variables altogether in some cases. In Figure 5.1 examples using the 'all mentions' dataset indicate that 'whether published in the English language or not' may have an effect on the response variable, 'time from close to publication', whereas 'international, limited international or single-country participation' looks unlikely to have an effect.

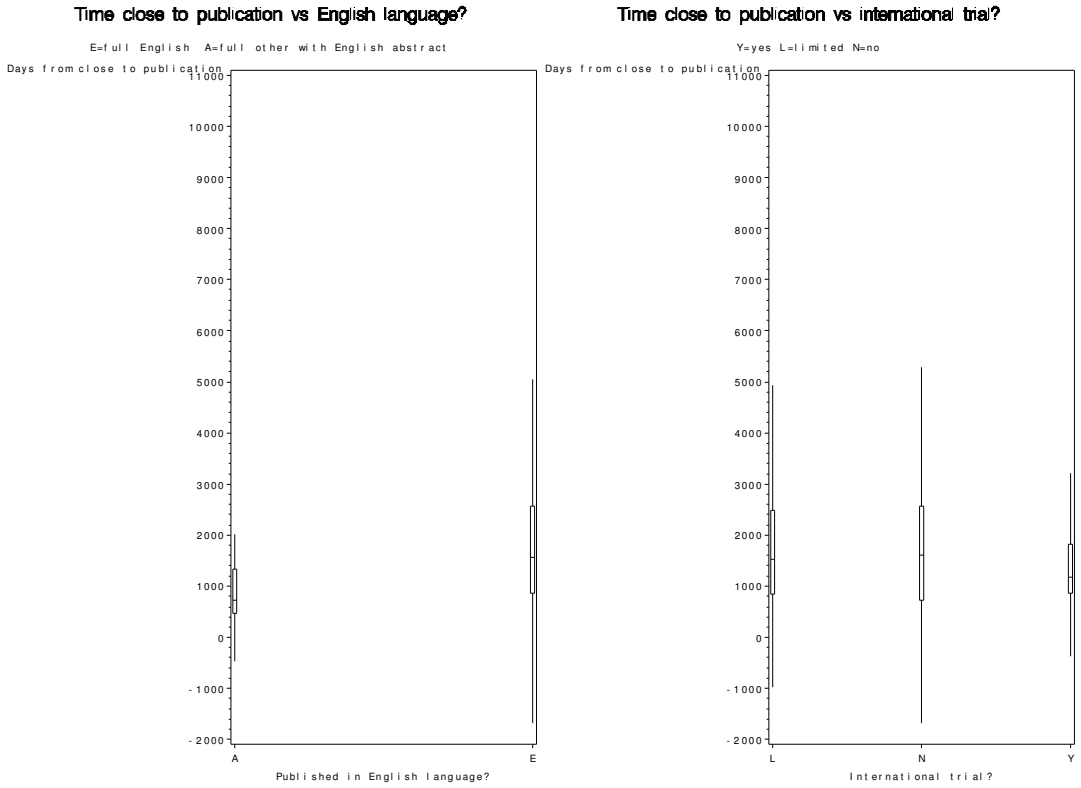


Figure 5.1 Use of box-plots to determine whether classes of a categorical variable can be combined. Examples: time from close to publication vs. whether or not article is published in the English language and vs. whether level of participation is international, international (limited) or single-country.

For each analysis the preliminary graphs and the proportion of missing values were used to help rule out variables for inclusion in the analysis. If a variable had more than 25% missing values, it was excluded.

5.3.3 Analysis

Once the list of possibly informative explanatory variables had been chosen, the preliminary analysis could be done. Multiple linear regression with backward elimination was used to obtain,

from these, the reduced set of significant variables. In this way the effect of each of the possible explanatory variables on the response variable was adjusted for that of the others. The method used is as follows:

- All variables were included in the model to begin with, and at each step the variable with the largest *p*-value was dropped, reducing to a set of variables all with $p<0.05$. Two pairs of variables are very highly correlated. These are number of trial arms (ARMS) and number of questions (NOQ), as discussed in Section 4.2, and start (NSTART) and close (NCLOSE) dates of accrual period. For the number of arms and number of questions the Pearson Correlation coefficient $\rho = 0.94412$ ($p<0.0001$), using the definitive records for all 243 randomizations. For the start and close dates of the accrual period the Pearson Correlation coefficient $\rho = 0.97210$, ($p<0.0001$) using the 219 of the 243 definitive records for randomizations for which both start and close date are present. Only one from each pair was used in the initial regression with backwards elimination, generally ARMS and NSTART. Then, at the stage that one of these was dropped from the model, as the least significant variable, its opposite number was tried. If the first of the pair remained in the reduced set, its opposite number was then tried to see whether this improved the model. If both elements of a pair were included at the same time, their correlation could make them appear less significant than they are. Methods for selecting the model that best fits the data, from several possibilities, are described at the end of this section.

- Many of the variables considered are categorical; some of which have more than two categories. To include a categorical variable with *n* categories in the analysis it is necessary to use (*n*-1) indicator variables.

- The backward elimination process was done 'by hand' rather than using an automated procedure. The reason for this is that if an indicator variable, specifying a category of one of the categorical variables, is the least significant variable, it can only be dropped if it makes sense to pool this category with the default category, not specified by an indicator

variable. Otherwise categories must be pooled in a different way, with reference to the box-plots. The automated method is far quicker but would not necessarily have made sensible pooling choices.

- Once the reduced set of significant variables was obtained, an attempt was made to improve on the fit of the model to the data. Each variable not used because of missing values and each variable thrown out during the elimination process was added in to the reduced set, one at a time. If a variable was dropped with a *p*-value of greater than 0.2 it was not tried again, as it is unlikely that it would be significant. Where a categorical variable with more than two categories was re-tried all indicator variables for it were put back in, so that all levels were represented. As an indicator variable became the least significant variable, categories were re-pooled.

- If two variables were highly correlated (with correlation coefficient greater than 0.4, say) and the first was included in the reduced set of significant variables and the second not, then the first was replaced by the second to see if this improved the fit of the model. If the second variable had more missing values than the first it was not tried.

- Once confident that the best main effects model has been found, interaction terms were tried. Only interactions of order 1 (i.e. involving two main terms only) were considered. If an interaction term is significant, but the *p*-value of one or both of the main terms exceeds 0.05, there are differing opinions as to whether these main terms should be retained in the model (hierarchical modelling) or not. In the analyses reported here they have been kept. For an interaction term to be considered it must be believed that it has an effect on the response variable over and above that of the main effects separately.

- Interactions were only tried if both the main terms were either significant or eliminated with a *p*-value <0.2, since if the main terms were not at all significant, it is unlikely that

their interaction would be. Also interactions were not introduced if they contained many missing values, the same rule applying here as for the inclusion of main terms.

- Where an interaction was found to be significant, the analysis was re-run trying marginally significant variables (the last few to be dropped in the elimination process).

Selecting the model that best fits the data

The large number of missing values in the data, despite best efforts to find these, meant that for each analysis there were several contenders for best-fitting model. Adding in a variable to the reduced set sometimes resulted in one or more of the existing variables becoming no longer significant. This necessitated the task of comparing two or more models with different numbers of explanatory variables and which use different numbers of observations.

There are three methods for doing this:

- Two (or more) models are compared using only those observations which have no missing values for any of the variables in either (any) model. A forward stepwise multiple regression is performed allowing variables from both models to be selected.

- The mean is calculated for each of the variables containing missing values and used where data are missing. A similar forward stepwise multiple regression procedure is then used.

- The problem is solved if both the above methods result in the same model being selected. Otherwise a more complicated imputation method could be tried. This involves estimating each explanatory variable with missing values by regressing on other explanatory variables, which are not in the model, and then using this estimate in the model, in the same way the mean is used.

In practice the first two of these three approaches were used, and this was a sufficient basis on which to select the best-fitting model. The chosen model was then applied to the real data, i.e. using all available observations for the combination of variables used in the model and without

substituting in estimates of missing values. In some cases a variable of borderline significance, but with a *p*-value less than 0.05, when the two methods for comparing models were used, was found to have a *p*-value greater than 0.05 once the real data were used, and so was then dropped.

5.3.4 Checking model assumptions (diagnostic plots)

Once the best-fitting model was selected, the assumptions of the model were checked.

1. The association between the response and the explanatory variables is linear

2. The assumption of constant variance

To check the above assumptions are satisfied, the standardised residuals were plotted against the fitted values. If the variance is constant the points will be scattered evenly above and below the *X*-axis and not form a wedge shape. An uneven spread could indicate that an important variable has been omitted, or that the relationship between the response and the explanatory variables is not linear. Ninety-five percent of observations should lie within two standard deviations. Observations lying outside three standard deviations are considered extreme outliers and have been investigated further, by looking at the studentised residuals (where each residual is divided by its standard error). The last few steps of the regression were re-run excluding the outlier(s) to see whether the significance of the variables changed.

For each regression a plot of Cook's distances (high influence points) against fitted values was obtained. This indicates how much effect any individual observation has on the model, i.e. how much the fitted values would change if that observation was omitted from the analysis. If an observation has an exceptionally large Cook's distance it could be atypical and have too great an influence on the model and should possibly be excluded, if reasonable to do so. The Cook's distances should be considered relatively rather than absolutely. One can appear large compared to the rest, but only because the others are small and an observation with a large Cook's distance need not necessarily be an outlier.

3. The assumption of normality

The normality assumption was checked using a normal probability plot. Here the residuals are plotted against their normal scores. The result is a straight line if the data are normally distributed. Any outliers show up at the extremities.

Examples of the three diagnostic plots used are given for the first of the 'time to first' analyses in Section 6.3.

4. The assumption of independence of observations

The final assumption is that of independence of observations. As was described in Section 3.4, since each randomization belongs to a trial, which may comprise more than one randomization, it is important to consider whether randomizations belonging to the same trial may be correlated and, if so, whether 'trial' should be built into the models. A decision was made not to include 'trial' in the model for the following reasons:

- An investigation was undertaken which indicated that analysing the data at 'trial' level and at 'randomization' level had similar effects on the time-period response variable. From this it was inferred that the correlations between randomizations belonging to the same trial are not of great importance. This is described in Section 8.20.
- The object of interest is the randomization rather than the trial, since it is the randomization that is used in meta-analyses.
- The 243 randomizations in this set of data are from a total of 195 trials. If 'trial' was to be incorporated into the model, this would need to be done using a categorical variable with 194 indicator variables (or as a class variable with 195 categories) which may lead to overparameterisation and hence unreliable findings.

The three regressions using the first mention for each randomization, and the three using the first results, have one record per randomization. For these independence can be assumed. Providing the other assumptions are also satisfied, the regressions performed as described previously are satisfactory. The results of these will be reported in the next chapter.

However the three regressions using all mentions and the three using all reportings of results can have multiple records per randomization. Therefore for these analyses independence of observations cannot be assumed, and this must be adjusted for. In order to do this, a repeated measures analysis must be applied to the data. However the independence model serves as a useful preliminary investigation and the results will be given in Chapter 7.

5.4 The ordinary linear regression model

In order to clarify the notation that will be used to describe the repeated measures model in Section 9.2, an outline of the basic method using ordinary linear regression model, used for the final models where independence can be assumed and for the preliminary analyses, is given below.

The ordinary linear regression, or independence, model is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where $\mathbf{y}$ is the $n \times 1$ column vector of values for the response variable for n observations

$\boldsymbol{\beta}$ is the $(p + 1) \times 1$ column vector of parameters to be estimated (There are p explanatory variables in the model)

$\mathbf{X}$ is the $n \times (p+1)$ matrix containing the values for the explanatory variables

$\boldsymbol{\varepsilon}$ is the $n \times 1$ column vector of values of the residuals for the n observations

It requires that the following assumptions are satisfied:

Y_i has distribution $N(\mu_i, \sigma^2)$, $i = 1 \ldots n$

i.e. the response variable is approximately normally distributed with the mean independent of the variance, and the variance is constant.

Since for a normal distribution, the probability density function (pdf) is

$$f(y / \mu, \sigma^2) = 1/\sqrt{(2\pi\sigma^2)}\ exp\ [-(y-\mu)^2 /2\sigma^2] \qquad -\infty < y < \infty,$$

the pdf for a single observation is

$$f(y_i / \mu_i, \sigma^2) = 1/\sqrt{(2\pi\sigma^2)}\ exp\ [-(y_i-\mu_i)^2 /2\sigma^2]$$

and therefore the joint pdf for n independent observations is the product

$$f(y_1 \ldots y_n / \boldsymbol{\mu}, \sigma^2) = \prod_{i=1}^{n} 1/\sqrt{(2\pi\sigma^2)}\, exp\,[-(y_i - \mu_i)^2 / 2\sigma^2]$$

$$= 1/\sqrt{(2\pi\sigma^2)}^{n}\, exp\,[-\sum_{i=1}^{n} (y_i - \mu_i)^2 / 2\sigma^2]$$

The likelihood function, L, is a measure of how likely values $y_1 \ldots y_n$ are to have come from a distribution based on given values of $\boldsymbol{\mu}$ and σ^2.

$$L(\boldsymbol{\mu}, \sigma^2 / y_1 \ldots y_n) = f(y_1 \ldots y_n / \boldsymbol{\mu}, \sigma^2) = 1/\sqrt{(2\pi\sigma^2)}^{n}\, exp\,[-\sum_{i=1}^{n} \varepsilon_i^2 / 2\sigma^2]$$

To maximize the likelihood is to minimize the sum of squares, $\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \mu_i)^2$.

Replacing $\boldsymbol{\mu}$ by the linear combination of explanatory variables,

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$$

For the i^{th} observation

$$\mu_i = \mathbf{X_i}\boldsymbol{\beta} \qquad \text{(the } i^{th} \text{ row of matrix } \mathbf{X})$$

Therefore it is necessary to minimise $\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \mathbf{X_i}\boldsymbol{\beta})^2$.

The estimating equations

$$\partial [\sum_{i=1}^{n} \varepsilon_i^2] / \partial \beta_j = 0, \; j = 0 \ldots p$$

solve to

$$\hat{\boldsymbol{\beta}} = (\mathbf{X^T X})^{-1} \mathbf{X^T Y}, \qquad \text{where } \mathbf{X^T} \text{ is the transpose of matrix } \mathbf{X}$$

σ^2 is estimated by $s^2 = RSS / [n-(p+1)]$, where RSS is the residual sum of squares. This is unbiased.

5.5 Some problems relating to the preliminary analyses

A summary of the problems encountered and possible solutions:

- Many variables were considered, some of which are highly correlated. Preliminary analyses were used to reduce the number. For each analysis, a correlation matrix for all the possible explanatory variables was produced. To decide between two highly correlated variables the criterion was to keep the most meaningful or that with fewer missing values.

- Many observations have missing values for some of the variables. This reduces the proportion of the dataset that can be analysed and can also result in not knowing which of several possible preliminary models should be used. The number of explanatory variables and number of observations used may vary between models. Solutions to these problems are as follows:

(i) Some missing values could be imputed using common sense, as described in Section 4.3.4.

(ii) Variables with more than 25% missing values were omitted from the regressions. The reason for this is that a variable missing for a very large proportion of observations is unlikely to have much influence on the response.

(iii) Variables with some missing values (but 25% or less) were included in the analysis, but not in the initial regression with backwards elimination. They were tried one by one, once a reduced set of variables, with p-values <0.05, had been found, to see whether the fit of the model could be improved. However this sometimes resulted in several possible models being produced, containing a varying number of explanatory variables and based on a varying number of observations.

(iv) If there is a choice between two models, differing by one variable only, similar with regards to quality of fit, and where the two variables in question are highly correlated, the criterion for inclusion in the model is the variable which is most meaningful, or that with fewer missing values, as previously mentioned.

(v) If a large number of observations are excluded due to missing values for one variable, that variable should probably be omitted. If it is highly correlated with another variable with fewer missing values, the latter would act as a surrogate.

(vi) If there is more than one possible 'best-fitting model', the contenders can be compared using the tests described in Section 5.3.3.

- Some less obvious variables were added at a later stage for the sake of covering all options. Fortunately, these did not have missing values and were tried one-by-one into the best fitting model found so far. These are discussed in Section 5.6.

- Multiple records per randomization, due to several reportings of some randomizations. Repeated measures analysis are used to compensate for this, as will be described in Chapters 8, 9 and 10.

5.6 New variables introduced at a late stage

The following variables, suggested by initial findings, were introduced at a late stage into the analysis. Also it was not feasible to consider further variables until some of the original variables had been dropped. In Chapters 6 and 7, where the results of the ordinary linear regression analyses are reported, the variables considered from the start are referred to as **'initial variables'** and those introduced later as **'second stage variables'**. Since the choice of interactions considered was also made on the basis of initial findings, these too are included in the second group of variables. Second stage variables were added in one-by-one to the best-fitting model obtained using the initial stage variables. Most of the analyses show a marked improvement in the proportion of variability explained by the model (R^2) once the second stage variables were incorporated.

'Not reported' variables and categories for results fields and journal impact factor

At first it appeared that the large number of missing values for statistical significance, clinical significance and direction of results meant that these important variables could not be included in the 'all mentions' and 'first mentions' analyses, using the '25% or less missing values' criterion for inclusion of variables. Later it became clear that 'not reported' is an important category. This was rectified in different ways depending on whether the variable in question is categorical or continuous.

For the categorical variables clinical significance (CLNSG) and direction of results (POSNG) this was resolved by introducing 'X = not reported' as an additional category.

Similarly the variable indicating whether the main question(s) stated in the paper had been answered in that paper (ANSWER) was given an 'X = not reported' category in addition to the yes/no options.

The continuous variable indicating statistical significance (LOGPEST) was set to zero if data were missing. In addition a new dichotomous variable (LOGPNR) was introduced to indicate whether statistical significance was reported or not. This was set to 1 if statistical significance was not reported, and to 0 if it was. Using the two variables in conjunction it is possible to tell which records have non-significant results and which do not report results, even though LOGPEST is zero for both categories.

A similar solution was used for the continuous variable 'impact factor' (IMPACT), which was missing for records from all books and from some journals and meeting abstracts. If no impact factor existed for an article, IMPACT was set to 0. A new variable NOIMPACT was also introduced. This was set to 1 if impact factor was missing and to 0 if present.

'Developing' country (DEVLPNG)

The categories for variable 'country group of trialists' are 'North America', 'Europe' and 'Other'. It became apparent from one of the analyses that this was not sufficient. 'Other' includes India, South Africa, Argentina, Japan, Peru, Brazil, Australasia, Taiwan and Israel. For the purposes of this research, Australasia and Japan are probably more similar to North America or Europe than to some of the other countries in the 'Other' category. Also Poland is the only Eastern European country in the 'Europe' category. For this reason the country group categories were left unchanged, but an additional variable (DEVLPNG) created. This was set to 1 if the country running the trial is considered a 'developing' country and to 0 if considered a 'developed' country. Countries considered to be 'developing' are all those in the 'Other' country group listed above, minus Japan and Australasia, plus Poland.

Number of trials (NTREP) and number of randomizations (NRREP) reported in article

Rather surprisingly, one of the variables found to be significant in several of the preliminary analyses was the number of randomizations (NRAND) belonging to a trial. This could be acting as a surrogate for the number of randomizations (or trials) reported in an article. This led to the creation of two new variables. Preliminary graphs indicated that a logarithmic transformation, LOGNTREP = log_{10}(NTREP) should be tried in addition to NTREP.

Number of co-authors (AUTHORS)

It had been suggested that a large number of co-authors might result in an increased time from close to submission. This was only included at a late stage. The number of authors named in an article range from one to 42. A logarithmic transformation, LOGAUTH = log_{10}(AUTHORS) was used. The reason for this is the hypothesis that the difference in effect on the response time variable of two authors collaborating compared with a sole author is greater than that when comparing, say, eleven authors with ten. Also, where an article names a very large number of authors, for example publication 267, [Tokyo Children's Cancer Study Group, Tsuchida et al (1991)] with 42 authors (the entire membership of the trial working party), it is likely that only a few of those named were actually involved in the writing. This variable was tried in all 12 regressions.

An additional complication is that early papers, which reported Medical Research Council trials, did not name the authors. Authorship was stated as 'Medical Research Council Working Party' and the working party members were listed elsewhere in each of the following; publications 277 [Medical Research Council (1971)], 5 [Medical Research Council (1973)], 1 [Medical Research Council (1975)], 2 [Medical Research Council (1977)], 3 [Medical Research Council (1978)], 16 [Medical Research Council (1982a)] and 15 [Medical Research Council (1982b)]. In some of these articles, members of the writing committee are indicated [Medical Research Council (1975), Medical Research Council (1977), Medical Research Council (1978)] but in others there is no indication of how many people were involved. In those analyses where LOGAUTH was found to be significant, or was eliminated from the regression at $p<0.2$, the regression was re-run, replacing

LOGAUTH with new variable LOGMRC. LOGMRC = log_{10}(AUTHMRC), where AUTHMRC is equal to the number of authors (AUTHORS) where given, but set to 30, a typical size of an Medical Research Council working party, where not given. The results of the two regressions were then compared.

A related variable, which was not considered, is the number of affiliations of the named authors. This is often stated on the paper. However, since it had not been used to compile references, it was not collected for this project. Also there is the added problem of not knowing whether 'the same affiliation' means 'working in the same building', since it is the latter which is the variable of interest. For example, working party members listed as belonging to the Medical Research Council are located around the UK.

Another possible variable to consider might be whether an article mentions working party collaboration or not. Again this has not been collected, and would require a further check of all the original articles.

5.7 Interaction terms

Interactions were only considered if the two main terms (or categories of terms, in the case of categorical variables) were either significant or eliminated with a *p*-value of <0.2.

If one of the main effects is categorical (with *n* classes) and the other continuous, the interaction is expressed by the (*n*-1) indicator variables which describe the categorical variable each multiplied by the continuous variable. Interactions of this type considered are:

- Conducted by trialists in North America and published in a journal with a high impact factor
- Multi-centre trial and results with smaller *p*-values
- International trial and results with smaller *p*-values
- Large trial and clinically significant
- International trial published in a journal with a high impact factor

- Large trial with a positive result
- Equivalence trial and results with larger *p*-values
- North American trial and results with smaller *p*-values
- North American trial and longer duration of randomization period. (This was tried because, surprisingly, the number of randomizations was found to be significant in several regressions, and some US trials were segmented into many randomizations due to arms being dropped and replaced throughout the course of the trial.)
- Large North American trial

For an interaction between two continuous variables, one is kept as continuous and the other converted into a categorical variable. For example the continuous variable LOGSIZE, log_{10} (number of patients), was converted into a categorical variable with two classes i.e. either 'greater than' or 'less than or equal to' the median number of patients (median=146). Similarly, the continuous variable expressing statistical significance, the distance of log_e (typical value of *p*) for each category from that for the not significant/not reported category, LOGPEST, was replaced by a categorical variable with three classes, highly significant ($p<0.0001$), significant ($p<0.05$ but >0.0001) and non-significant ($p>0.05$). An interaction dealt with in this way is:

- Large trial and results with smaller *p*-values

This was tried in two ways, taking size of trial as the continuous variable and statistical significance as the categorical, and vice versa.

In the case of an interaction between two categorical variables, the interaction is the specified criteria versus all other combinations. Interactions of this type are:

- Conducted by North American trialists and published by a North American journal
- Multi-centre trial with a clinically significant result
- International trial with a clinically significant result
- Multi-centre trial with a positive result
- International trial with a positive result

- Equivalence trial with a clinically significant result
- Equivalence trial with either a null or a negative result
- North American trial with a clinically significant result
- North American trial with a positive result
- Multi-centre North American trial
- International North American trial

To begin with the choice of interaction between categorical variables was made with the intention of investigating the category most likely to result in a shorter time to publication. However, later it was decided also to look at the opposite category, that which may result in a delay to publication. This yielded some interesting results. This latter group of interactions considered included:

- Small trial with results with larger *p*-values
- Single-centre trial with results with larger *p*-values
- Single-country trial with results with larger *p*-values
- Small trial with results not clinically significant
- Single-centre trial with results not clinically significant
- Single-country trial with results not clinically significant
- Small trial with direction of results other than positive
- Single-centre trial with direction of results other than positive
- Single-country trial with direction of results other than positive
- Trial conducted outside North America and Europe with results with larger *p*-values
- Trial conducted outside North America and Europe with results not clinically significant
- Trial conducted outside North America and Europe with direction of results other than positive
- Small trial conducted outside North America and Europe
- Single-centre trial conducted outside North America and Europe
- Single-country trial conducted outside North America and Europe

Since there are many interactions that could be considered, it was decided, in the case of an interaction between two categorical variables, each with several categories, to try only the interactions using those categories where the significance of the main term strongly suggested an effect was likely. For example, if for direction of results, the response variable did not differ significantly between positive and null results,' non-positive results' would not have been tried as an interaction with another variable.

5.8 Criteria for including variables, or categories of variables, in the analyses

For all twelve 'How long?' analyses, a variable was included only if there were 25% or less observations with missing values.

For categorical variables, preliminary graphs (box-plots) were also used to decide which variables might have an effect on the response, and which categories could be pooled from the start. The categorical variables were represented by indicator (or dummy) variables, one less than the number of categories for each variable. The first three letters of the name of each dummy variable are DUM. The last part of the name is descriptive. Generally, variable names are consistent between analyses. However this is not so for the dummy variables. For example variable CGROUP is the country group of the trialists, and has categories North America (A), Europe (E) and Other (O). For one analysis, preliminary graphs may indicate that North American randomizations should be compared with those from European and other countries combined. In this case DUMCGRP is set to 1 if CGROUP= E or O and to 0 if CGROUP= A. However another analysis may involve comparing randomizations conducted in other countries with those from North America and Europe combined. In this case DUMCGRP is set to 1 if CGROUP=O and to 0 if CGROUP = A or E. If all three country groups need to be represented separately then two dummy variables DUMCGRP1 and DUMCGRP2 are used. The name and meaning of each dummy variable is given at the start of the output for each regression in Appendices VIII and IX.

For continuous variables, preliminary graphs (scatter plots) were produced. Provided that there were 25% or less missing values, all continuous variables were included in the initial regression. There were three reasons for this: firstly, it is more difficult to assess whether a continuous (as opposed to a categorical) variable will affect the response merely by looking at the graph. Secondly there is no extra work involved in order to try continuous variables in the regression (unlike with categorical variables where indicator variables must be created). Thirdly there are relatively few continuous variables. Scatter plots were also used to decide whether any of the continuous variables required the use of a transformation.

In the initial stage, variables with no, or few, missing values, and which might affect the response were included in the initial regression using a backwards elimination procedure. Once a reduced set of significant variables was obtained, the variables with missing values, but not more than 25% missing, were added in one-by-one to see if a better fitting model could be found.

The second stage variables, none of which have any missing values, were then tried in turn. Finally, those interactions thought possibly to have an effect were also tried.

Appendix VII lists all variables tried in each of the twelve 'How long?' analyses in tabular form. The following details are given:

- The stage in the analysis at which each variable was tried
- The proportion of missing values for those variables for which this is of note
- For categorical variables, the pooling choices made
- Box-plots were used to decide which categorical variables to use and how to pool classes. In most cases the indication was clear, but cases where the decision to include a particular variable or pooling was borderline, are also shown in the table
- For interaction terms tried, the combination selected versus all others is specified

The alphabetical list of variables and their meanings given in Appendix IV is useful for interpreting the table.

5.9 Range and mean of the response variable for each analysis

For each analysis, the mean and the range of the response variable is given. Since open randomizations can be reported, for example by reference to a subsequent trial starting, in the discussion section of an article, it is possible that the minimum time from close of randomization to submission or minimum time from close of randomization to publication may be negative. This may be exacerbated (or lessened) by the fact that some of the dates of both publication and close of randomization are inaccurate, i.e. only the year is known and, for analysis purposes, 30 June has been used.

5.10 Typical change in the response variable caused by each explanatory variable

The typical change in the response variable caused by each explanatory variable is given for the final choice of model for each analysis of a 'time to' response variable. These are reported in Chapter 6 for the analyses of first mentions and first resportings of results, and in Chapter 11 for the analyses of all mentions and all reportings of results. For each variable the typical change was obtained using the coefficient (β estimate) given in Appendix VIII in the case of analyses of first mentions and first reportings of results, and in Appendices XV and XVIII for the analyses of all mentions and all reportings of results.

The following method was used to calculate the typical effects:

- In the case of an indicator variable specifying the effect of a class of a categorical variable, the typical effect (in days) is given by the coefficient from the appropriate analysis
- In the case of a continuous variable an estimate is calculated by multiplying its standard deviation as given in Appendix V by the coefficient.

From Appendix V it can be seen that time from close to publication (mean 4 years 11 months) is largely made up from the time from close to submission (mean 5 years 3 months), with time from receipt to publication making a very small contribution (mean 10 months). The reason that the

mean time from close to submission appears longer than the mean time from close to publication is that both date of close of randomization and date of submission of article are known for only 209 publication records, whereas both date of close of randomization and date of publication are known for 582 publication records, and so latter is more representative of the whole set of reportings. Since the time from receipt to publication is far shorter than the other two time intervals, the typical effect any variable this is much smaller.

5.11 Atypical observations

From the diagnostic plots an observation is considered atypical if one or both of the following apply:

- It is an extreme outlier i.e. the magnitude of its studentised residual is greater than 3.
- It has a much larger Cook's distance than that of most other records

Some articles generate records which are all atypical. In one case (record 967) an individual record is atypical although all other records from the same publication are not. Certain outlying publications and individual records appear in several of the analyses.

Record 967 is an important outlier in two of the 'time to first' analyses and also in two of the 'repeated measures' analyses, and so is described here. Other atypical publications and individual records apply to either one or the other and so are discussed at the beginning of the chapter reporting their results. The purpose of this is to show patterns common to several regressions and to allow the publication ID and/or record ID to be used without further explanation when reporting the results.

For all atypical articles and individual records, the following checks were made for typographical errors:

- The dates of submission/receipt, acceptance and publication held in the computer were checked against the original article.

- The date of close of randomization (and date of start) was checked against the information held in the ACCESS database for all randomizations which had one or more atypical records. The reason for this is that the date used in the analysis was taken from the definitive record for each randomization, which was compiled using information from all publication records for that randomization.

No typographical errors were found.

As described Section 3.13, if an observation was found to be atypical, (i.e. an outlier with studentised residual of magnitude greater than or equal to 3 and/or a large Cook's distance relative to those of other observations), any notes codes attached to that observation were examined. If other non-outlying observations were also found to have that notes code, the notes code could be ignored safely. Otherwise it was investigated further. For all twelve 'How long?' analyses the notes codes attached to atypical observations were also attached to others and so could be ignored.

Record 967

Record 967 represents the data on trial 2199's sole randomization 2199, reported in publication 284 [Sackmann-Muriel et al (1998)], and is an outlier in many analyses.

It appeared as an outlier in two of the 'time to first' analyses, 'time from close to submission' and 'close to publication' for 'first mentions'. Publication 284 generated seven records in total, of which records 974, 973 and 972 were also used in these two analyses and not found to be atypical. Record 967 was not used in any of the other four 'time to first' analyses because it has a missing value for one of the variables in the final model for 'time from receipt to publication for first mentions', country group of publisher (JGROUP), and it does not contain results and so could not be used in any of the three 'first results' analyses.

Similarly it is also an outlier in the 'time from close to submission' and 'time from close to publication for all mentions' analyses. All seven records from publication 284 were used in these analyses, and again none, apart from 967, were atypical, with the exception of 955 in the former.

Record 967 could not be included in the other four preliminary analyses for the 'repeated measures' data for exactly the same reasons as for the 'time to first' analyses; that JGROUP is missing for this record, and this variable is in the model for 'time from receipt to publication for all mentions', and that the record contains no results and so is excluded from the three 'all results' analyses.

Randomization 2199 was not found during the routine identification process conducted by the Cancer Overviews Group. The GATLA Protocol 11-ALL-67 is an early Argentinian trial, open between November 1967 and September 1970, and does not appear to have been published until 1998. It was received by the journal on 21 September 1996. It accrued only 38 patients and compares two maintenance treatments: Methotrexate given at doses 30mg/m^2 versus 15mg/m^2, twice weekly, with both arms receiving 6-mercaptopurine + vincristine + prednisolone pulses every six months. The duration of treatment was five years. Other studies conducted by this group at around the same time were not randomized and are not included in this thesis. The time from close to submission is 26 years 2 months, and from close to publication is 27 years 9 months.

This observation has been included in the analyses. It is certainly atypical, lying well into the tail of the normal distribution, but since there is no reason to believe that the randomization procedure used in the trial is invalid, the randomization would be eligible for inclusion in meta-analyses. The regressions for which it is an outlier have also been re-run excluding it. Where its exclusion would have led to a change in the conclusions this is discussed in the appropriate section.

It is worth noting the enormity of the typical effect of variable DEVLPNG (trial conducted in a developing country), in the reportings of the analyses that follow. From Appendices VIII, XV and XVIII it can be seen this variable has a large coefficient, which has a large standard error. Also only 15 of the 243 randomizations were conducted in developing countries. While undoubtably this is an important explanatory variable, its typical effect is likely to have been over-emphasised by the inclusion of record 967. To a lesser degree variable CGROUP (country group of trialists) is affected similarly. This should be borne in mind when interpreting the findings.

6 FINAL RESULTS OF THE SIX 'TIME-TO-FIRST' ANALYSES

6.1 Introduction

This chapter begins by describing two outlying observations from the analysis of time from close to publication for first mentions. The results from the six 'time to first' analyses are then reported. These use the three time period response variables: 'time from close to submission', 'time from receipt to publication', and the sum of these, 'time from close to publication', firstly for the first mentions and then for the first results for each randomization. Since all the assumptions of the linear regression model, including that of independence of observations, are satisfied, these are the final results. Details of the variables tried in these analyses are given in Appendix VII and the output for the final choice of model in each case is given in Appendix VIII.

6.2 Outlying articles and records

The atypicality of the following observations can probably be safely disregarded.

Publication 135

Publication 135 [Otten et al (1988b)], generated two records, 696 and 697, for the two randomizations 1901 and 1902 belonging to trial 1901, EORTC 58741. This is the sole reporting of this trial. Both records appeared as outliers in the 'time to publication for first results' analysis (with standardised residuals of approximately 3.3 and 3.4 respectively) and the second and third largest Cook's distances respectively. The randomization was open between May 1971 and January 1979. It was published as an abstract for the *Medical and Pediatric Hematology (SIOP)* meeting and was referred to in the title as 'a long term evaluation'. For a first publication and in abstract form the time from close to publication is unusually long at 9 years 8 months.

Record 716

Publication 148 [Sackmann-Muriel et al (1978)] generated three records, 716, 717 and 718, relating to randomizations 2101, 2102 and 2103 belonging to trial 2101, GATLA 72, open between October

1972 and December 1975. The first two are used in the 'time from close to publication for first results' analysis, with the first having the largest Cook's distance. It was not an outlier but had a low studentised residual (-2.3). The article was published in October 1978 in *Cancer* which has impact factor 2.864. The results were null, and neither statistically nor clinically significant. Time from close to publication is short, 2 years 10 months.

6.3 For the first mention of each randomization, which trial characteristics affect the time from close to submission?

The best fitting model using the initial set of variables

Time from close to submission is longer for randomizations with the following characteristics:

- Shorter duration of accrual period (DURRAN) ($p<0.0001$)
- Conducted outside North America (CGROUP) ($p<0.0001$)
- Trial comprises fewer randomizations (NRAND) ($p=0.0032$)

($R^2=0.401878$, F- statistic=13.21, p-value<0.0001, based on all 63 observations)

Comments on diagnostic plots

This analysis resulted in one extreme outlier; record 967, which is described in Section 5.10. It had by far the largest Cook's distance and a studentised residual of 4.28.

This was removed and the regression re-run. This gave a similar result; $R^2=0.398250$, $F=12.80$, p-value<0.0001 using 62 observations. The three variables remained in the model each acting in the same direction, the significance of the first two remaining at <0.0001, and that of NRAND dropping to 0.0128.

The three diagnostic plots are given in Figures 6.1, 6.2 and 6.3.

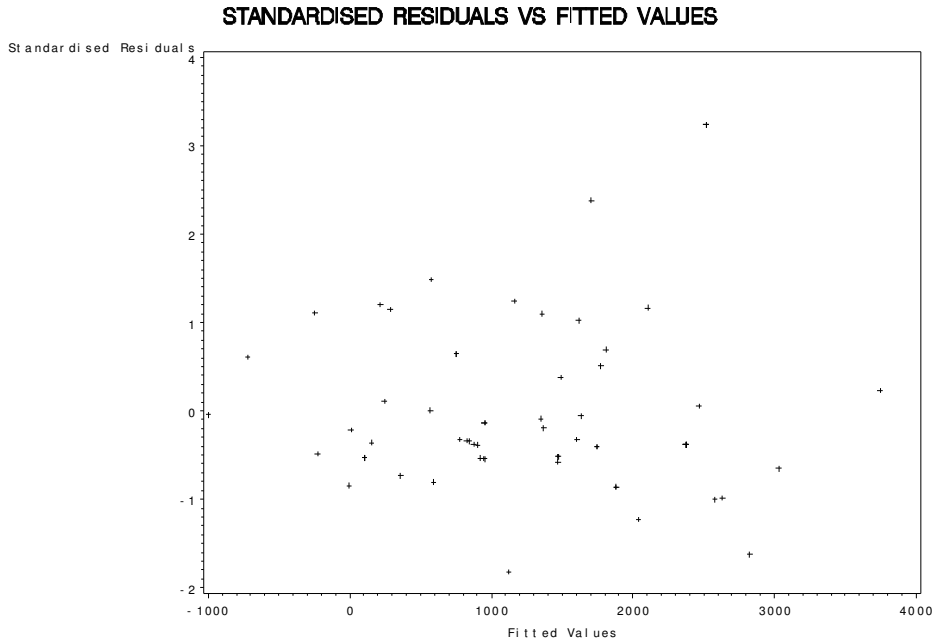


Figure 6.1 Linear regression model using the initial set of variables for time from close to submission for first mentions. Diagnostic plot to check model assumption of constant variance of response variable: standardised residuals vs. fitted values

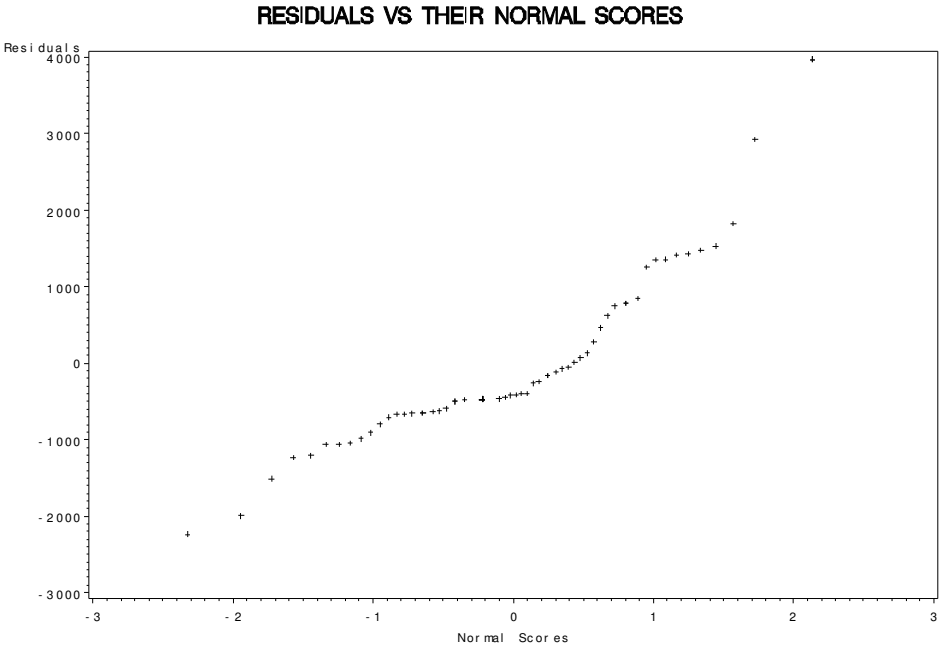


Figure 6.2 Linear regression model using the initial set of variables for time from close to submission for first mentions. Diagnostic plot to check model assumption of normality of response variable: residuals vs. their normal scores

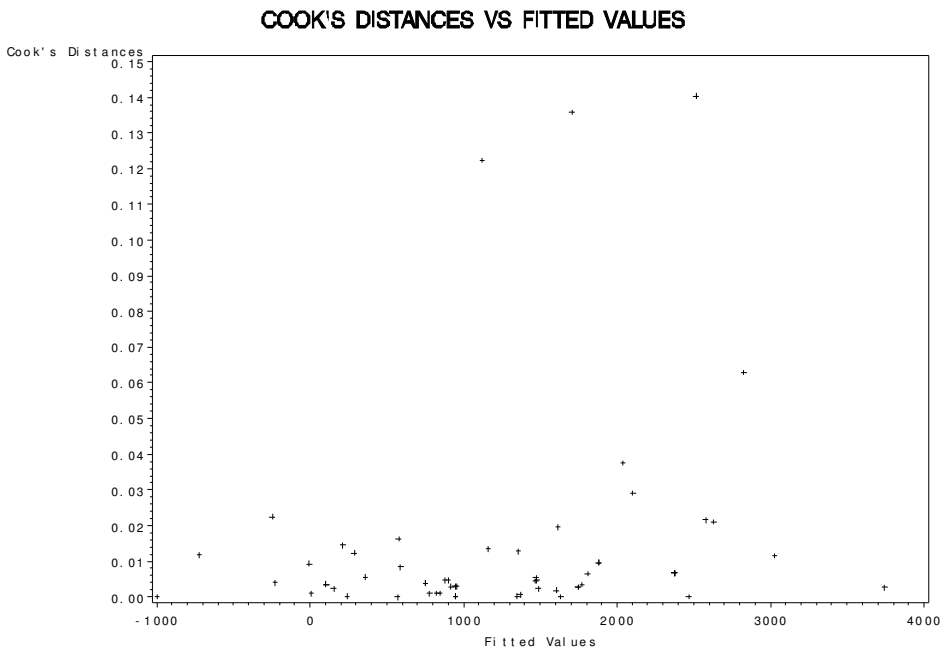


Figure 6.3 Linear regression model using the initial set of variables for time from close to submission for first mentions. Diagnostic plot to highlight atypical observations: Cook's distances vs. fitted values

There were now two new outliers from trial 2198 reported in publication 284; record 972 (randomization 2198) and record 973 (randomization 2197), both had studentised residuals equal to 3.32 and large Cook's distances. These, like 967, are newly found randomizations. The third largest studentised residual (2.49) is record 974 (randomization 2104) also from publication 284. Record 908 (trial 30501, which has sole randomization 30501 and is reported in publication 269) also had a large Cook's distance and a fairly high but negative studentised residual (-1.94). In order of magnitude the four largest Cook's distances are for records 972 and 973 (superimposed), 974 and 908. The table in Appendix I provides further details of these randomizations.

All three records belong to the 'Other' category for country group of trialists. The first two randomizations were conducted in Argentina, a 'developing' country, and the third in Australasia, 'developed'. Also residuals were large and positive for the first two and fairly large and negative for the third, indicating that the third record should not be grouped with the first two. This led to the introduction of a new variable to distinguish between randomizations conducted by trialists in 'developing' countries and those in 'developed' countries (DEVLPNG). This new variable was subsequently tried in other analyses.

Comments on the final model

The only new significant variable to be added to the model once the second stage variables were included is whether the randomization was conducted in a 'developing' or a 'developed' country (DEVLPNG). When this, the country group of trialist (DUMCGRP) and the interaction term for North American trials published in North American journals were all in the model, the vector of interactions was almost identical to that of one of the main effects, indicating that the interaction term practically reproduces one of the columns in the design matrix.

No outliers remained in the final model, although observation 967, which has previously been described, has a large Cook's distance. When removed the model remained unchanged (F=36.21 on 62/63 observations). It was decided that this record should be left in. Although there is not a suitable formal test to compare the model developed using only the initial set of variables and the final model, it is worth noting that the value of R^2, the measure of how much variability in the response variable, 'time from close to submission', is explained by the model, has increased substantially from 0.40 to 0.77.

Final results

Time from close to submission is longer for randomizations with the following characteristics:

- Shorter duration of accrual period (DURRAN) (p<0.0001), typical effect 2 years 7 months
- Conducted outside North America (CGROUP) (p<0.0001), typical effect 3 years 8 months
- Conducted in a 'developing' country (DEVLPNG) (p<0.0001), typical effect 13 years 8 months
- Trial comprises fewer randomizations (NRAND) (p=0.0125), typical effect 11 months

(R^2=0.774776, F- statistic=49.88, p-value<0.0001 based on all 63 observations)

Mean and range of response variable: 1412 (–1113 to 9503) days, i.e. approximately 3 years 10 months (–3 years 1 month to 26 years 0 months).

6.4 For the first mention of each randomization, which trial characteristics affect the time from receipt to publication?

The best fitting model using the initial set of variables

Time from receipt to publication is longer for randomizations with the following characteristics:

- Trial comprises fewer randomizations (NRAND) (p<0.0001)
- Published in a North American or European journal (JGROUP) (p=0.0002)

(R^2=0.400376, F-statistic=21.37, p-value<0.0001, based on 67 of the 72 observations)

Comments on the final model

Before trying the interactions between country group and duration of accrual period, and between country group and whether there was international participation or not, the following model was obtained:

Time from receipt to publication is longer for randomizations with the following characteristics:

- Trial comprises fewer randomizations (NRAND) (p<0.0001)
- Published in a North American or European journal (JGROUP) (p<0.0001)
- Reported in articles which mention fewer randomizations (NRREP) (p=0.0002)
- No clear indication given of whether results are clinically significant or not (DUMCLN) (p=0.0328)

(R^2=0.530205, F-statistic=17.49, p-value<0.0001 based on 67 out of 72 observations (93%))

There were no outliers or observations with large Cook's distances. The observations omitted are due to missing values for publisher country group (JGROUP).

Whether considering main effects alone, or including interactions, this has been the most difficult analysis to model, since missing values have caused instability. Introduction of the interactions between country group and duration of accrual period, and between country group and whether there was international participation or not resulted in the creation of four other possible models. These are all based on smaller numbers of observations. Given that the dataset is small, the response variable only present for 72 observations, and the models unstable, conservation of

observations is important. The analyses using the other four models are based on 59 observations (82%) (three) and 54 (75%) (one).

The model selected was the only one which added terms but did not result in any of the four variables NRAND, JGROUP, NRREP or CLNSG being dropped. The additional main effects CGROUP and DURRAN are non-significant (p=0.0710 and p=0.6869 respectively) but are retained because of the moderate significance of the interaction between country group and duration of accrual period (p=0.0263). It is unclear as to whether the model chosen is the most appropriate, but it is the cautious choice.

Analysis of diagnostic plots indicates no outliers or observations with large Cook's distances. The additional observations omitted are due to missing values for duration of accrual period (DURRAN).

Final results

Time from receipt to publication is longer for randomizations with the following characteristics:

Main effects:

- Trial comprises fewer randomizations (NRAND) (p=0.0135), typical effect 1 month
- Published in a North American or European journal (JGROUP) (p<0.0001), typical effect 8 months
- Fewer randomizations reported in each article (NRREP) (p=0.0010), typical effect 1 month
- No clear indication given of whether results are clinically significant or not (DUMCLN) (p=0.0175), typical effect 2 months

Interaction:

- For trials conducted outside North America, longer duration of accrual period (CGROUP*DURRAN) (p=0.0263), typical effect 2 months

[In addition the typical effects of the non-significant main effects CGROUP and **shorter** duration of randomization period DURRAN are 4 months and 1 week respectively.]

(R^2=0.610919, F-statistic=11.44, p-value<0.0001 based on 59 out of 72 observations (82%))

Mean and range of response variable: 231 (18 to 497) days, i.e. approximately 8 months (1 month to 1 year 4 months).

6.5 For the first mention of each randomization, which trial characteristics affect the time from close to publication?

The best fitting model using the initial set of variables

Time from close to publication is longer for randomizations with the following characteristics:

- Earlier randomizations (NSTART) ($p<0.0001$)
- Shorter duration of randomization period (DURRAN) ($p=0.0042$)
- Treatment type is immunotherapy as opposed to radiotherapy or chemotherapy (TXCHEMO) ($p=0.0352$)
- Trials conducted outside North America and Europe (CGROUP) ($p<0.0001$)
- Randomization has not been presented at a meeting (PRESENTED) ($p=0.0005$)
- Published in a journal with a low or no impact factor (IMPACT) ($p=0.0352$)
- Main question stated in article is clearly either answered or not answered in the article, as opposed to when this is not reported (ANSWER) ($p=0.0137$)
- Not an equivalence trial (EQUIV) ($p=0.0287$)

($R^2=0.405110$, F-statistic=15.66, p-value<0.0001 based on 193 out of 195 observations)

Comments

A preliminary graph indicated that publication type (PUBTYPE) i.e. whether the article was published in a book, journal or meeting abstract may be important. Book publication appeared to take longer. However, this variable was abandoned when it became clear that there was only one book publication in the dataset used. The spread of data for some of the other variables used is also uneven, but all categories of all other variables contain at least five observations. To show how much more extreme the problem with publication type is than with other variables, the categorical variables with the most uneven spread of data in this analysis are shown here:

Publication type (PUBTYPE):

154 (79%) journals, 1 book (0.5%) 40 abstracts (20.5%)

Treatment type (TXCHEMO):

5 (2.6%) immunotherapy, 188 (96.4%) other categories, 2 (1%) missing

Target number of centres reached? (CTARGET):

7 (3.6%) no, 152 (78%) yes, 36 (18.4%) missing

Published in English language? (ENGLISH):

185 (94.9%) yes, 10 (5.1%) no

First-line treatment? (FIRSTL):

170 (87.2%) first-line, 17 (8.7%) relapse/refractory, 8 (4.1%) missing

Table 4.3 lists all variables for which one category (excluding the missing values category) contains less than 10% of the total, when all 610 records are included.

Comments on the final model

Incorporating the second stage variables resulted in treatment type (TXCHEMO) (immunotherapy versus all other categories) being dropped (p=0.2719) and two new variables entering the model; trial conducted by 'developing'/'developed' country (DEVLPNG) and number of trials reported in article (NTREP). Note that TXCHEMO has a correlation coefficient >0.3 with both country group of trialists (CGROUP) and DEVLPNG. None of the interaction terms tried was significant.

There was one extreme outlier (record 967 see Section 5.9). The studentised residual for this observation was >4 and it had by far the largest Cook's distance. However, when the regression was re-run omitting this observation, the model remained unchanged, i.e. all the variables in it remained significant. (F=15.57 on 194/195 observations). Therefore it was decided to leave this record in the analysis. There were also two borderline outliers; record 533, the only record of the nine from publication 57 included in this analysis, with studentised residual 3.02, and record 564, one of a pair from publication 72, the other not used in this analysis, with studentised residual 3.09.

Final results

Time from close to publication of article is longer for randomizations with the following characteristics:

- Earlier randomizations (NSTART) (p<0.0001), typical effect 1 year 3 months
- Shorter duration of accrual period (DURRAN) (p=0.0083), typical effect 7 months
- Conducted outside North America and Europe (CGROUP) (p=0.0434), typical effect 1 year 9 months
- Not presented at a meeting (PRESENTD) (p=0.0003), typical effect 1 year 9 months
- Reported in a journal with a low or no impact factor (IMPACT) (p=0.0089), typical effect 8 months
- A clear indication is given as to whether the main questions in the paper are answered in that paper (ANSWER) (p=0.0002), typical effect 1 year 9 months
- Not an equivalence trial (EQUIV) (p=0.0094), typical effect 1 year 5 months
- Conducted in a 'developing' country (DEVLPNG) (p<0.0001), typical effect 5 years 3 months
- Reported in articles which mention a greater number of trials (NTREP) (p<0.0001), typical effect 8 months

(R^2=0.489958, F-statistic=19.75, p-value<0.0001 based on all 195 observations)

Mean and range of response variable: 1262 (-1679 to 10150) days, i.e. approximately 3 years 5 months (-4 years 7 months to 27 years 9 months).

6.6 For the first reporting of results for each randomization, which trial characteristics affect the time from close to submission?

The best fitting model using the initial set of variables

Time from close to submission is longer for randomizations with the following characteristics:

- Earlier close date (NCLOSE) (p=0.0116)

- Multi-centre participation (MULTIC) (p=0.0065)
- Conducted in a European country (CGROUP) (p=0.0207)
- Conducted in a country other than North America or Europe (CGROUP) (p=0.0063), the latter having a greater effect than the former.

(R^2=0.298416, F statistic=5.00, p-value=0.0019 based on all 52 observations)

Comments on final model

In the final analysis no interactions were significant.

Diagnostic plots showed no outliers and no observations with large Cook's distances.

Final results

Time from close to submission is longer for randomizations with the following characteristics:

- Earlier close date (NCLOSE) (p=0.0036), typical effect 1 year 4 months
- Multi-centre participation (MULTIC) (p=0.0063), typical effect 1 year 11 months
- Conducted in a European country (CGROUP) (p=0.0008), typical effect 3 years 6 months
- Conducted in a country outside North America and Europe (CGROUP) (p=0.0007), typical effect 5 years 5 months

(A clearer interpretation of the previous two results is that results of North American trials are submitted fastest, followed by those from European trials, with those from trials conducted elsewhere having the longest time from close to submission).

- Published in a journal with a high impact factor (IMPACT) (p=0.0242), typical effect 1 year 10 months
- Direction of result is not negative (POSNG) (p=0.0147), typical effect 2 years 6 months
- Direction of result is not positive or null POSNG (p=0.0107), typical effect 2 years 2 months

(A clearer interpretation of the latter two results is that randomizations with negative results are submitted more quickly than those with positive or null results, with those with opposite results or where the direction of results is not reported taking the longest time.)

(R^2=0.460099, F-statistic=5.36, p-value=0.0002 based on all 52 observations)

Mean and range of response variable: 1294 (-191 to 4026) days, i.e. approximately 3 years 5 months (-6 months to 11 years 0 months).

6.7 For the first reporting of results for each randomization, which trial characteristics affect the time from receipt to publication?

The best fitting model using the initial set of variables

Time from receipt to publication is longer for randomizations with the following characteristics:

- Trials comprises fewer randomizations (NRAND) (p=0.0030)
- Funded by charity as well as Government money (FUNDG) (p=0.0153)
- An indication of whether the paper is answering the main trial questions as specified in the paper, or not, as opposed to not reporting this (ANSWER) (p=0.0337)
- Treatment type is immunotherapy or chemotherapy rather than radiotherapy (TXCHEMO) (p=0.0449)

(R^2=0.327469, F-statistic=5.36, p-value=0.0013 based on 49 of the 60 observations)

Comments on the final model

The effects of number of trials and number of randomizations mentioned in an article (NTREP and NRREP) are in opposite directions.

There were no outliers or observations with large Cook's distances.

Final results

Time from receipt to publication is longer for randomizations with the following characteristics:

- Funded by charity as well as Government money (FUNDG) (p=0.0003), typical effect 3 months
- Reported in an article which mentions a **larger** number of trials (NTREP) (p=0.0001), typical effect 1 month

- Reported in an article which mentions a **smaller** number of randomizations (NRREP) (p<0.0001), typical effect 2 months
- Has been presented at a major meeting (PRESENTD) (p=0.0032), typical effect 2 months
- Results not clearly reported as clinically significant (CLNSG) (p=0.0056), typical effect 3 months
- Results not clearly reported as **not** clinically significant (CLNSG) (p=0.0040), typical effect 3 months,

  the effect of the latter being only marginally greater than that of the former.

(A clearer interpretation of the latter two results is that results clearly not clinically significant are published fastest, closely followed by those which clearly are clinically significant. Where clinical significance is not reported, the time from receipt to publication is longest.)

(R^2=0.583463, F-statistic=9.81, p-value<0.0001 based on 49 out of 60 observations)

Mean and range of response variable: 232 (18 to 421) days, i.e. approximately 8 months (1 month to 1 year 2 months).

6.8 For the first reporting of results for each randomization, which trial characteristics affect the time from close to publication?

The best fitting model using the initial set of variables

Time from close to publication is longer for randomizations with the following characteristics:

- Earlier randomizations (NSTART) (p=0.0007)
- Shorter duration of randomization period (DURRAN) (p<0.0001)
- Treatment type is immunotherapy as opposed to radiotherapy, chemotherapy or antibiotic (TXCHEMO) (p=0.0498)
- Conducted outside North America and Europe (CGROUP) (p=0.0083)

(R^2=0.221724, F-statistic=11.68, p-value<0.0001 based on 169 out of 170 observations)

Comments

The one missing value is for treatment type (TXCHEMO), a variable of borderline significance.

Comments on the final model

Incorporating the second stage variables resulted in treatment type (TXCHEMO) which was of only borderline significance, being dropped (p=0.0651). New significant main effects are impact factor (IMPACT), whether the journal had an impact value associated with it (NOIMPACT), clinical significance (CLNSG), statistical significance (LOGPEST) and direction of results (POSNG). Two interactions are also significant. These are European trials published in journals with a high impact factor, and null results from randomizations conducted outside North America and Europe. The introduction of the latter caused start of accrual period (NSTART), which was of borderline significance (p=0.0439) to be dropped (p=0.1098). When country group of trialists was split into 'North America' versus 'Europe' versus 'Other', so that the categories of this variable used in the interactions were also included as main effects, the two main terms were not significant. The interaction between statistical significance and null result was also significant, but this was found to be based on a subset of size 1. Record 697, (randomization 1902) is the only occurrence of a record with statistical significance 'p<0.05' and direction of results 'null' associated with it. This interaction term is therefore not included in the model. This observation is discussed in more detail in Section 4.4.

Diagnostic plots now show a pair of outliers, records 696 and 697, from randomizations 1901 and 1902 respectively, from publication 135. These have studentised residuals of 3.3 and 3.4 and the second and third largest Cook's distances, with record 716 one of a pair from publication 148 having the largest.

Note the increase in the proportion of variability explained by the model once the second stage variables are incorporated, from a poor R^2 value of 0.22 to a moderate 0.36.

The revised model is now given.

Final results

Time from close to publication of article is longer for randomizations with the following characteristics:

Main effects:

- Shorter duration of accrual period (DURRAN) (p=0.0002), typical effect 9 months
- No clear indication of whether clinically significant or not is given (CLNSG) (p=0.0060), typical effect 1 year 3 months
- Reported in a journal with an impact factor associated with it (NOIMPACT) (p<0.0001), typical effect 2 years
- Results with larger p-values associated with them (LOGPEST) (p=0.0153), typical effect 6 months
- Published in a journal with a low impact factor (IMPACT) (p=0.0062), typical effect 9 months
- Direction of results not null (POSNG) (p=0.0127), typical effect 1 year 2 months

Interactions:

- Conducted by European trialists and reported in a publication with a high impact factor (CGROUP*IMPACT) (p=0.0006), typical effect 1 year 9 months
- Conducted outside North America and Europe and results are null (CGROUP*POSNG) (p=0.0023), typical effect 4 years 6 months

[The typical effects of the non-significant main effects CGROUP are: conducted **outside** Europe 7 months, and conducted outside North America and Europe 1 month.]

(R^2=0.364055 F-statistic=9.10, p-value<0.0001 based on all 170 observations)

Mean and range of response variable: 1310 (-972 to 4326) days, i.e. approximately 3 years 7 months (–2 years 8 months to 11 years 10 months).

7 PRELIMINARY RESULTS OF THE 'ALL MENTIONS' AND 'ALL REPORTINGS OF RESULTS' ANALYSES

7.1 Introduction

This chapter begins by explaining why, if the assumption of independence is not satisfied, the correlations must be compensated for. This is followed by a description of atypical publications and individual observations which were outliers in one or more of the 'all mentions' and 'all reportings of results' analyses. The findings from these six analyses are then described. These are the analyses with the three time period response variables; 'time from close to submission', 'time from receipt to publication' and the sum of these, 'time from close to publication', firstly for all records (i.e. all mentions of all randomizations) and then for all records which contain results. Since there can be more than one record relating to each randomization, the assumption of independence is not satisfied, and so the correlations should be accounted for. Therefore the results of the analyses reported in this chapter are preliminary findings only. Details of the variables tried in these analyses are given in Appendix VII and the output for the final choice of preliminary model in each case is given as Appendix IX.

7.2 The need to account for repeated measures

In the three analyses involving all mentions and the three involving all results, some randomizations generate several records, and so records are not independent. Therefore repeated measures need to be compensated for. However before this stage it is necessary to perform preliminary analyses which use the same methods as the final analyses for the 'time to first' questions reported in the previous chapter.

Reasons for using preliminary analyses with multiple linear regression (the independence model):

- To get a feel for the data
- To reduce the number of variables that need investigating with repeated measures analysis.

- To obtain preliminary graphs and diagnostic plots. The diagnostics from multiple linear regression, for analysing residuals, are good and well trusted.
- Repeated measures analysis involves an iterative method, which uses much more computer time than multiple linear regression.
- To avoid the risk of over parameterisation. When repeated measures are used and correlation structures are introduced, more parameters are introduced. If there are too many explanatory variables, there is a risk of over-parameterising. This can lead to the likelihoods being flat and the algorithms that maximise the likelihoods running into convergence difficulties.

There are two aspects of repeated measures which need to be investigated:

- clustering, where each randomization is a cluster
- serial correlation, where for a randomization the time to publication of the first article is correlated with that of the second, that of the second with that of the third, that of the first with that of the third and so on

7.3 Outlying articles and records

Publication 61

Publication 61 [Jones et al (1991)] generated three records, 546, 547 and 548, reporting randomizations 1407, 1408 and 1409 respectively, all from trial 1407 (CLB 7111). This article is the fifth mention and fourth reporting of results for randomization 1407, the seventh mention and third reporting of results for randomization 1408 and the fourth mention and third reporting of results for randomization 1409. Randomizations 1407 and 1408 opened 5 February 1971, and 1408 opened 12 July 1971. All three closed in March 1974. The article was received by the publisher 16 April 1990, accepted 21 February 1991 and published sometime during 1991, the day and month unknown. This trio of records was atypical in four out of six of the following analyses.

For ‘close to publication’ for ‘all mentions’ and for ‘close to submission’ and ‘close to publication’ for ‘all results’, all three records were atypical. For ‘close to submission’ for all ‘mentions’ 546 and 548 were outliers. The time from close to submission for this article was long (16 years 1 month), also resulting in a long time from close to publication. The time from receipt to publication was not exceptional for this trio of records.

Record 547 reports results that are highly statistically significant ($p<0.001$) and both records 546 and 547 report clinically significant results, All three randomizations recruited large numbers of patients, 646, 467 and 493 respectively. Also, randomization 1407 has eight treatment arms. It appears that this large trial, with randomizations which have yielded results of statistical and clinical significance, was considered worthy of long-term follow-up 16 years later. Clinical and/or statistical significance is present in all four models for which these records are outliers or have large Cook’s distances. Also either number of arms or number of questions is present in three out of the four models, the exception being ‘time from close to publication for all mentions’.

Publication 100

Publication 100 [Bleyer et al (1991)] also generated three records, 628, 629 and 630, reporting 1618, 1619 and 1621, the three randomizations from trial 1618 (CCG-161). This article is the fourth mention and second reporting of results for randomization 1618, the fifth mention and second reporting of results for randomization 1619 and the eighth mention of randomization 1621. It did not report the results of randomization 1621. The article was received by the publisher 13 February 1989, accepted 20 November 1990 and published in June 1991. The long delay between receipt and publication (2 years 4 months) meant that all three records from this article were atypical in the ‘time from receipt to publication for all mentions’ analysis, and 628 and 629 in the ‘time from receipt to publication for all results’ analysis. The journal in question is *Journal of Clinical Oncology*, an American publication with a high impact factor of 6.922 associated with it. The trial is multi-centre and was conducted in the US. The numbers of patients accrued for each

randomization was large, 529, 625 and 285 respectively. Record 628 reports results of clinical significance and record 629 results of statistical significance ($p<0.01$).

Publication 95

Publication 95 [Lange et al (1996)] generated one record only, 615 (for randomization 1611, the sole randomization of trial 1611 (CCG-139)). It is the second mention and the second reporting of results for this randomization. Again the time from receipt to publication was exceptionally long (2 years 2 months). The article was submitted/received 2 May 1994, accepted 15 December 1994 and published some time in 1996, month and day unknown. However, for the purpose of analysis, where only the year of a date is known, the publication date is set to 30 June 1996 and could actually be up to six months earlier (or later). Again this record is extreme in two analyses, 'time from receipt to publication' for 'all mentions' and for 'all reportings of results'. However, it is possible that this observation is less (or more) extreme than it appears.

Publication 64

This publication [Hill et al (1994)], generated one record, 551 (for randomization 1416, the sole randomization from trial 1416 (CLB 7611)). It is the eighth mention and fifth reporting of results for this randomization. This record was atypical in two analyses, 'time from close to publication' for 'all mentions' and for 'all results'. The accrual period was from 12 November 1976 to late 1979 and the article was published as a meeting abstract in March 1994, 14 years 5 months after close of randomization. Dates of receipt and acceptance are unknown, and so this record could not be included in the 'time from close to submission' and 'time from receipt to publication' analyses. Randomization 1416 accrued a large number of patients (506) and this record reports a statistically significant result ($p<0.01$). Again, this is a large randomization with a statistically significant result, and in the opposite direction to expected, which may explain why it was considered worthy of reporting 14 years after it closed. However, perhaps it is unusual in that it was mentioned in an abstract so long after closure.

Publication 160

This publication [Koizumi et al (1991)] also generated a single record, 740 (for randomization 2601, the sole randomization from trial 2601 (JCCLSG S-811)). It is the fourth mention and third reporting of results for this randomization and was found to be an atypical observation for the 'time from receipt to publication for all mentions' analysis. The accrual period was from late 1980 until early 1984. The article was received by the publisher 16 November 1990, accepted 4 March 1991 and published some time during 1991, so the time from receipt to publication is short (just under 7 ½ months). As with publication 95, since the date of publication is not precisely known, this observation could be less (or more) extreme than it appears.

The atypical nature of the following observations can probably be disregarded:

Records 528, 529 and 531

Publication 57 [Bleyer 1990] generated nine records, three of which appeared as atypical observations in two analyses. These are 528 and 529, randomizations 1410 and 1411 from trial 1410 (CLB 7112 relapse) and 531, randomization 1413 from trial 1412 (CLB 7211 relapse). This article is the second mention and the second reporting of results for randomizations 1410, 1411 and 1413. The two analyses in which is was found to be an outlier are 'time from close to publication' for 'all mentions' and for 'all results'. The former used all nine records, the latter records 526, 527, 528, 529 and 531. The atypical records are for two early trials reported in the article, and those which recruited the fewest patients. The randomizations from trial 1410 were open during 1971, and that from trial 1412 was open during 1972, precise dates unknown. This is a book publication with dates of receipt and acceptance unknown and date of publication some time during 1990. Again, since the date of close of accrual period and publication are not precisely known, these observations could be less (or more) extreme than they appear.

Record 508

This record is one of nine generated by publication 52 [Holland and Glidewell (1972a)]. It is the first mention of randomization 1407 from trial 1407 (CLB 7111) and does not report its results. It

only appears as an outlier in the 'time from close to submission for all mentions' analysis. All records from publication 52 are used in the analysis, with the exception of 1068 (a newly found randomization linked to trial 1402). No other record from this publication except 508 is atypical, including 509 and 510, which are also linked to trial 1407. Randomization 1407 has a large number of arms (8) and this is a variable used in the model.

Record 955

This record, like 967, is from publication 284. However, whilst 967 is for a very early randomization (accrual period November 1967 to September 1970), 955 is for randomization 2105 from trial 2105 (GATLA 7 LLA-87), open between July 1987 and December 1989, so of short duration and the most recent trial reported in this article. This is the second mention of record 955 and does not report its results. This record shows up as atypical once only, in the 'close to submission' analysis using all mentions. All seven records from the publication are used and no others are exceptional.

7.4 For all mentions of each randomization, which trial characteristics affect the time from close to submission?

The best fitting model using the initial set of variables

Time from close to submission is longer for randomizations with the following characteristics:

- Larger number of patients accrued (LOGSIZE) (p=0.0247)
- Shorter duration of accrual period (DURRAN) (p=0.0006)
- Conducted outside North America (CGROUP) (p=0.0078)
- Not presented at a major meeting (PRESENTD) (p=0.0021)

(R^2=0.143962, F-statistic=7.61, p-value<0.0001 based on 186 out of 209 observations)

Comments

An alternative model was also being considered at this point. (R^2=0.1755, F=8.73, p-value<0.0001 using 169 of the 209 observations). This used variable impact factor (IMPACT) instead of country

group (CGROUP), but was abandoned due to the 9 % missing values for variable IMPACT. Variable CGROUP had no missing values. The model using IMPACT would have used 169 observations out of a possible 209, as opposed to 186. However once the not reported category for impact factor was introduced, IMPACT had no missing values. It was tried in the model again with the second stage variables. It is worth noting that country group can be thought of as a surrogate for impact factor. At this initial stage the two have a high negative correlation of ρ=-0.51

Comments on the final choice of preliminary model

There were two possible choices for the best preliminary model. The first is as follows:

Time from close to submission is longer for randomizations with the following characteristics:

Main effects:

- Shorter duration of accrual period (DURRAN) (p=0.0001)
- From a publication with no or a low impact factor (IMPACT) (p=0.0201)
- Results have smaller p-values associated with them (LOGPEST) (p=0.0032)
- Conducted in a 'developing' country (DEVLPNG) (p<0.0001)
- Reported in an article which mentions a **greater** number of trials (LOGNTREP) (p=0.0313)
- Reported in an article which mentions **fewer** randomizations (NRREP) (p=0.0002)
- Conducted outside North America (CGROUP) (p=0.0165)

Interaction terms:

- For trials conducted outside North America, a smaller number of patients accrued (CGROUP*LOGSIZE) (p=0.0214)

(R^2=0.414273, F-statistic=13.83, p-value<0.0001 based on 186/209 observations)

$\log_{10}$(number of patients accrued) (LOGSIZE) is of borderline significance (p=0.0522), but is retained in any case due to the significance of its interaction with country group of trialists. This main effect acts in the opposite direction from that of the interaction, i.e. a longer time from close to submission is associated with the accrual of a larger number of patients.

Both country group (CGROUP) and impact factor (IMPACT) remain in the model, although highly correlated, as does duration of randomization (DURRAN). Whether presented at a meeting (PRESENTD) is now no longer significant (eliminated at p=0.0700).

Diagnostic plots of this model indicate an outlier, record 955 with a studentised residual of -3.4 and the largest Cook's distance, with record 967 from the same publication having the second largest.

When the regression was run omitting record 967, from publication 284, the interaction between LOGSIZE and CGROUP was no longer significant (p=0.1065). For this reason, the only moderate significance of the interaction term (p=0.0214) and the loss of 13 observations (6%) its inclusion would have caused, the model without the interaction term and main effect LOGSIZE was chosen. The simpler model is robust to the inclusion/exclusion of record 967, and is based on all observations for which the response variable is present.

For the final model the diagnostic plots showed records 548 from publication 61 and 955 from publication 284 to be outliers, with studentised residual of approximately 3 and -3.5 respectively. Record 546, also from publication 61 also had a studentised residual of almost 3. Records 955 and 967 had the two largest Cook's distances, in that order.
Hence the model of choice is as follows. Note the improved fit of the model once the second stage variables were included, R^2=0.427824, compared with the poor value of R^2=0.143962 when only the initial set of variables were used.

Preliminary results

Time from close to submission is longer for randomizations with the following characteristics:

- Shorter duration of accrual period (DURRAN) (p=0.0001)
- From a publication with no or a low impact factor (IMPACT) (p=0.0012)
- Results have smaller p-values associated with them (LOGPEST) (p=0.0006)
- Conducted in a 'developing' country (DEVLPNG) (p<0.0001)

- Reported in an article which mentions a **greater** number of trials (LOGNTREP)) (p=0.0090)
- Reported in an article which mentions **fewer** randomizations (NRREP) (p<0.0001)

(R^2=0.427824, F-statistic=25.17, p-value<0.0001 based on all 209 observations)

Mean and range of response variable: 1923 (-1113 to 9503) days,i.e. approximately 5 years 3 months (–3 years 1 month to 26 years 0 months).

7.5 For all mentions of each randomization, which trial characteristics affect the time from receipt to publication?

The best fitting model using the initial set of variables

Time from receipt to publication is longer for randomizations with the following characteristics:

- Later start date of accrual period NSTART (p<0.0001)
- Trials comprises fewer randomizations (NRAND) (p<0.0001)
- Published in a non-European journal (JGROUP) (p<0.0001)
- Published in a North American or European journal (JGROUP) (p<0.0001), the latter having a greater effect than the former.

  (A clearer interpretation of the above two results is that the time from receipt to publication is shortest for randomizations reported in non-American, non-European publications, followed by those reported in European publications, with those reported in American publications taking the longest time.)
- Not presented at a major meeting (PRESENTD) (p=0.0065)

(R^2=0.323848, F-statistic=18.58, p-value<0.0001 based on 200 of the 218 observations)

Comments

Once second stage variables were incorporated duration of randomization period (DURRAN) became significant and whether presented at a major meeting (PRESENTD) and number of randomizations (NRAND) were dropped. DURRAN is highly correlated with NRAND (ρ>0.4).

Also the number of randomizations reported in article (NRREP) was significant. This may explain some of the variation in NRAND not accounted for by DURRAN. The relationship between NRAND and DURRAN is interesting. Within most trials, randomizations address totally different questions. For example one may be between induction treatments and another radiotherapy doses. In other trials there can be several randomizations from the same question. For example, a randomization between maintenance treatments may start by comparing two therapies, then add a third arm, then drop one of the original arms. This is recorded as three randomizations belonging to the same trial. In this case duration of randomization period (DURRAN) is likely to be highly negatively correlated with number of randomizations (NRAND). Unfortunately there is not a variable in the data to distinguish between what could be referred to as **'separate randomizations'** and **'serial randomizations'**. Many of the serial randomizations are from trials performed in the US. Separate randomizations are more common than serial randomizations, and are more likely to affect publication, since if one randomization is reported the other is also likely to be mentioned and vice versa. Serial randomizations are likely to be written up together, quoting joint results.

In order to try interactions, indicator variables were used that divided country group (CGROUP) further to 'North America' versus 'Europe' versus' Other'. Although none of the interactions tested were significant, the variable distinguishing European trial groups from others was, and its introduction resulted in country group of publisher (JGROUP) 'European', being dropped.

Identical regressions were produced whether LOGAUTH or LOGMRC was used. This shows that for the few early articles describing Medical Research Council trials, which do not list the authors, whether an approximate working party size is used, or an estimate of the number of people on the writing committee, makes little difference.

A possible model is as follows: time from receipt to publication is longer for randomizations with the following characteristics:

- Later start date of accrual period (NSTART) ($p<0.0001$)
- Published in a North American or European journal (JGROUP) ($p<0.0001$)

- Longer duration of accrual period (DURRAN) (p=0.0002)
- Reported in articles which mention fewer randomizations (NRREP) (p<0.0001)
- Conducted outside Europe (CGROUP) (p<0.0001)
- Larger number of co-authors (LOGAUTH/LOGMRC) (p=0.0339)

(R^2=0.451353, F-statistic=25.78, p-value<0.0001 based on 195 out of 218 observations)

Diagnostic plots are given as Figures 7.1 and 7.2. The former shows that three extreme outliers remain; records 628, 629 and 630, the trio from publication 100, which all have studentised residuals of approximately 4, and record 615 which has studentised residual of approximately 3. The four largest Cook's distances in order are 629, 630, 615 and 628, although these are not extremely large. The graph of residuals against fitted values (Figure 7.1) produced a wedge-shape rather than a random spread, indicating that the variance is dependent on the mean.

Both logarithmic, to base 10, and square root transformations of the response variable were tried. Diagnostic plots indicated that both models are a better fit to the data than with the untransformed response, and that the square root transformation is the one to use. The plots for the chosen model are given as Figures 7.3 and 7.4.

The scatter plot of standardised residuals versus fitted values (Figure 7.3) shows a more even spread, indicating that the square root of the response variable 'time from receipt to publication' has constant variance. The line in the graph of residuals versus their normal scores (Figure 7.4) is now straighter, showing that the square root of the response variable 'time from receipt to publication' is approximately normally distributed.

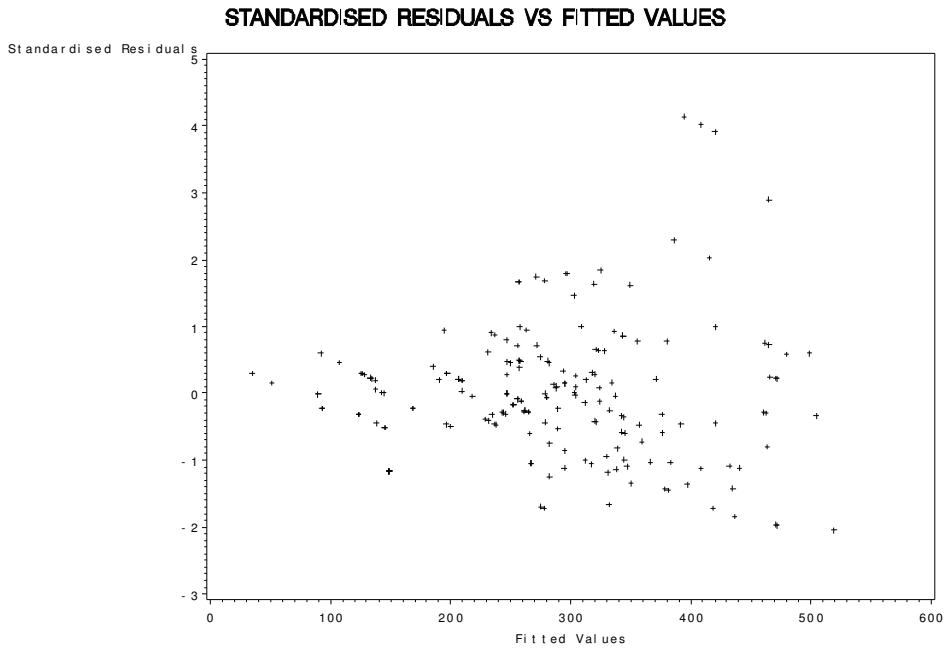


Figure 7.1 Linear regression model of time from receipt to publication for all mentions before transformation. Standardised residuals vs. fitted values

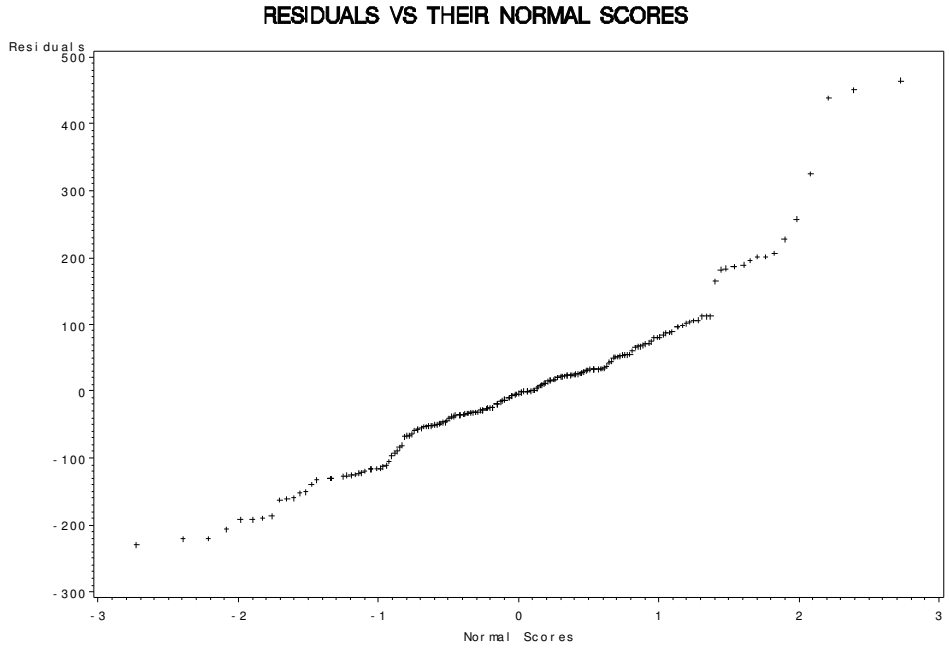


Figure 7.2 Linear regression model of time from receipt to publication for all mentions before transformation. Residuals vs. their normal scores

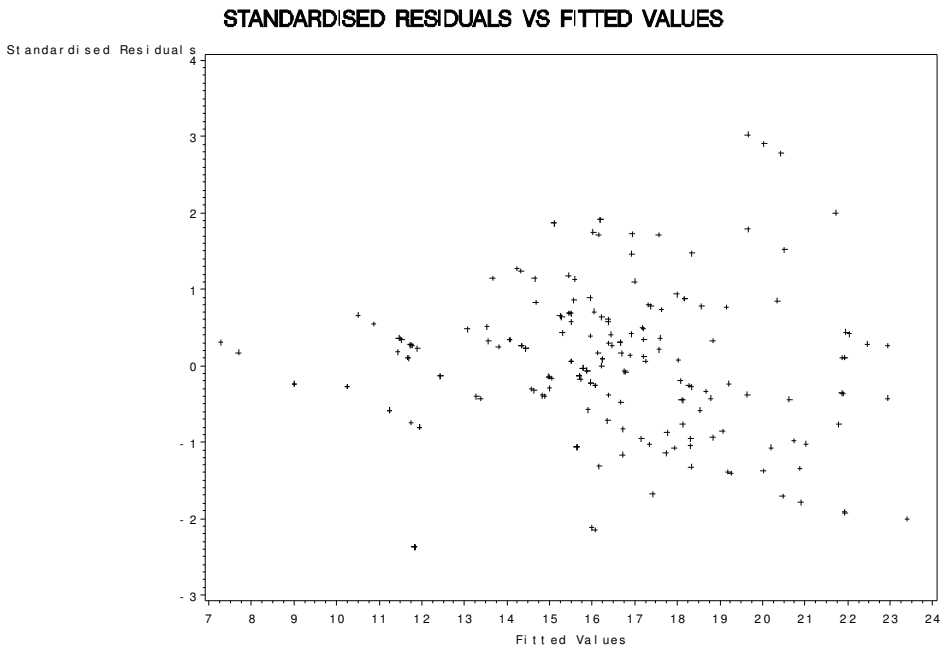


Figure 7.3 Linear regression model of time from receipt to publication for all mentions after square root transformation. Standardised residuals vs. fitted values

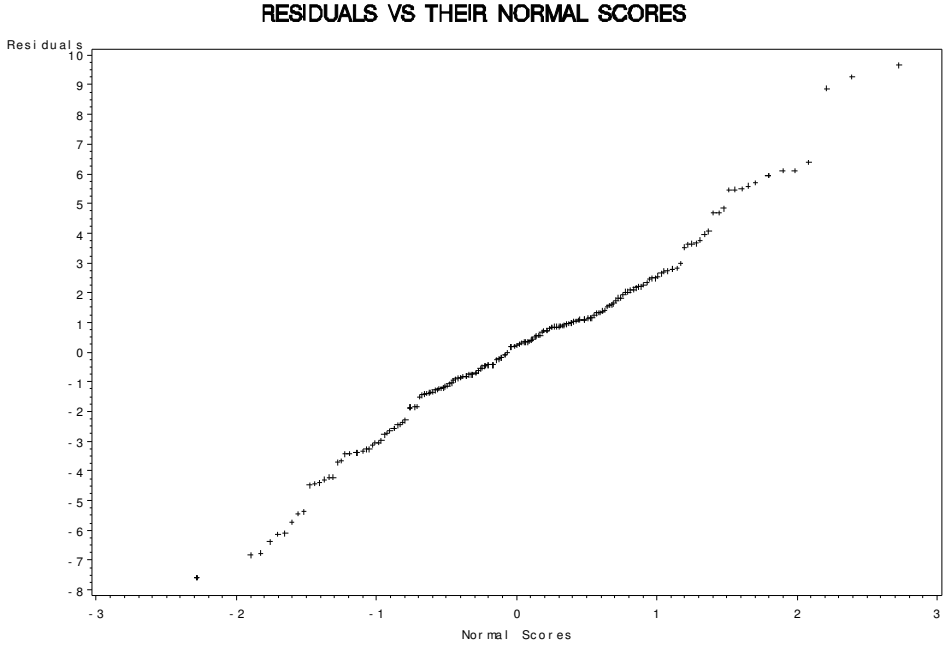


Figure 7.4 Linear regression model of time from receipt to publication for all mentions after square root transformation. Residuals vs. their normal scores

Preliminary results

√(time from receipt to publication) is longer for randomizations with the following characteristics:

- Later start date of accrual period (NSTART) ($p<0.0001$)

- Published in a North American or European journal (JGROUP) (p<0.0001)
- Longer duration of accrual period (DURRAN) (p=0.0002)
- Reported in articles which mention fewer randomizations (NRREP) (p<0.0001)
- Conducted outside Europe (CGROUP) (p<0.0001)
- Larger number of co-authors (LOGAUTH/LOGMRC) (p=0.0121)

(R^2=0.502887, F-statistic=31.70, p-value <0.0001 based on 195 out of 218 observations)

Mean and range of response variable: 284 (18 to 859) days, i.e. approximately 9 months (1 month to 2 years 4 months).

7.6 For all mentions of each randomization, which trial characteristics affect the time from close to publication?

The best fitting model using the initial set of variables

Time from close to publication is longer for randomizations with the following characteristics:

- Earlier start of accrual period (NSTART) (p<0.0001)
- Shorter duration of randomization period (DURRAN) (p<0.0001)
- Participation of five or more centres (MULTIC) (p=0.0003)
- Conducted outside North America and Europe (CGROUP) (p<0.0001)
- Published in full in the English language (ENGLISH) (p=0.0045)
- Has not been presented at a meeting (PRESENTD) (p<0.0001)
- Trial comprises fewer randomizations (NRAND) (p=0.0444)

(R^2=0.218208, F-statistic=22.85, p-value<0.0001 based on 581 out of 582 observations)

Comments

Another possible model consisted of the above variables but with log_{10} (number of patients accrued) (LOGSIZE) replacing MULTIC and NRAND (R^2 = 0.1945, F-statistic = 20.73 using 522 of the 582 observations). However this was abandoned since LOGSIZE had missing values for 10% of observations, whereas MULTIC had less than 1% missing values. However, MULTIC may be seen as a surrogate for LOGSIZE since their correlation coefficient is moderately high at 0.41.

Comments on the final choice of preliminary model

There is a choice between two possible models. The first is as follows; time from close to publication of article is longer for randomizations with the following characteristics:

Main effects:

- Earlier start of accrual period (NSTART) ($p<0.0001$)
- Short duration of accrual period (DURRAN) ($p=0.0031$)
- Published in full in the English language (ENGLISH) ($p=0.0037$)
- Not presented at a major meeting (PRESENTD) ($p<0.0001$)
- Results are clearly reported as clinically significant (CLNSG) ($p=0.0051$)
- Clinically significance is not reported (CLNSG) ($p=0.0051$),

  the former having a slightly greater effect than the latter.
- A clear indication is given as to whether the main questions in the paper are answered in that paper (ANSWER) ($p<0.0001$)
- Conducted in a 'developing' country (DEVLPNG) ($p<0.0001$)
- Reported in articles which mention a **greater** number of trials (NTREP) ($p=0.0003$)
- Reported in articles which mention **fewer** randomizations (NRREP) ($p=0.0272$)
- Reported in a publication with a low or no impact factor (IMPACT) ($p=0.0348$)
- Reported in a journal article or book chapter as opposed to a meeting abstract (PUBTYPE) ($p=0.0403$)
- Conducted outside Europe (CGROUP) ($p=0.0018$)
- Conducted outside North America and Europe (CGROUP) ($p=0.0127$),

  the latter having a greater effect than the former.
- Participation of five or more centres (MULTIC) ($p<0.0001$)

Interactions:

- For trials conducted outside North America, the effect of the participation of less than five centres is to delay publication (CGROUP*MULTIC) ($p=0.0012$)

- In the case of trials conducted outside North America and Europe, a smaller number of patients accrued is also significant (CGROUP*LOGSIZE) (p=0.0117)

(R^2=0.344253, F-statistic=14.64, p-value<0.0001 based on all 521 out of 582 observations)

It is worth noting the following points: This model contains sixteen main terms and two interactions. The effects of number of trials and number of randomizations mentioned in an article (NTREP and NRREP) are in opposite directions. The introduction of the non-significant variable LOGSIZE (p=0.3962) because of its moderately significant interaction with country group (p=0.0117), results in a loss of 60 observations.

Diagnostic plots show that several outliers remain (records 546, 547 and 548, the trio from publication 61, records 528, 529 and 531 from publication 57 and record 551 from publication 64). All have studentised residuals greater than 3. Record 967 from publication 284 has a large Cook's distance.

When record 967 was omitted, the interaction between LOGSIZE and CGROUP became non-significant, and hence variable LOGSIZE could be dropped also, meaning that the 60 observations could now be used. The model without this interaction was robust to the inclusion/exclusion of record 967. For reasons of this robustness, the additional 60 observations that could be included (581 out of the 582 observations for which the response variable was present can now be used) and the only moderate significance of the interaction term (p=0.0117) the simpler model was chosen.

Diagnostic plots for the final choice of model show that several outliers remain (records 546, 547 and 548, the trio from publication 61, records 528, 529 and 531 from publication 57, record 551 from publication 64 and record 967 from publication 284). All have studentised residuals greater than 3. Record 967 from publication 284 has a large Cook's distance.

Preliminary results

Hence, the model of choice is: time from close to publication of article is longer for randomizations with the following characteristics:

Main effects:

- Earlier start of accrual period (NSTART) (p<0.0001)
- Shorter duration of accrual period (DURRAN) (p<0.0001)
- Published in full in the English language (ENGLISH) (p=0.0041)
- Not presented at a major meeting (PRESENTD) (p<0.0001)
- Results are clearly reported as clinically significant (CLNSG) (p=0.0030)
- Clinically significance is not reported (CLNSG) (p=0.0028),

  the former having a marginally greater effect than the latter.

(A clearer interpretation of the above two findings is that results clearly reported as **not** or only possibly clinically significant are published fastest, followed by those where no indication of clinical significance is given, with results clearly reported as clinically significant having longest time to publication.)

- A clear indication is given as to whether the main questions in the paper are answered in that paper (ANSWER) (p<0.0001)
- Conducted in a 'developing' country (DEVLPNG) (p<0.0001)
- Reported in articles which mention a **greater** number of trials (NTREP) (p<0.0001)
- Reported in articles which mention **fewer** randomizations (NRREP) (p=0.0021)
- Reported in a publication with a low or no impact factor (IMPACT) (p=0.0252)
- Reported in a journal article or book chapter as opposed to a meeting abstract (PUBTYPE) (p=0.0253)
- Conducted outside Europe (CGROUP) (p=0.0004)
- Participation of five or more centres (MULTIC) (p<0.0001)

Interaction:

- For trials conducted outside North America the effect of participation of less than five centres is to delay publication (CGROUP*MULTIC) (p=0.0006)

(R^2=0.341276, F-statistic=18.26, p-value<0.0001 based on 581 out of 582 observations)

Mean and range of response variable: 1793 (–1679 to 10150) days, i.e. approximately 4 years 11 months (–4 years 7 months to 27 years 9 months).

7.7 For all reportings of results for each randomization, which trial characteristics affect the time from close to submission?

The best fitting model using the initial set of variables

Time from close to submission is longer for randomizations with the following characteristics:

- Greater number of questions (NOQ) (p=0.0053)
- Multi-centre participation rather than single-centre (MULTIC) p=0.0238)
- Limited international or single-country rather than truly international participation (INTERNL) (p=0.0353)
- Not presented at a major meeting (PRESENTD) (p=0.0045)

(R^2=0.196965, F-statistic=7.48, p-value<0.0001 based on 127 out of 129 observations)

Comments

Another possible model at this stage was as above but with statistical significance (LOGPEST) replacing MULTIC. (F=11.10, p-value =0.0001, R^2= 0.2894 using 114/129 observations) However this was abandoned because of the large number of missing values, despite the superior value of R^2.

As explained in Section 5.5, during the initial stage none of the results variables had a 'not reported' category, leading to a high proportion of missing values for each. This was rectified at the second stage.

It was thought that MULTIC might be considered a surrogate for LOGPEST. However, the correlation coefficient between the two is only 0.13277 (p=0.0559 on 208/209 observations). The interaction between these two variables will be tried, i.e. results with smaller p-values from large multi-centre trials versus the rest.

Comments on the final choice of preliminary model

There are two outliers (records 547 and 548 with studentised residuals 3.5 and 3.9 respectively). Record 546 has a large Cook's distance. The trio is from publication 61 and is left in the analysis.

None of the interactions tried were significant.

Note that this model includes variable LOGPEST (and MULTIC has been dropped), as in the alternative model using the initial set of variables only, but since the 'not reported' category for LOGPEST was introduced with the second stage variables, all 129 observations are now used in the analysis. The value of R^2 has improved from a poor 0.196965 for the chosen initial model, and 0.2894 for the alternative model for which there was a loss of 12% of observations, to 0.323121 using all available data.

Preliminary results

Time from close to submission is longer for randomizations with the following characteristics:

- Greater number of questions (NOQ) (p=0.0011)
- Not presented at a major meeting (PRESENTD) (p=0.0003)
- From a publication with no or a low impact factor (IMPACT) (p=0.0100)
- Results have smaller p-values associated with them (LOGPEST) (p=0.0034)
- Clinical significance is 'yes', 'possibly' or not reported as opposed to 'no' (CLNSG) (p=0.0363)
- Direction of results is not reported (POSNG) (p=0.0475)

(R^2=0.323121, F-statistic=9.71, p-value<0.0001 on all 129 observations)

Mean and range of response variable: 1867 (–191 to 5876) days, i.e. approximately 5 years 1 month (–6 months to 16 years 1 month).

7.8 For all reportings of results for each randomization, which trial characteristics affect the time from receipt to publication?

The best fitting model using the initial set of variables

Time from receipt to publication is longer for randomizations with the following characteristics:

- Trial comprises fewer randomizations (NRAND) (p=0.0035)
- Multi-centre participation (MULTIC) (p<0.0001)
- Published in a non-European journal (JGROUP) (p=0.0003)
- Published in a North American or European journal (JGROUP) (p=0.0034),
  the latter having a greater effect than the former.
- Not presented at a major meeting (PRESENTD) (p=0.0014)

(R^2=0.286860, F-statistic=9.98, p-value<0.0001 based on 130 of the 137 observations)

Comments on the final choice of preliminary model

Once second stage variables are incorporated, duration of randomization period (DURRAN) becomes significant and number of randomizations (NRAND) is dropped. The two have a correlation coefficient of –0.29. The relationship between these two variables has been discussed in Section 7.5. Other new main effects to enter the model at the second stage are whether an international trial (INTERNL) and whether the main questions have been answered (ANSWER). In addition the interaction between whether an international trial (INTERNL) and impact factor (IMPACT) is also significant.

Diagnostic plots indicate one slightly large Cook's distance (record 615 from publication 95) and two outliers with studentised residuals of approximately 3 (records 628 and 629, two of the three records from publication 100). These are not extreme and all are left in.

Again there is a large improvement in the proportion of variability explained by the model once the second stage variables are included, R^2 increased from 0.29 to 0.52.

Preliminary results

Time from receipt to publication is longer for randomizations with the following characteristics:

Main effects:

- Multi-centre participation (MULTIC) ($p<0.0001$)
- Published in a non-European journal (JGROUP) ($p<0.0001$)
- Published in a North American or European journal (JGROUP) ($p<0.0001$), the latter having a greater effect than the former.

(A clearer explanation of the two findings above is that results published in non-US, non-European journals are published more quickly than those published in European journals, with those published in North American journals having the longest time to publication.)

- Longer duration of accrual period (DURRAN) ($p<0.0001$)
- The main questions as stated in the paper are answered in the paper (ANSWER) ($p=0.0042$)
- Not presented at a major meeting (PRESENTD) ($p=0.0113$)
- Trials with limited international or single-country, as opposed truly international, participation (INTERNL) ($p=0.0002$)

Interaction term:

- Truly international trials with results published in a high impact factor journal (INTERNL*IMPACT) ($p=0.0016$)

($R^2=0.516723$, F-statistic=13.19, p-value<0.0001 based on 121 out of 137 observations)

Mean and range of response variable: 284 (18 to 859) days, i.e. approximately 9 months (1 month to 2 years 4 months).

It is interesting to note that the effect of being an international trial reduces the time from receipt to publication, but where the results of an international trial are published in a journal with a non-zero impact factor the reduction is not as great. For journals with an impact factor of less than 8, the

time to publication is reduced. For journals with a higher impact factor, the effect is reversed and the time to publication is increased.

7.9 For all reportings of results for each randomization, which trial characteristics affect the time from close to publication?

The best fitting model using the initial set of variables

Time from close to publication is longer for randomizations with the following characteristics:

- Larger number of questions (NOQ) (p=0.0008)
- Shorter duration of randomization period (DURRAN) (p<0.0001)
- Randomization has not been presented at a meeting (PRESENTD) (p=0.0019)
- Direction of results is not null (POSNG) [4] (p=0.0066)

(R^2=0.140677, F-statistic=13.22, p-value<0.0001 based on 328 out of 372 observations)

Comments

There was another possible model at this stage. This used clinical significance (CLNSG) instead of direction of results (R^2=0.1574, F=14.39, p-value<0.0001 using 313 out of 372 observations). The first model was chosen as a starting point, in order to conserve observations used. It is interesting to note that once the second stage variables are included and non-reported options for results variables are used, clinical significance remains in the final model and direction of results is eliminated.

Comments on the final choice of preliminary model

The diagnostic plots for the final model show that there is one observation with a large Cook's distance (record 546), and records 547 and 548 are outliers. This is the trio of records from publication 61. Also outliers are records 528, 529 and 531 from publication 57 (the other records from publication 57 are not outliers) and record 551 (the sole record from publication 64). All

[4] As was explained in Section 5.5, during the initial stage none of the results variables had a 'not reported' category, leading to a high proportion of missing values for each. This was rectified at the second stage.

outliers have studentised residuals greater than 3 and those for records 547 and 548 are approximately 4.

The low value of R^2 (0.216225), the proportion of variability of the response variable that is explained by the model and a guide of the appropriateness of the model, is of some concern. A value of at least 0.3 is aimed for.

Preliminary results

Time from close to publication of article is longer for randomizations with the following characteristics:

Main effects:

- Greater number of questions (NOQ) (p=0.0190)
- Earlier close date of accrual period (NCLOSE) (p=0.0002)
- Not presented at a meeting (PRESENTD) (p=0.00012)
- Results are clearly reported as clinically significant or not reported at all (CLNSG) (p=0.0076)
- A clear indication is given as to whether or not the main questions in the paper are answered in that paper (ANSWER) (p=0.0045)
- Treatments for relapse or refractory disease (DUMFL) (p=0.0081)
- Reported in articles which mention a greater number of trials (NTREP) (p=0.0069)
- Results with smaller p-values (LOGPEST) (p<0.0001)

Interaction:

- Positive results with larger p-values (POSNG*LOGPEST) (p=0.0002)
- Negative results with larger p-values (POSNG*LOGPEST) (p=0.0082),

  the first of these having a marginally greater effect than the second.

(R^2=0.216225, F-statistic=8.00, p-value<0.0001 based on 361 out of 372 observations)

Mean and range of response variable: 1830 (-972 to 6940) days, i.e. approximately 5 years 0 months (-2 years 8 months to 19 years 0 months).

8 REPEATED MEASURES ANALYSIS

8.1 Introduction

This chapter starts by explaining the structure of the data and why the observations are not independent. There follows a summary of the assumptions of linear regression models and how all except that of independence are satisfied. This is followed by an explanation that an adjustment must be made to the independence model in order to compensate for the repeated measures and that this involves incorporating the covariances between values of the response variable, and hence between the residuals, for first, second, third etc. publications, from the analysis using the independence model. For each of the six analyses, the covariance matrix is obtained.

The rest of the chapter describes a preliminary investigation into modelling the covariance matrix. Five possible correlation structures are described; independence, unstructured, exchangeable, stationary *m*-dependent and autoregressive. The reason for trying to find a plausible correlation structure rather than using the unstructured model is to avoid over-parameterisation. Therefore exploratory analyses are performed using the residuals from the independence model. This is done for the analyses using the two largest datasets, 'time from close to publication' for 'all mentions' and for 'all reportings of results'. For brevity these analyses will be referred to as 'all mentions' and 'all results' respectively in this chapter.

The variance inflation factor method is then introduced. This simple model, although too much of an over-simplification for these data, serves the purpose of giving some feel for the order of correction to be expected from the repeated measures analysis. The variance inflation factor method is equivalent to the first iteration of the generalised estimating equation (GEE) process in the case of the exchangeable correlation structure.

A preliminary investigation using a similar method is then undertaken to obtain empirical correlation matrices using other structures.

8.2 Structure of the data

Some randomizations are reported once only and others several times. An article may report more than one trial, each comprising several randomizations. In the publications database each mention of a randomization counts as a record. Thus there can be multiple records relating to each randomization, i.e. clustering where each randomization is a cluster, and so records are not independent. Therefore, for each of the three analyses using all mentions of a randomization, and the three which use all results, repeated measures analysis must be incorporated.

Since several publications report the same randomization, it is necessary to investigate for that randomization the serial correlations i.e. how the second publication is related to the first in terms of the response variable, the response variable being a time period; from close to submission, from receipt by publisher to publication or the sum of these, from close to publication. Similarly the relationship of the third publication to the first, the third to the second, the fourth to the third and so on should be investigated, and whether this is common to all randomizations. There may be, for example, a 'first publication effect' in as much as the first reporting may be generally quicker or slower than subsequent reportings. By using a repeated measures analysis, the correlation between the responses is built into the model obtained from the preliminary analysis, the independence model.

If the data are clustered but this is not accounted for (i.e. the records are assumed to be independent, as in the preliminary analyses) the estimation of the variance for each coefficient may be smaller than it is in reality, in which case the variables in the model would appear to be more significant than they are. A set of non-independent records might be considered equivalent to a smaller sample of independent records, which would have a larger variance.

The significance of variable X_i is calculated from the value of $t = \hat{\beta}_i / s.e.(\hat{\beta}_i)$,

where $\hat{\beta}_i$ is the estimate of parameter β_i and *s.e.*($\hat{\beta}_i$) is the standard error of $\hat{\beta}_i$. The effect that the incorporation of repeated measures will have on the preliminary model is to increase *s.e.*($\hat{\beta}_i$). However it is possible for the algorithm to give a different weighting in estimating $\hat{\beta}_i$, which could increase it. Therefore $\hat{\beta}_i$ could become large enough to reduce its *p*-value but it is unlikely that it would change much from the independence case, and this would not justify including explanatory variables dropped from the preliminary model, unless their exclusion was borderline.

8.3 Assumptions of linear regression models

For the preliminary analysis, linear regression with backwards elimination was used to obtain a reduced set of significant variables. Linear models require the following assumptions:

- The response variables (and hence the residuals) are independent.
- The response variables (and hence the residuals) have a normal distribution with constant variance. The residuals have mean equal to zero i.e. distribution N (0, σ^2).

Diagnostic plots indicate that the second assumption is satisfied. When the residuals are plotted against normal scores the line produced is fairly straight, indicating a normal distribution. When (standardised) residuals are plotted against fitted values the random pattern observed suggests constant variance. However, clearly the data cannot satisfy the first assumption.

There are two aspects to consider; correlation between the response variable for publication records relating to (i) the same randomization, and (ii) randomizations belonging to the same trial. In order to compensate for the former an adjustment is needed. This involves incorporating the covariances between values of the response variable (and hence between the residuals) for first publications of each randomization, second publications and so on into the findings from the preliminary regression analyses. A investigation into correlation between randomizations belonging to the same trial was also undertaken. This is reported in Section 8.20.

The algorithm for the program used to produce the correlation matrix for the residuals from the independence model is given as Appendix X (i). The correlation matrices obtained from the residuals from the six analyses are given as Appendix XI.

8.4 Method for obtaining the correlation matrix for each of the six analyses

When looking at all mentions, variable ALLORDER specifies the order of the record for the randomization to which it refers, by date of publication. There is an equivalent variable RESORDER for the dataset containing only those records which report results.

The regression is run using the best fitting independence model. Residuals are output for the purpose of model testing and also to form the covariance matrix. The order number and the randomization ID are known for each residual. The residuals from the first reportings are renamed RESID1, those from second reportings RESID2 and so on. This is done up to and including the fifth reportings. Although some randomizations have been reported more than five times, (one randomization was published thirteen times), there are not enough sixth and subsequent reportings to warrant continuing the process. A correlation (standardised covariance) matrix is then calculated for variables RESID1, RESID2 … RESID5. There are a few cases where articles reporting the same randomization share the same publication date, and so their order will tie. There are eight cases of ties between pairs of records and one case of a tie of a trio of records. These are listed below. For the purpose of obtaining the covariance matrix for residuals, the ties are not treated as such, but will be in order of publications ID (the number assigned to the article as it was entered into the computer). As well as being the simplest solution, it is also logical to do this since if, for example, the 2nd and 3rd articles relating to a randomization are published on the same day, the time between those two publications, i.e. zero days, is included in the calculation of the matrix.

Some of the ties are due to the exact date of publication being unknown i.e. day or day and month missing. As was discussed in Section 5.3.1, it was a time-consuming task to discover which article

was the first to mention or report results for each randomization, and it was not considered worth the extra effort needed to break ties between the 2nd and 3rd publications, 3rd and 4th etc..

The following randomizations are reported in articles with tied publication dates. Note that where the 1st and 2nd tie, this is a genuine tie, which has been fully investigated, with both articles appearing in the same issue of a journal.

1205: 3rd and 4th	(publications 32 and 47)
1206: 2nd, 3rd and 4th	(publications 32, 41 and 47)
1207: 2nd and 3rd	(publications 32 and 41)
1208: 2nd and 3rd	(publications 32 and 47)
1603: 2nd and 3rd	(publications 82 and 84)
1604: 2nd and 3rd, 4th and 5th	(publications 78 and 79, 82 and 84)
2701: 1st and 2nd	(publications 4 and 5)
2708: 1st and 2nd	(publications 15 and 16)

The references for these publications are listed in publication number order in Appendix I.

Correlation matrices were calculated for the residuals from each of the six analyses, which contained repeated measures. Only the correlation coefficients from the residuals from the analyses on the two largest datasets are discussed here. Those from the residuals from the other four analyses are based on small numbers of observations, and so are less reliable. However all six are given in Appendix XI.

The elements forming the leading diagonal of the correlation matrix are, of course, unity. The other elements are examined, grouped by their distance from the leading diagonal. The purpose of this is to try to identify a plausible correlation structure so that this can be built into the repeated measures analysis. The structure developed for the two largest datasets will be applied to the other four analyses also. The estimates for the correlation coefficients will be different for the six analyses,

but the most suitable correlation structure found, using the two largest datasets, will be imposed on all six.

8.5 Possible correlation structures

These are described in Zeger and Liang (1986), and a more comprehensive explanation is given in Diggle et al (1994).

Independence

If none of the correlation coefficients are found to be significant then repeated measures need not be used. The preliminary analysis will suffice. Incorporation of the identity matrix has no effect on the original model.

$$\begin{matrix} 1 & 0 & 0 & - \\ 0 & 1 & 0 & - \\ 0 & 0 & 1 & - \\ - & - & - & - \end{matrix}$$

Unstructured (Non-independence)

All correlation coefficients are different. GEE work best on large datasets and convergence problems arise from datasets which are too small for the method. This structure is not considered because of the high degree of parameterisation it requires.

$$\begin{matrix} 1 & \rho_{12} & \rho_{13} & - \\ \rho_{21} & 1 & \rho_{23} & - \\ \rho_{31} & \rho_{32} & 1 & - \\ - & - & - & - \end{matrix}$$

where $\rho_{ji} = \rho_{ij}$

Exchangeable

All correlation coefficients are assumed to be equal. The advantage of this is that it produces the minimum parameterisation. The disadvantage is that the correlation structure is not likely to be a good approximation to the real covariance matrix, but it does give a useful guide to the effect of allowing for correlation. In this structure clustering is modelled, but not serial correlation.

$$\begin{matrix} 1 & \rho & \rho & - \\ \rho & 1 & \rho & - \\ \rho & \rho & 1 & - \\ - & - & - & - \end{matrix}$$

Stationary m-dependent

Here all elements distance 1 from the leading diagonal have the same correlation coefficient, ρ_1, all those distance 2 have coefficient ρ_2 and so on.

m = the number of diagonals where ρ_k is non-zero

$$\begin{matrix} 1 & \rho_1 & \rho_2 & - \\ \rho_1 & 1 & \rho_1 & - \\ \rho_2 & \rho_1 & 1 & - \\ - & - & - & - \end{matrix}$$

Autoregressive

This is similar to the stationary *m*-dependent correlation structure, except ρ_2, ρ_3 and so on are functions of ρ_1, so that only one new parameter need be introduced.

$$\begin{matrix} 1 & \rho^{|t2-t1|} & \rho^{|t3-t1|} & - \\ \rho^{|t1-t2|} & 1 & \rho^{|t3-t2|} & - \\ \rho^{|t1-t3|} & \rho^{|t2-t3|} & 1 & - \\ - & - & - & - \end{matrix}$$

where $|t_i - t_j| = |t_j - t_i|$

For the purpose of analysing these data *i* and *j* are integers representing the order of publication for articles reporting a particular randomization.

Therefore $t_i = i$ and $|t_i - t_j| = |i - j|$, giving the simplified correlation structure:

$$\begin{matrix} 1 & \rho & \rho^2 & - \\ \rho & 1 & \rho & - \\ \rho^2 & \rho & 1 & - \\ - & - & - & - \end{matrix}$$

8.6 The correlation matrix for the residuals from the independence model analysis of time from close to publication for all mentions

All the 1-step correlation coefficients are highly significant:

ρ_{12} =0.35271 p<0.0001 (n=128)

ρ_{23} =0.46667 p<0.0001 (n=90)

ρ_{34} =0.66755 p<0.0001 (n=68)

ρ_{45} =0.75220 p<0.0001 (n=44)

The 2-step correlation coefficients are less significant and less correlated:

ρ_{13} =0.16092 p=0.1297 (n=90) n/s

ρ_{24} =0.40235 p=0.0007 (n=68)

ρ_{35} =0.45929 p=0.0017 (n=44)

The 3-step correlation coefficients are:

ρ_{14} =0.16406 p=0.1813 (n=68) n/s

ρ_{25} =0.20546 p=0.1809 (n=44) n/s

The 4-step correlation coefficient is:

ρ_{15} = -0.05110 p=0.7418 (n=44) n/s

The pattern that emerges is one of highly significant strong correlations between successive groups of residuals. As the orders of publication become further apart the correlation coefficients between them become weaker and less significant. The strong pattern in the correlation matrix suggests use of the stationary m-dependent or autoregressive correlation structure.

Note that as i and j increase, the ρ_{ij} calculated will be less accurate since the number of observations, n_{ij}, on which it is based is smaller. For example; for correlation coefficients distance 2 from the leading diagonal, the value of coefficient ρ_{13} is more representative of those lying on that diagonal than that of ρ_{35}, although the latter is stronger and has a higher significance level. This is worth bearing in mind when trying to decide on a suitable correlation structure. Also the higher the value of a $\rho_{ij,}$ the more likely it is to be significant. For small correlations, the dataset used may be too small to show significance.

8.7 Estimation of the correlation coefficient for an exchangeable correlation structure using the variance inflation factor method: all mentions

The variance inflation factor method of Rao and Scott (1992) is an adaptation to binary data of the effective sample size effect outlined by Kish (1965).

The formula for calculating an estimate, ρ_{est}, of the correlation coefficient for an exchangeable correlation structure of dimensions $m \times m$, and hence the variance inflation factor is as follows:

Let the correlation matrix produced from the residuals from the independence model have elements $\rho_{ij,}$ where the subscripts refer to the i^{th} and j^{th} repeated measure and n_{ij} is the number of pairs of observations used to calculate ρ_{ij}.

For n_{ij} pairs of observations

$$z_{ij} = ln \sqrt{} [(1 + \rho_{ij}) / (1 - \rho_{ij})]$$

$$z_{est} = \frac{\sum_{i=2}^{m} \sum_{j=1}^{i-1} (n_{ij} - 3)\, z_{ij}}{\sum_{i=2}^{m} \sum_{j=1}^{i-1} (n_{ij} - 3)}$$

where $n_{ij} > 3$

and the estimate of the pooled correlation is

$$\rho_{est} = \frac{e^{2Z_{est}} - 1}{e^{2Z_{est}} + 1}$$

For the 'time from close to publication for all mentions' analysis, the estimate of the correlation coefficient for an exchangeable correlation structure is

$$\rho_{est} = 0.37837$$

8.8 The variance inflation factor (VIF)

For each of the six analyses involving repeated measures, the output from the preliminary (independence) model is given in Appendix IX. In each case, the parameter (β) estimates together with the standard error (σ), the *t*-statistic and the *p*-value for each are output.

The variance inflation factor compensates for the correlation by enlarging the variance. This is the traditional method for dealing with repeated measures, but is only suitable for use with an exchangeable correlation structure.

Since the β estimates have a *t*-distribution, $t=(x - \mu) / s$, the estimated variance, s^2, must be divided by the variance inflation factor.

The variance inflation factor = $1 - \rho_{est}^2$

Therefore the estimated standard error, *s*, must be divided by $\sqrt{VIF}$, i.e. by $\sqrt{(1 - \rho_{est}^2)}$.

The exchangeable correlation structure is easily incorporated into the preliminary model using the variance inflation factor.

To incorporate the exchangeable correlation matrix

$$s_{new} = s_{old} / \sqrt{VIF}$$

$$t_{new} = \hat{\beta} / s_{new}$$

Since the *t*-value for each variable is divided by the same constant, it will be the least significant variables, if any, which will be dropped from the model.

8.9 Calculation of revised estimates of the standard errors of the parameter (β) estimates, *t*- and *p*-values imposing an exchangeable correlation structure calculated using the variance inflation factor method: all mentions

Now the variance inflation factor=0.85684

The standard error, *s*, of each β-estimate is divided by $\sqrt{VIF}$=0.92566 and the *p*-values calculated.

A 2-tailed *t*-test is used for the regression, since the β-estimates can be positive or negative. The original and new values for the standard error of β, *t* and *p* are given in the Table 8.1.

VARIABLE	BETA	SE_OLD	T_OLD	PROB_OLD	SE_NEW	T_NEW	PROB_NEW
INTERCPT	2344.83	266.143	8.81042	0.00000	287.517	8.15545	0.00000
NSTART	-0.15	0.029	-5.28143	0.00000	0.031	-4.88881	0.00000
DURRAN	-0.44	0.104	-4.22610	0.00003	0.112	-3.91193	0.00010
DUMCGRP1	-535.11	149.114	-3.58863	0.00036	161.089	-3.32185	0.00095
DUMCGRP2*	23.51	253.833	0.09262	0.92624	274.219	0.08573	0.93171
DUMENG	1048.55	364.272	2.87849	0.00415	393.527	2.66451	0.00793
PRESENTD	-637.95	118.341	-5.39076	0.00000	127.844	-4.99001	0.00000
DUMCLN1	444.48	149.161	2.97985	0.00301	161.140	2.75833	0.00600
DUMCLN2	438.23	145.795	3.00575	0.00277	157.504	2.78231	0.00558
DUMANS	523.72	122.078	4.29001	0.00002	131.882	3.97109	0.00008
DEVLPNG	2182.69	423.091	5.15890	0.00000	457.070	4.77539	0.00000
NTREP	243.98	51.860	4.70453	0.00000	56.025	4.35480	0.00002
NRREP	-91.03	29.393	-3.09703	0.00205	31.753	-2.86680	0.00430
IMPACT	-25.62	11.417	-2.24402	0.02522	12.334	-2.07720	0.03823
DUMPUBT3	-354.65	158.095	-2.24328	0.02527	170.792	-2.07652	0.03830
DUMMULT	813.22	163.454	4.97524	0.00000	176.581	4.60538	0.00001
ICGMUL2	1226.46	355.925	3.44585	0.00061	384.509	3.18968	0.00150

Indicator variables

DUMMULT = 1 if MULTIC = Yes
DUMMULT = 0 if MULTIC = Limited or No

DUMCGRP1 = 1 if CGROUP = Europe
DUMCGRP1 = 0 if CGROUP = America or Other

DUMCGRP2 = 1 if CGROUP = Other
DUMCGRP2 = 0 if CGROUP = America or Europe
* This main term was not significant before the exchangeable correlation matrix was imposed. It was retained because of the significance of the interaction term.

DUMENG = 1 if ENGLISH = A (i.e. published in a language other than English, with an English abstract)
DUMENG = 0 if ENGLISH = E (i.e. published in full in English)

DUMCLN1 = 1 if CLNSG = Yes
DUMCLN1 = 0 if CLNSG = No, Possibly or not reported

DUMCLN2 = 1 if CLNSG = not reported
DUMCLN2 = 0 if CLNSG = Yes, No, Possibly

DUMANS = 1 if ANSWER = Yes or No
DUMANS = 0 if ANSWER = not reported

PUBTYPE3 = 1 if PUBTYPE = Meeting abstract
PUBTYPE3 = 0 if PUBTYPE = Journal or Book

Interaction term

ICGMUL2 = 1 if non-US trial with less than 5 centres
ICGMUL2 = 0 otherwise

Table 8.1 Revised estimates of the standard errors of the parameter (β) estimates, *t*- and *p*-values imposing an exchangeable correlation structure calculated using the variance inflation factor method: all mentions

All variables significant in the independence model remained significant. None of the terms in the independence model were borderline, so applying the variance inflation factor was unlikely to result in terms being dropped. However, this may not be the case with some of the other analyses.

The algorithm for the program to calculate: the correlation coefficient ρ for a matrix with an exchangeable correlation structure, the variance inflation factor and revised *t*-statistics and *p*-values for parameter estimates is given as Appendix X (ii). The output, from each of the six analyses is given as Appendix XII.

The method of variance inflation factor with an exchangeable correlation structure is not satisfactory for use with these data. The assumption that all correlation coefficients are the same is not tenable in the light of the correlation structure which emerged from the residuals of the independence model. However, it is worth doing in order to indicate that it is unlikely that a repeated measures approach will change the basic model.

8.10 Generalised estimating equations (GEE)

An alternative method is that of generalised estimating equations (GEE), a type of mixed effect model [Zeger and Liang (1986)]. This can be used with other correlation structures such as the stationary *m*-step dependent and autoregressive structures and so is more suited for use with these data.

GEE use an iterative procedure to calculate the correlation coefficient, ρ. Initially ρ is set to zero for the preliminary analysis. The residuals are used to calculate the first estimation of ρ. The regression is then re-run, this time incorporating ρ. New residuals are obtained and hence a second, more accurate, estimate of ρ. The process is repeated until ρ convergence is adequate. As well as estimating ρ the GEE produces the new parameter estimates, standard errors, *t*-statistics and *p*-values.

A limitation of GEE is that they may not converge well on small datasets. Two of the six datasets cannot be regarded as large, these being 'time from close of randomization to submission', and 'time from receipt to publication', for all records which report results, which use 129 and 121 observations respectively. However the response variables are normally distributed, which is an advantage when using GEE because the likelihood is then exact.

The results of applying GEE to the data using the stationary *m*-step dependent and autoregressive correlation structures will be discussed in Chapter 9. Before this, a preliminary investigation using these structures is undertaken. This uses a method similar to that for calculating the variance inflation factor, used in estimation of the correlation coefficient for an exchangeable correlation structure. This is in fact the first approximation to the correlation structure that the GEE method will use. There is an argument that for small datasets, except for the case of the exchangeable correlation structure, the process should not be continued past this first step. This is because convergence problems may occur, producing unreliable results due to flat likelihoods, and so for small datasets this section of the procedure alone may be optimal. Whatever the size of dataset, this preliminary investigation is essential as a check that the extra parameterisation introduced is not producing unreliable results, due to convergence problems, when the full GEE is used.

8.11 A preliminary investigation to impose a stationary *m*-dependent correlation structure: all mentions

The second correlation structure tried was the stationary *m*-dependent. This requires that a separate correlation coefficient be calculated for each diagonal. The formula used to calculate these is that used to estimate the correlation coefficient for the exchangeable correlation structure. However, here it is applied separately to each diagonal and so four different correlation coefficients result and four new parameters are introduced. As with the exchangeable correlation structure, the residuals from the independence model are used to calculate the *m*-step coefficients, from which the correlation matrix is formed. This is, in effect, reproducing the first iteration of the GEE procedure. The algorithm for the program to do this is given as Appendix X (iii), and the correlation

coefficients estimated for the four analyses for which there are sufficient observations are given in Appendix XIII.

The estimates for the correlation coefficients for the analysis using the largest dataset, 'time from close to publication for all mentions' are as follows:

$\rho_1 = 0.51840$ $\rho_2 = 0.31154$ $\rho_3 = 0.18015$ $\rho_4 = -0.05110$

where ρ_d = the correlation coefficient for all elements distance *d* from the leading diagonal.

8.12 The Mantel test

This correlation matrix was then tested against the matrix formed from the residuals from the independence model using the Mantel test [Manly (1986)], a permutation test. The program for applying the Mantel test is given as Appendix XIV. However a brief description of the process follows:

Let **M** be the correlation matrix formed from the residuals from the independence model.
Let **E** be the matrix constructed using the correlation coefficients estimated, using a stationary *m*-dependent correlation structure:

$$\mathbf{E} = \begin{bmatrix} 1.00000 & 0.51840 & 0.31154 & 0.18015 & -0.05110 \\ 0.51840 & 1.00000 & 0.51840 & 0.31154 & 0.18015 \\ 0.31154 & 0.51840 & 1.00000 & 0.51840 & 0.31154 \\ 0.18015 & 0.31154 & 0.51840 & 1.00000 & 0.51840 \\ -0.05110 & 0.18015 & 0.31154 & 0.51840 & 1.00000 \end{bmatrix}$$

A Mantel's *Z* value is then calculated from the elements of matrices **M** and **E** using the formula:

$$Z = \sum_{i=2}^{n} \sum_{j=1}^{i-1} m_{ij} e_{ij}$$

Z is the sum of the products of the elements in the lower diagonal parts of the matrices **M** and **E**.

Then the 120 possible permutations of the elements of the estimated matrix **E** are obtained. For each a *Z*-value is calculated using a permutation of **E** and the unpermutated original matrix **M**. In this way a distribution of *Z* is obtained.

It is the position of the *Z*-statistic for matrix **E** within the distribution of the *Z*-values which is of interest, rather than the value of *Z*. If there is a positive correlation between the estimated matrix **E** and the original matrix **M** the *Z*-value obtained will tend to be greater than others in the distribution. If matrices **M** and **E** are not correlated then the *Z*-value would be expected to be typical of the distribution.

Therefore the null hypothesis can be stated as:

> *H_0: Estimated matrix **E** is no better correlated with matrix **M** than random matrices made from the permutations of elements of **E**.*

The proportion of permutations with a *Z*-statistic greater than or equal to that for the estimated matrix is equal to the *p*-value.

For the 'time from close to publication for all mentions' analysis, the *Z*-value for the unpermutated matrix **E** with **M** is 1.54851, which is high compared to that for other permutations of **E**. Of the 120 *Z* values obtained, that for the unpermutated estimated matrix **E** was joint highest with one other. Due to the symmetrical nature of the matrix **E** and its permutations there are two of every *Z*-value produced i.e. 60 pairs. Therefore the significance of this result is p=2/120=0.01667, which is sufficient to reject H_0. Thus the estimated matrix **E** can be considered a good approximation to **M**.

Next the correlation between the lower elements of matrices **M** and **E** (Pearson's correlation coefficient) was calculated.

$$r = \frac{Z - n(n-1)\,\bar{m}\,\bar{e}\,/2}{\sqrt{\left(\sum_{i=2}^{n}\sum_{j=1}^{i-1} m_{ij}^2 - n(n-1)\,\bar{m}^2/2\right)\left(\sum_{i=2}^{n}\sum_{j=1}^{i-1} e_{ij}^2 - n(n-1)\,\bar{e}^2/2\right)}}$$

$\bar{m}$ = mean of the m_{ij} values
$\bar{e}$ = mean of the e_{ij} values
$n \times n$ = dimensions of matrices **M** and **E**
$n(n-1)/2$= number of lower diagonal elements in matrices **M** and **E**

The correlation coefficient r = 0.84893, a high positive correlation.

8.13 A preliminary investigation to impose an autoregressive correlation structure: all mentions

Lastly an autoregressive correlation structure was tried. Here, elements distance 1 from the leading diagonal have the value ρ, those distance 2 from the leading diagonal have value ρ^2, those distance 3 have value ρ^3 and so on. The advantage of this over the stationary *m*-dependent structure is that only one new parameter is introduced rather than four. If the correlation between this structure and the original covariance matrix is almost as good as that between the stationary *m*-dependent structure and the original covariance matrix, then this may be the structure to use. However, a disadvantage of the autoregressive structure is that it does not allow a limited number of non-zero coefficients to be set.

The estimate of the correlation coefficient for elements distance 1 from the leading diagonal was obtained in the same way as for the stationary *m*-dependent structure, and so has the same value, $\rho = 0.51840$.

The estimates for the correlation coefficients for the other elements are as follows:

distance 2 from the leading diagonal: $\rho^2 = 0.26874$

distance 3 from the leading diagonal: $\rho^3 = 0.13931$

distance 4 from the leading diagonal: $\rho^4 = 0.07222$

Mantel's test gives $Z = 1.48336$ (p=0.01667 using 120 observations).

Pearson's correlation coefficient between matrix ***E*** estimated using an autoregressive structure and matrix ***M*** is $r = 0.82115$.

8.14 Conclusion: all mentions

As expected, both models are good approximations to the matrix formed from the residuals from the independence model, **M**. Estimated matrix **E** using an autoregressive structure is not as highly

correlated with **M** as **E** using the stationary *m*-dependent structure. However this is offset against the need for only one new parameter rather than *m*.

In matrix **M**, the correlation coefficients distance 3 or more from the leading diagonal are small and not statistically significant, and those distance 1 are both greater and far more significant than those distance 2, suggesting that the stationary *m*-dependent structure with either *m*=2 or *m*=1 is the most appropriate. Since the autoregressive structure does not allow the setting of all coefficients distance 3 and greater to be zero this is less suitable.

8.15 The correlation matrix for the residuals from the independence model analysis of time from close to publication for all results

The investigations performed on the largest dataset, where the response variable is 'time from close of randomization' and all records are used, are now repeated for the second largest dataset, that where the response variable is the same but where only records reporting results are used. Again see Appendix XI for the correlation matrix obtained from the residuals from the independence model.

All the 1-step correlation coefficients are highly significant:

ρ_{12} =0.53121 p<0.0001 (n=91)

ρ_{23} =0.75606 p<0.0001 (n=54)

ρ_{34} =0.73189 p<0.0001 (n=29)

ρ_{45} =0.89130 p<0.0001 (n=13)

The 2-step correlation coefficients are less significant and less correlated:

ρ_{13} =0.56003 p<0.0001 (n=54)

ρ_{24} =0.38425 p=0.0396 (n=29)

ρ_{35} =0.63482 p=0.0198 (n=13)

The 3-step correlation coefficients are:

ρ_{14} = 0.08945 p=0.6445 (n=29) n/s

ρ_{25} = -0.10232 p=0.7394 (n=13) n/s

The 4-step correlation coefficient is:

ρ_{15} = -0.13473 p=0.6608 (n=13) n/s

The pattern of the correlation matrix for the 'results only' dataset with response variable 'close to publication' is similar to when all records are used, except the 1- and 2-step correlation coefficients here are greater. Again, correlations and significance levels decrease with distance apart. Also the numbers of observations on which the coefficients are calculated diminish sharply.

8.16 Estimation of the correlation coefficient for an exchangeable correlation structure and revised standard errors of the parameter (β) estimates, *t*- and *p*-values using the variance inflation factor method: all results

The correlation coefficient for an exchangeable correlation structure for the 'time from close to publication for all results' analysis was estimated as ρ_{est} =0.55167.

Hence the variance inflation factor (VIF) was calculated to be 0.69566.

Dividing the standard error of each β estimate by $\sqrt{VIF}$ = 0.83406 does increase the p-values considerably, proclaiming a weakening in the evidence for variable number of questions (NOQ), as shown in Table 8.2.

Note that main terms specifying direction of results (DUMDIR1 and DUMDIR2) were not significant in the independence model and are retained only due to the significance of their interaction with statistical significance (LOGPEST).

VARIABLE	BETA	SE_OLD	T_OLD	PROB_OLD	SE_NEW	T_NEW	PROB_NEW
INTERCPT	2036.56	301.806	6.74791	0.00000	361.852	5.62816	0.00000
NOQ	47.86	20.309	2.35650	0.01899	24.350	1.96546	0.05014
NCLOSE	−0.10	0.027	−3.77133	0.00019	0.032	−3.14552	0.00180
PRESENTD	−461.68	141.864	−3.25439	0.00125	170.089	−2.71436	0.00697
DUMCLN2	−405.87	151.038	−2.68719	0.00755	181.088	−2.24128	0.02563
DUMANS	416.20	145.726	2.85604	0.00454	174.719	2.38211	0.01774
DUMFL	692.59	259.924	2.66458	0.00806	311.637	2.22242	0.02689
NTREP	136.21	50.084	2.71971	0.00686	60.049	2.26840	0.02391
LOGPEST	330.63	83.769	3.94693	0.00010	100.436	3.29198	0.00110
DUMDIR1	276.15	171.197	1.61308	0.10762	205.257	1.34541	0.17936
DUMDIR2	339.43	220.885	1.53669	0.12527	264.831	1.28169	0.20079
LPDIR1	−390.87	103.790	−3.76600	0.00019	124.439	−3.14107	0.00183
LPDIR2	−386.69	145.513	−2.65741	0.00823	174.464	−2.21644	0.02730

Indicator variables

DUMCLN2 = 1 if CLNSG = No or Possibly
DUMCLN2 = 0 if CLNSG = Yes or not reported

DUMANS = 1 if ANSWER = Yes or No
DUMANS = 0 if not reported

DUMFL = 1 if FIRSTL = Treatment for relapse or refractory disease
DUMFL = 0 if FIRSTL = First-line treatment

DUMDIR1 = 1 if POSNG = Positive
DUMDIR1 = 0 if POSNG = Negative, Null, Opposite or not reported

DUMDIR2 = 1 if POSNG = Negative
DUMDIR2 = 0 if POSNG = Positive, Null, Opposite or not reported

Interaction terms

LPDIR1 = LOGPEST * DUMDIR1
LPDIR2 = LOGPEST * DUMDIR2

Table 8.2 Revised estimates of the standard errors of the parameter (β) estimates, *t*- and *p*-values imposing an exchangeable correlation structure calculated using the variance inflation factor method: all results

8.17 A preliminary investigation to impose a stationary *m*-dependent correlation structure: all results

The estimates for the correlation coefficients for the analysis using the 'time from close to publication for all results' dataset were calculated as:

$$\rho_1 = 0.66706 \quad \rho_2 = 0.52135 \quad \rho_3 = 0.03624 \quad \rho_4 = -0.13473$$

where ρ_d = the correlation coefficient for all elements distance *d* from the leading diagonal.

Hence Mantel's *Z* value = 2.78240 (*p*=0.01667 using 120 observations) and the correlation between the lower elements of the matrix formed from the residuals from the independence model and those from the estimated *m*-dependent correlation structure is $r = 0.94739$.

8.18 A preliminary investigation to impose an autoregressive correlation structure: all results

Since the correlation coefficients for all elements distance 1 from the leading diagonal were calculated as $\rho = 0.66706$, those for elements distance 2, 3 and 4 from the leading diagonal were set to $\rho^2 = 0.44497$, $\rho^3 = 0.29682$ and $\rho^4 = 0.19800$ respectively.

Hence Mantel's Z value =2.61361 (p=0.01667 using 120 observations) and the correlation between the lower elements of the matrix formed from the residuals from the independence model and those from the estimated m-dependent correlation structure is $r = 0.90111$.

8.19 Conclusion: all results

The results are similar to those for the 'all mentions' dataset. Again both models are good approximations to the correlation matrix formed from the residuals from the independence model, **M**. The correlation of **M** with the matrix **E** having a stationary m-dependent structure is slightly higher and slightly more significant, than that with matrix **E** having an autoregressive structure. Again, in matrix **M** only the correlation coefficients distance 1 and 2 from the leading diagonal are statistically significant, the former much more so that the latter, indicating that the stationary m-dependent with m=2 or perhaps m=1 is the correlation structure of choice.

8.20 An investigation into the many-to-one relationship between randomizations and trials

Having established that the incorporation of repeated measures analysis will solve the problem of non-independence of records due to multiple reportings of the same randomization, there remains a further problem. There is clustering, since several randomizations may belong to the same trial. This applies to all twelve analyses, not just the six, which have repeated measures per randomization.

To investigate whether this has an important effect on the response, correlation matrices were again produced for the residuals from the analyses on the two largest datasets, but this time by trial rather

than randomization. The object was to compare these by eye with those for the randomizations. If the patterns are very similar, the randomization-trial clustering can be ignored. The method used is approximate and the results are not stated, just the overall impression.

For the largest dataset, that for analysing time from close to publication for all mentions, the correlation matrix obtained for trials echoes that for randomizations although higher correlations remain at the 3rd and 4th steps. The associated significance levels are also higher. So correlations extend further back with the trials correlation matrix than with the randomizations correlation matrix. However since the patterns of the two matrices are similar, the randomizations-trial clustering can be safely ignored.

When applied to the second dataset, that for analysing close to publication for all reportings of results, the trials correlation matrix produced was very similar to that for randomizations. Therefore, again, the clustering of randomizations-trial can be safely ignored.

9 APPLICATION OF GENERALISED ESTIMATING EQUATIONS (GEES) TO THE LARGEST DATASET

9.1 Introduction

This chapter begins by describing the full method of generalised estimating equation (GEE) analysis for use with correlated data. This can be used to obtain not only the coefficients for any correlation structure, but also the revised *t*-statistics and *p*-values for the parameter estimates. Firstly the general model is discussed and then its application to these particular data. This method is then applied to the analysis using the largest of the six datasets, 'time from close to publication for all mentions', imposing different correlation structures in turn. There follows a section on how, for each variable in the model, the *p*-value alters when different structures are applied. Diagnostic plots from the analyses with different correlation structures imposed are also compared. Finally, a linear mixed effects model is run, in order to confirm the findings from the GEE analyses.

9.2 Generalised linear models (GLM)

The notation established here will be used for the section on GEE (Section 9.3).

Let Y be a random variable whose probability function depends on a single parameter, θ. If the distribution of Y can be written in the form

$$f(y;\theta) = exp\,[a(y)b(\theta) + c(\theta) + d(y)], \qquad \text{where } a, b, c \text{ and } d \text{ are functions,}$$

then the distribution of Y belongs to the exponential family.

If $a(y)=y$ then the distribution is in the canonical form and $b(\theta)$ is the natural parameter of the distribution. Any other parameters in the expression are treated as nuisance parameters. The Poisson, normal, and binomial distributions all belong to the exponential family and can be written in the canonical form. [McCullagh and Nelder (1986), Dobson (1990)]

In generalised linear models the maximum likelihood estimates are obtained for the parameters using an iterative procedure.

In generalised linear models the link function is defined as

$$g(\mu_i) = \alpha_i + \sum_{j=1}^{p} \beta_{ij} x_{ij}$$

$$= \mathbf{X_i \beta} \text{ for the } i^{th} \text{ observation.}$$

The link function describes the model used.

Examples

- *Normal distribution*

  For the i^{th} observation, $E[Y_i] = \mathbf{X_i \beta}$

  Therefore $g(\mu_i) = \mu_i$

  i.e. the link function for the ordinary linear regression, which is used to model data where the response variable has a normal distribution, is the identity function.

- *Logistic regression*

  For the i^{th} observation, $log\,[p_i / (1-p_i)] = \mathbf{X_i \beta}$,

  where p_i = the probability of success of 1 trial.

  i.e. where $P(X_i=1) = p_i$ and $P(X_i=0) = (1- p_i)$

  and $E(X_i) = p_i = \mu_i$

  Therefore $g(\mu_i) = logit(\mu_i)$

  i.e. the link function for the logistic regression model is the *logit* function. The logistic regression model will be used in Chapter 14 to model whether or not a randomization is ever published.

- *Poisson regression*

  For the i^{th} observation, the parameter of the Poisson distribution is modelled by

  $$log(\mu_i) = \alpha_i + \sum_{j=1}^{p} \beta_{ij} x_{ij} = \mathbf{X_i \beta},$$

  where μ_i is the Poisson parameter for the i^{th} observation.

  Therefore $g(\mu_i) = log(\mu_i)$

  i.e. the link function for the Poisson regression model is the logarithmic function. The Poisson model will be used in Chapter 15 for modelling frequency of publication.

9.3 The GEE method

Specify the link function, g (μ_i)

In GEE the link function is the same as for the GLM, as described in Section 9.2.

Specify the mean-variance relationship

Generalised estimating equations use a quasi-likelihood which has the form of a normal likelihood, with the variance expressed as a function *h,* say, of the mean. The variance of Y_i , is written in the form $V(Y_i) = h(\mu_i) / \varphi$, where φ is a scale parameter, which is treated as a nuisance parameter. A quasi-likelihood is used because the exact likelihood function cannot be written down in non-independent cases, except for the normal distribution. Non-normality is accounted for by incorporating a relationship between the variance and the mean. [Zeger SL and Liang K (1986)]

Examples

- *Poisson distribution*

  If $E(Y_i) = \mu_i$ then $V(Y_i) = \mu_i$, and so $h(\mu_i) = \mu_i$

- *Normal distribution*

  $E(Y_i) = \mu_i$

  $V(Y_i) = \sigma^2$, which is constant and so independent of μ_i

  In the case of the normal distribution, there is no relationship between the mean and the variance and the likelihood is exact. In fact, in this thesis, GEE are only used for normally distributed data.

Specify the correlation structure

As described in Section 8.5, possible correlation structures are independence, unstructured, exchangeable, stationary *m*-dependent and autoregressive.

In the independence model (See Section 5.4)

$$L = 1/\sqrt{(2\pi\sigma^2)}^{n} \, exp \, [-\sum_{i=1}^{n} \varepsilon_i^2 / 2\sigma^2]$$

where $\varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \mu_i)^2$

$$= [y_1-\mu_1 \quad y_2-\mu_2 \quad \ldots \quad y_n-\mu_n] \begin{matrix} y_1-\mu_1 \\ y_2-\mu_2 \\ \\ .. \\ y_n-\mu_n \end{matrix}$$

$$= \mathbf{(y-\mu)^T (y-\mu)}$$

So $L = 1/\sqrt{(2\pi\sigma^2)^n}\, exp\, [-\mathbf{(y-\mu)^T I (y-\mu)}/2\sigma^2]$

For the non-independence model, the likelihood function is

$$L^* = 1/\sqrt{(2\pi\sigma^2)^n}\, exp\, [-\mathbf{(y-\mu)^T S^{-1} (y-\mu)}/2\sigma^2],$$

where **S** is the variance/covariance matrix for the residuals.

For non-normal data, this likelihood is used as a quasi-likelihood, non-normality being accounted for by the mean-variance relationship, i.e. if E $(Y_i) = \mu_i$ then $V(Y_i) = h(\mu_i) / \varphi$.

Using imposed correlation structure **R**, model parameter estimates with their variances and covariances are obtained by using the quasi-likelihood equations.

In the special case of normal data, the quasi-likelihood is the likelihood proper

$$L_Q = 1/\sqrt{(2\pi\sigma^2)^n}\, exp\, [-\mathbf{(y-\mu)^T S^{-1} (y-\mu)}/2\sigma^2] ,$$

where the covariance matrix for the i^{th} group is $\mathbf{S_i = V_i^{½} R_i V_i^{½}}$ where $\mathbf{R_i}$ is the correlation matrix and $\mathbf{V_i^{½}}$ is a diagonal matrix with j^{th} entry $\sqrt{[V(\mu_{ij})/\varphi]}$

L_Q is maximum when $\mathbf{(y-\mu)^T S^{-1} (y-\mu)}$ is minimum.

So to maximize the quasi-likelihood, this expression is differentiated with respect to α and β_j where $j = 1 \ldots p$ (p+1 parameters)

$$\partial/\partial\beta_j [\mathbf{(y-\mu)^T S^{-1} (y-\mu)}] = 0$$

$$-\partial\boldsymbol{\mu}^{\mathbf{T}}/\partial\beta [\mathbf{S^{-1} (y-\mu)}] + [\mathbf{(y-\mu)^T S^{-1}}. - \partial\boldsymbol{\mu}^{\mathbf{T}}/\partial\beta] = 0$$

$$-2\, \partial\boldsymbol{\mu}^{\mathbf{T}}/\partial\beta [\mathbf{S^{-1} (y-\mu)}] = 0$$

which gives the generalized estimating equations

$$\partial\boldsymbol{\mu}^{\mathbf{T}}/\partial\beta [\mathbf{S^{-1} (y-\mu)}] = 0$$

Now the link function is $g(\mu_i)$ where

$$g(\mu_i) = \alpha_i + \sum_{j=1}^{p} \beta_{ij} x_{ij} \ , i = 1 \ldots n$$

Therefore

$$\mu_i = g^{-1} [\alpha_i + \sum_{j=1}^{p} \beta_{ij} x_{ij}], i = 1 \ldots n$$

where $g^{-1}(\alpha_i, \beta_{i1,} \beta_{i2, \ldots} \beta_{ip})$ is the inverse link function

Substitute the inverse of the link function for μ_i in the equations and solve for α and β_j, j=1 … p, and their variances and covariances.

These equations must be solved iteratively.

The iterative process is as follows:

- Corrected parameter estimates α and β_j, j=1 … p, are obtained and hence new residuals
- Hence a corrected estimation of correlation matrix **R** is calculated. The structure of **R** chosen is kept the same, but the coefficients are corrected.
- Hence the corrected variance/covariance matrix **S** is calculated
- The inverse of the corrected matrix **S**, $\mathbf{S^{-1}}$ is calculated and substituted into the quasi-likelihood expression.
- The new generalized estimating equations are the solved.

The first iteration was the independence model.

This process is repeated until convergence is reached for the parameter (β) estimates, their variances and covariances and the correlation matrix estimate **R**.

9.4 Application of the GEE method

Before applying GEE analysis, the independence model was run in order to obtain the residuals which were used in an exploratory procedure for investigating plausible structures for the correlation matrix **R**. This was described in Chapter 8. Those structures were then imposed upon **R** and the GEE analysis was run.

9.4.1 Application of the GEE method to these particular data

Link function

In each of the analyses, the 'time to' response variable is approximately normally distributed, as was indicated by the diagnostic scatter plot of residuals versus their normal scores. An example is given as Figure 6.2. Therefore the link function used in the GEE analysis is the identity.

The mean-variance relationship

Since the response variable is normally distributed the variance is not a function of the mean, and the assumption of constant variance is satisfied. The latter was confirmed by the diagnostic plot of standardised residuals versus fitted values. An example is given as Figure 6.1. It is a reasonable assumption that the variance for the first publications is the same as that for the second, third and so on. Since the response is normal, the likelihood is exact.

Specifying the correlation structure

The preliminary investigations in Chapter 8 suggested that the stationary m-dependent m=2 and m=1 structures are likely to be most suitable for these data. This is what one might intuitively expect. However, for comparison only the exchangeable, stationary m-dependent m=3 and autoregressive will also be tried. In addition the independence model will also be run in order to check that the findings are the same as those from the preliminary model, which were reported in Section 7.6.

9.4.2 Using SAS to run GEE analysis

The GENMOD procedure in SAS fits generalized linear models. One of its applications is to use generalised estimating equations (GEE) to fit models to correlated data resulting from repeated measures. The GEE method is recommended for use with files containing a large number of fairly small clusters where the correlation matrix itself is not the object of interest. In this case it is the parameter (β) estimates which are of interest. From here onwards the term 'β estimate' will be used

in preference to ‘parameter estimate’ since the coefficients of the correlation matrix are also parameters.

The SAS commands for running the GEE analysis are given as Appendix X (iv). This is run on the largest dataset, that for the ‘time from close to publication for all mentions’ analysis. In the REPEATED statement the correlation structure is specified. Options include independent, unstructured (not applicable), exchangeable, stationary *m*-dependent and autoregressive correlation structures. It is also possible to specify the actual correlation coefficients of the matrix, although this option is not used. The clustering variable, the randomization ID (RANID), is also specified. The program requests that the working correlation matrix is printed out and that the *Z*- and *p*-values obtained using standard error estimates based on Fisher’s (expected) information are given. Estimates based on observed information are also obtained for the purpose of confirming the findings from the former, and are discussed briefly in Section 11.8.

9.5 Application of GEE analysis to the largest of the six datasets

The dataset used is that for ‘time from close to publication for all mentions’.

Generalised estimating equation analysis was used with the following structures in turn:

- Independence
- Exchangeable
- Autoregressive
- Stationary *m*-dependent with *m* set to 3, 2 and 1 respectively.

From the preliminary investigations the most likely correlation structures appear to be the stationary *m*-dependent, with $m = 1$ or 2. The generalised estimating equation analysis was run on the others as well for comparison only. Although the preliminary investigations suggested that the autoregressive correlation structure might be plausible, since this involved the introduction of a single new parameter, ρ, as opposed to *m* new parameters for the stationary *m*-dependent correlation structure, this was based on a 5×5 working correlation matrix, with 4 bands, plus the

leading diagonal of unity in each case. The number of bands can be specified when the stationary *m*-dependent correlation structure is imposed, setting all other correlation coefficients to zero. However, for all other correlation structures i.e. for the unstructured, exchangeable and autoregressive structures, the working correlation matrix is forced to be the full 13×13 matrix.

To return to the original data; the correlation matrix for the residuals from the independence model indicated that the correlation coefficients between the response for consecutive reportings were strong and positive, and, more importantly, highly significant:

ρ_{12} =0.35271 p<0.0001 (n=128)

ρ_{23} =0.46667 p<0.0001 (n=90)

ρ_{34} =0.66755 p<0.0001 (n=68)

ρ_{45} =0.75220 p<0.0001 (n=44)

They became less strong and less significant for those publications distance 2 apart:

ρ_{13} =0.16092 p=0.1297 (n=90) n/s

ρ_{24} =0.40235 p=0.0007 (n=68)

ρ_{35} =0.45929 p=0.0017 (n=44)

Correlation coefficients for publications distance 3 or more apart were both weak and have *p*-values > 0.05 associated with them.

This would suggest that a correlation structure with a limited number of non-zero bands is appropriate. Since this is not so for the autoregressive correlation structure, the stationary *m*-dependent structure, with either m = 1 or 2, may be a better choice.

9.6 An investigation into how far the *p*-values of the parameter (β) estimates alter when different correlation structures are imposed

Appendix XV gives the output from the generalised estimating equation analysis applied to the largest dataset, with each correlation structure imposed. Each variable from the independence

model is taken in turn and the significance of its parameter (β) estimate, compared when each of the correlation structures is imposed. This listing is given as Appendix XVI.

There are three comparisons to be made:

- Whether there is a clustering effect, where each randomization is a cluster. This involves comparing the *p*-values when the exchangeable correlation structure is applied with those from the independence model.
- Whether there is a serial correlation effect. This is done by comparing the *p*-values from the stationary *m*-dependent and autoregressive correlation structures with those from the exchangeable one.
- Finally how much the *p*-values for the parameters vary, when the four correlation structures representing a serial correlation effect are applied.

The findings are as follows:

- Eight of the 17 variables were very highly significant, $p<0.0001$ in the independence model. Of these, four including the intercept remained so under all correlation structures imposed.
- Of the other 13 variables, the *p*-value of five remained roughly the same under the exchangeable, but increased under the stationary *m*-dependent and autoregressive structures.
- For six variables the *p*-value was largely unchanged no matter which correlation structure was used.
- For the only interaction term in the model, the *p*-value did not differ much between the exchangeable, stationary *m*-dependent and autoregressive models, but was greater for these than for the independence model.
- The final term was non-significant in the independence model and remained so whichever correlation structure was applied

- Notably, it was seen that for each variable, the results are similar whichever of the three most appropriate correlation structures, i.e. the autoregressive and *m*-step dependent with *m*=2 and *m*=1, are imposed.

9.7 Variables significant in the independence model but no longer significant once a correlation structure is imposed

All variables in the independence model remained significant no matter which correlation structure was imposed.

The variable, 'country group of trialists is Europe or Other, as opposed to North America', (CGROUP) was not statistically significant in the independence model (p=0.9262). However, this was retained due to the significance of the interaction between country group of trialist and whether participation was multi-centre or single-centre (p=0.0006). As expected, this main term remained non-significant when the various correlation structures were applied.

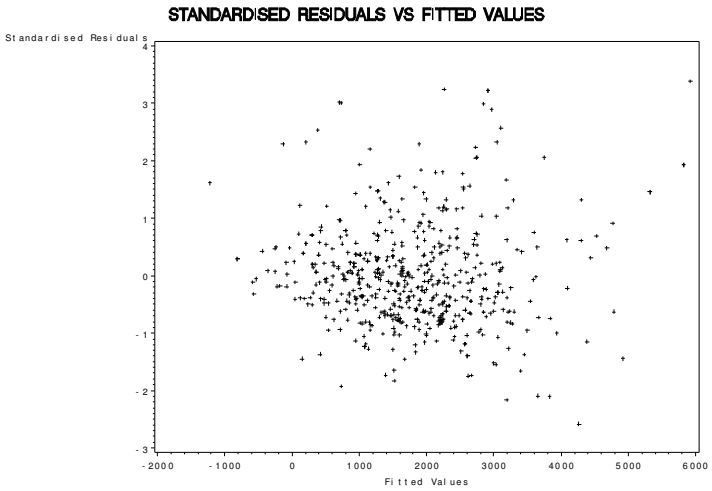
9.8 Comparing diagnostic plots for the various correlation structures

Standardised residuals versus fitted values

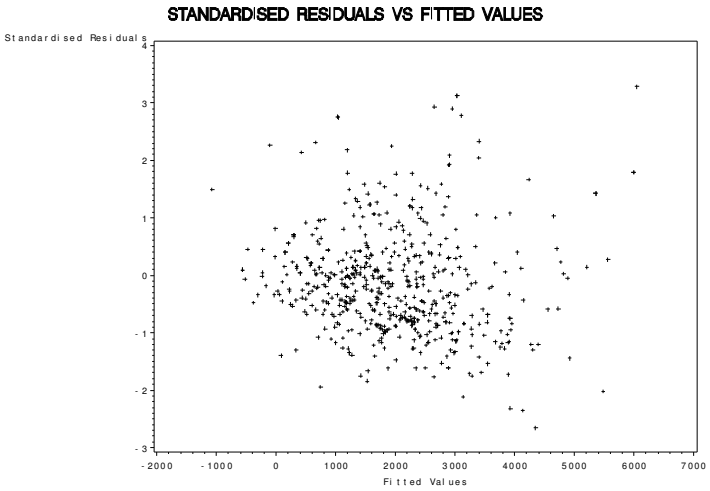
Whichever correlation structure is imposed, including the independence model, the plot of standardised residuals versus fitted values does not alter noticeably. The diagnostic plots using the independence, exchangeable and stationary *m*-step, *m*=2, dependent correlation structures are given in Figure 9.1.

The result in each case is a random spread, indicating constant variance, especially if those observations with a negative fitted value are ignored. It is reasonable to ignore them since this group of observations are those for which the date of publication is before the date of close of randomization i.e. those randomizations which are mentioned in a publication while the randomization is still open for accrual, or even prior to opening for accrual, and so are atypical. If these are left in, the spread is slightly more wedge-shaped.

(i)



(ii)



(iii)

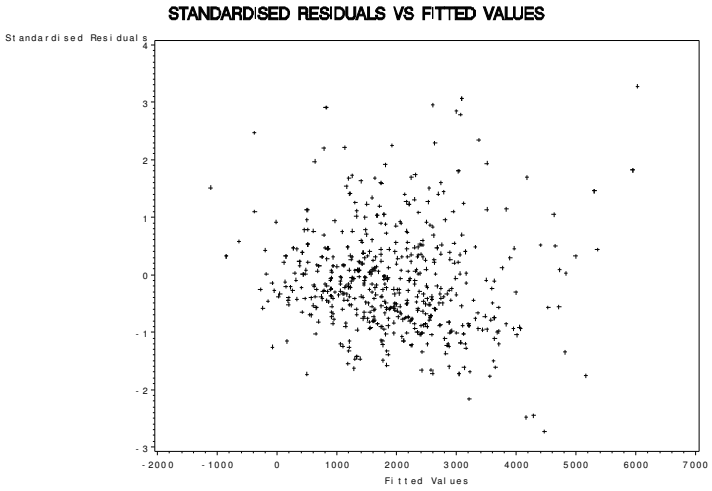
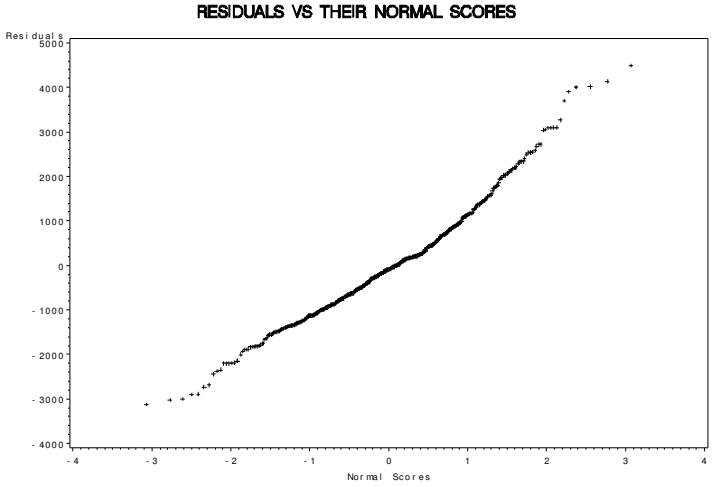
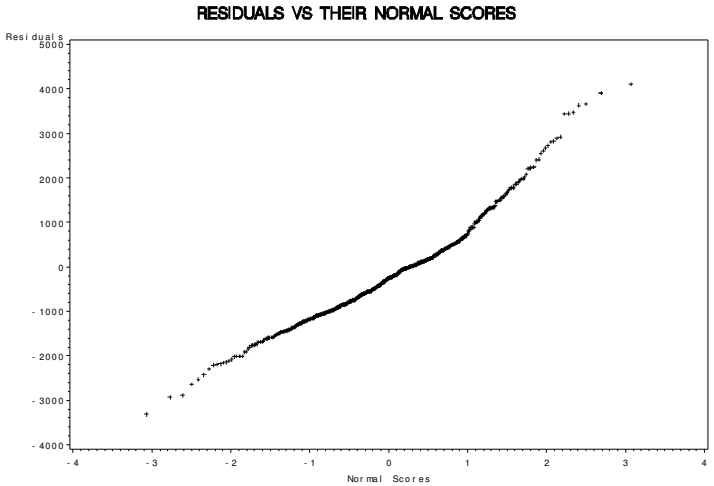


Figure 9.1 Time from close to publication for all mentions. Standardised residuals vs. fitted values: (i) independence model (ii) using generalized estimating equation analysis with an exchangeable correlation structure and (iii) with a stationary *m*-dependent, *m*=2, correlation structure

(i)



(ii)



(iii)

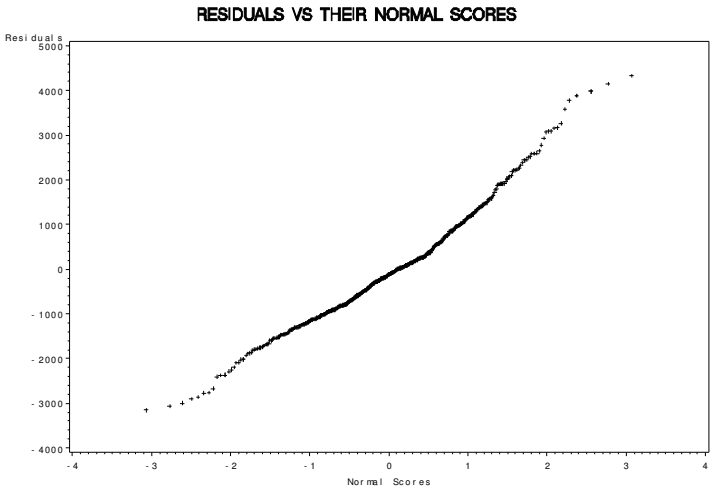


Figure 9.2 Time from close to publication for all mentions. Residuals vs. their normal scores
(i) independence model (ii) using generalized estimating equation analysis with an exchangeable correlation structure and (iii) with a stationary *m*-dependent, *m*=2, correlation structure

Residuals versus normal scores

The plots for residuals versus their normal scores for independence, exchangeable and stationary *m*-dependent, *m*=2, correlation structures are given in Figure 9.2. Again, whichever correlation structure is used, the residuals versus normal scores plot remains largely unaltered. The line is fairly straight, indicating a normal distribution with a slight kink for the largest residuals. The line is marginally straighter when either of the correlation structures are used, indicating a slightly better fit to the data, but the difference is negligible.

Conclusion

Whichever correlation structure is applied, both constant variance and normality are reasonable assumptions, as they were for the independence model.

9.9 The linear mixed effects model with grouped data

As described in Section 5.4, the ordinary linear regression model (independence model) can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$$

where $\boldsymbol{\beta}$ is the vector of fixed-effects parameter estimates

For the i^{th} group (randomization in this case)

$$\mathbf{y_i} = \mathbf{X_i}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$$

A linear mixed effects model for the i^{th} group is

$$\mathbf{y_i} = \mathbf{X_i}\boldsymbol{\beta} + \mathbf{Z_i}\mathbf{b} + \boldsymbol{\varepsilon}_i$$

where:

$\mathbf{b}$ is the unknown vector of random-effects such that $\mathbf{b} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_b)$, i.e. $\mathbf{b}$ is normally distributed with mean vector zero and variance/covariance matrix $\boldsymbol{\Sigma}_b$.

The columns of $\mathbf{Z}$ are usually a subset of those of $\mathbf{X}$, i.e. values of a subset of the explanatory variables, acting on each group separately.

$\boldsymbol{\varepsilon}$ is the vector of residuals and is generally of the form $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$.

Random effects $\mathbf{b}_i$ and residuals $\boldsymbol{\varepsilon}_i$ are assumed to be independent. However, a useful extension enables the modelling of non-constant variances and also special within-group correlation structures. In such cases $\boldsymbol{\varepsilon}$ is the vector of within-group errors and has the form $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, where the variance/covariance matrix $\boldsymbol{\Sigma}$ is to be estimated. The general case is the unstructured form. However, the same correlation structures as were used in GEE analysis can be imposed on the linear mixed effects models, and for the same reason – to avoid over-parameterisation.

9.10 Using a linear mixed effects model to confirm the findings from the GEE analysis

A linear mixed effects model was run using the MIXED procedure in SAS with the commands given in Appendix X (v). Different correlation structures were imposed and the β estimates, estimates of the standard errors of these and *p*-values obtained compared with those from the generalised estimating equations. Although the linear mixed effect model calculates the β estimates using a different method from the generalised estimating equations, the purpose of using this method in addition to GEE is as a mutual check. If the β estimates and the estimates of their standard errors found using the linear mixed effects model are very similar to those from the GEE analysis when a particular correlation structure is imposed, this gives more confidence that the correlation structure chosen is appropriate. The output from the program running the linear mixed effects model is given as Appendix XVII. The findings are summarized below.

Independence correlation structure

As expected, the β estimates, standard errors and *t* values obtained using the linear mixed effects model are identical to the β estimates, standard errors and *Z* values obtained using GEE. The *p*-values are almost identical, slightly higher in the linear mixed effects model, the difference negligible. The output from the linear mixed effects model is also identical to that from the GLM procedure used for the preliminary analysis. This provides a good starting point for any comparisons.

Exchangeable correlation structure

The β estimates, standard errors of these, *t*-values and *p*-values obtained using the linear mixed effects model are all very similar to the estimates obtained using GEE analysis. The *p*-values of all variables remain <0.05 with the linear mixed effects model also.

Autoregressive correlation structure

The β estimates, standard errors and *t*-values obtained using the linear mixed effects model are broadly similar to those obtained from the GEE analysis, but the comparative significance of the variables is not retained in all cases. 'Number of randomizations reported in a publication' (NRREP) has *p*-value 0.0375 when GEE are used but *p*=0.0700 under the linear mixed effects model. The *p*-values of the other variables remain <0.05 when the linear mixed effects model is applied.

Stationary m-dependent (m=3) correlation structure

Estimates from the linear mixed effects analysis differed substantially from those from the GEE analysis. Some variables of moderate or borderline significance when GEE were used were no longer significant when a linear effects model was used.

- Conducted in Europe (CGROUP): *p*=0.0060 increased to *p*=0.0736
- Article published in a language other than English (ENGLISH): *p*=0.0433 increased to *p*=0.3811
- Reported in articles that mention fewer randomizations (NRREP): *p*=0.0487 increased to *p*=0.1700
- Impact factor of journal (IMPACT): *p*=0.0099 increased to *p*=0.3162
- Publication type: meeting abstracts versus journal articles and book chapters (PUBTYPE3): *p*=0.0453 increased to *p*=0.1845

However, for 'indication as to whether the main questions, as stated in the paper, are answered in the paper' (ANSWER) the *p*-value of 0.0092 increased to only 0.0237.

So again, the comparative significance of the variables is not retained in all cases, between the independence model and when a correlation structure is imposed using GEE, between the independence model and when the correlation structure is imposed using a linear mixed effects model, or between the same correlation structure being imposed using GEE and a linear mixed effects model.

In the independence model, impact factor of journal (IMPACT) and publication type; meeting abstracts versus book chapters and journals (PUBYTPE3) had by far the largest *p*-values, p=0.0252 and 0.0253 respectively, apart from one of the indicator variables for country group (CGROUP)(p=0.9262), which was retained only due to the significance of its interaction with multi-centre versus single-centre participation (MULTIC). The two indicator variables representing clinical significance (CLNSG) are next, p=0.0030 and 0.0028. These remain significant under both the GEE and linear mixed effects models, p=0.0021 and 0.0314, and 0.0015 and 0.0302 respectively. Note how the *p*-value for the former has become slightly more significant under both, and the latter far less so.

Stationary m-dependent (m=2) correlation structure

The estimates obtained using the two procedures are broadly similar. The comparative significance of the variables when the GEE and the linear mixed effects methods are used was not retained in all cases.

When the linear mixed effects model was used, variable 'published in full in a language other than English' (ENGLISH) had *p*-value 0.1065, whereas with the GEE model p=0.0349. The *p*-values of all other variables remained at <0.05 under both procedures.

Stationary m-dependent (m=1) correlation structure

The estimates obtained using the two procedures are broadly similar. Again, the comparative significance of variables, between the GEE findings and the linear mixed effects model, is not retained for all variables. The results when the latter was used are almost identical to those using the independence model. *p*-values for all variables remained at <0.05 under both models.

Summary of findings

Running the linear mixed effects model broadly confirms the GEE findings when imposing a correlation structure of type exchangeable, autoregressive and stationary *m*-dependent with *m*=2 and with *m*=1.

10 APPLYING THE GENERALISED ESTIMATING EQUATION ANALYSIS WITH STATIONARY *M*-DEPENDENT, *M* = 2 AND *M* = 1, CORRELATION STRUCTURES TO ALL SIX DATASETS INVOLVING REPEATED MEASURES

10.1 Introduction

As discussed in the previous chapter the stationary *m*-dependent correlation structures with *m*=2 and *m*=1 appear to be most appropriate for the data. It was decided to apply both to the remaining five datasets and compare the results before deciding on the final model. The findings are summarized below, together with those for the largest dataset, on which the various correlation structures were piloted.

10.2 Time from close to submission for all mentions

For each variable the *p*-values obtained under the stationary *m*-dependent correlation structure with *m*=2 and with *m*=1 are very similar. All variables remain significant, i.e. $p<0.05$, under both models.

10.3 Time from receipt to publication for all mentions

Again, for each variable the *p*-values obtained under the stationary *m*-dependent correlation structure with *m*=2 and with *m*=1 are very similar. All variables remain significant under both models.

10.4 Time from close to publication for all mentions

There was very little difference between the *p*-values obtained for each variable, using the stationary *m*-dependent correlation structure with *m*=2 and with *m*=1. The *p*-values of all variables remained <0.05 [with the exception of country group of trialists (CGROUP): Other versus North America and Europe', which was non-significant under the independence model, and retained only due to the significance of its interaction with single-centre versus multi-centre participation (MULTIC)].

10.5 Time from close to submission for all results

Again, the *p*-values are very similar using the stationary *m*-dependent correlation structure with *m*=2 and with *m*=1. In both cases there is only marginal evidence for the retention of variable 'clinical significance' (CLNSG) (*p*=0.1051 and *p*=0.1030 respectively). Direction of results is of borderline significance (*p*=0.0450 and *p*=0.0449 respectively) but remains in the model.

The analysis, therefore, will be re-run omitting variable CLNSG.

10.6 Time from receipt to publication for all results

p-values were similar for most variables, but with a notable difference:

Under the stationary *m*-dependent correlation structure with *m*=2, presented at a major meeting (PRESENTD) is dropped from the model at *p*=0.1257 and impact factor of journal (IMPACT) is retained (*p*=0.0244).

Under the stationary *m*-dependent correlation structure with *m*=1, PRESENTD is retained, with *p*=0.0435. However the *p*-value for IMPACT increases to 0.0836, but will remain in the model due to the significance of its interaction with whether there was international participation or not (INTERNL).

Under the independence model, PRESENTD was retained (*p*=0.0113) and IMPACT was non-significant (*p*=0.1759), but again retained due to the significance of its interaction with INTERNL. Apart from IMPACT, PRESENTD was the least significant variable in the model.

If it is decided that the *m*=2 model is to be used, the analysis must be re-run, omitting variable PRESENTD.

10.7 Time from close to publication for all results

Again, when the stationary *m*-dependent, *m*=2 and *m*=1, correlation structures are applied the *p*-values obtained are similar. The *p*-values for the same five variables increase to >0.05. These are:

- Number of questions (NOQ): p=0.1771 and p=0.1566 respectively
- Clinical significance (CLNSG): p=0.1015 and p=0.0755 respectively
- First-line versus relapse treatment (FIRSTL): p=0.0817 and p=0.0753 respectively
- The two indicator variables representing direction of results (POSNG): p=0.2560 and p=0.2659 respectively, and p=0.0669 and p=0.1000 respectively.

When the independence model was used, the p-values for the two indicator variables representing POSNG were also >0.05, and were retained only because of the significance of their interaction with statistical significance (LOGPEST). Under the stationary m-step dependent correlation structures, m=2 and m=1, again these are retained only because of the significance of the interaction terms.

A backwards elimination procedure must now be used beginning by dropping the least significant main term, NOQ.

10.8 Conclusion

The two correlation structures, which looked most appropriate from the original data, the stationary m-dependent, with m=2 and with m=1, give very similar results. For the correlation matrices formed from the residuals from the independence model for the two largest datasets, the correlations for publications distance 1 apart are strong and highly statistically significant, and those distance 2 apart are less strong but still mostly statistically significant. Those distance 3 or more apart were neither strong nor statistically significant. Therefore it has been decided to use the m=2 correlation structure.

For the three analyses where imposing the stationary m-dependent, m=2, correlation structure resulted in one or more terms becoming no longer significant, a backwards elimination procedure was used to obtain a reduced set of significant variables. The three datasets in question are the three time period response variables for all reportings of results. The findings from all six analyses with repeated measures are described in the following chapter.

11 FINAL RESULTS FOR THE SIX REGRESSIONS INCORPORATING REPEATED MEASURES ANALYSIS USING GENERALISED ESTIMATING EQUATIONS (GEES)

11.1 Introduction

As discussed previously, six datasets contain more than one record for some randomizations, i.e. repeated measures for the same observation. This has now been compensated for by using generalised estimating equations, imposing a stationary m-dependent, m=2, correlation structure. The final results are now given and the output for these is in Appendix XVIII, with the exception of those for the analysis of time from close to publication for all mentions, which was given in Appendix XV. This chapter ends with a short section showing that whether the standard error estimates were obtained using the expected or observed information makes little difference to the outcome.

11.2 For all mentions of each randomization, which trial characteristics affect the time from close to submission?

Time from close to submission is longer for randomizations with the following characteristics:

- Shorter duration of accrual period (DURRAN) (p=0.0014), typical effect 1 year
- From a publication with no or a low impact factor (IMPACT) (p=0.0007), typical effect 1 year 7 months
- Results have smaller p-values associated with them (LOGPEST) (p=0.0002), typical effect 10 months
- Conducted in a 'developing' country (DEVLPNG) (p<0.0001), typical effect 12 years
- Published in an article which mentions a **greater** number of trials (LOGNTREP) (p=0.0356), typical effect 6 months
- Published in an article which mentions **fewer** randomizations (NRREP) (p=0.0012), typical effect 10 months

This is based on all 209 observations. The mean and range of response variable are approximately 5 years 3 months (–3 years 1 month to 26 years 0 months). The correlation coefficients used in the working correlation matrix were calculated as 0.5276 for those distance 1 from the leading diagonal and 0.4617 for those distance 2. This means that the model used assumes that, for any randomization, the correlation between the time from close to submission of one article and that for the next is 0.5276, the correlation between that for the article and the next-but-one is 0.4617 and that there is no correlation between time from close to submission for articles further apart.

11.3 For all mentions of each randomization, which trial characteristics affect the time from receipt to publication?

√(time from receipt to publication) is longer for randomizations with the following characteristics:

- Later start date of accrual period (NSTART) ($p<0.0001$)
- Published in a North American or European journal (JGROUP) ($p<0.0001$)
- Longer duration of accrual period (DURRAN) ($p=0.0002$)
- Reported in articles which mention fewer randomizations (NRREP) ($p<0.0001$)
- Conducted outside Europe (CGROUP) ($p<0.0001$)
- Larger number of co-authors (LOGAUTH/LOGMRC) ($p=0.0089$)

This is based on 195 out of 218 observations. The mean and range of time from receipt are approximately 9 months (1 month to 2 years 4 months). The correlation coefficients used in the working correlation matrix were calculated as 0.0006 for those distance 1 from the leading diagonal and -0.2442 for those distance 2.

11.4 For all mentions of each randomization, which trial characteristics affect the time from close to publication?

Time from close to publication of article is longer for randomizations with the following characteristics:

Main effects:

- Earlier start of accrual period (NSTART) (p<0.0001), typical effect 1 year
- Shorter duration of accrual period (DURRAN) (p=0.0008), typical effect 8 months
- Published in full in the English language (ENGLISH) (p=0.0349), typical effect 1 year 10 months
- Not presented at a major meeting (PRESENTD) (p<0.0001), typical effect 1 year 5 months
- Results are clearly reported as clinically significant (CLNSG) (p=0.0030), typical effect 1 year 2 months
- Clinical significance is not reported (CLNSG) (p=0.0169), typical effect 10 months

(A clearer interpretation of the above two findings is that results clearly reported as **not** or only possibly clinically significant are published fastest, followed by those where no indication of clinical significance is given, with results clearly reported as clinically significant having the longest time to publication.)

- A clear indication is given as to whether the main questions in the paper are answered in that paper (ANSWER) (p=0.0054), typical effect 9 months
- Conducted in a 'developing' country (DEVLPNG)(p<0.0001), typical effect 6 years 2 months
- Reported in articles which mention a **greater** number of trials (NTREP) (p=0.0002), typical effect 7 months
- Reported in articles which mention a **smaller** number of randomizations (NRREP) (p=0.0362), typical effect 4 months
- Reported in a publication with a low or no impact factor (IMPACT) (p=0.0128), typical effect 5 months
- Reported in a journal article or book chapter as opposed to a meeting abstract (PUBTYPE) (p=0.0316), typical effect 10 months
- Conducted outside Europe (CGROUP) (p=0.0076), typical effect 1 year 4 months

- Participation of five or more centres (MULTIC) (p=0.0001), typical effect 2 years 2 months

Interaction:

- However, for trials conducted outside North America, participation of less than five centres also increases time to publication (CGROUP*MULTIC) (p=0.0066), typical effect 3 years 4 months

[The typical effect for the non-significant main term CGROUP (conducted outside North America and Europe) is 5 months].

This is based on 581 out of 582 observations. The mean and range of response variable are approximately 4 years 11 months (–4 years 7 months to 27 years 9 months). The correlation coefficients used in the working correlation matrix were calculated as 0.4514 for those distance 1 from the leading diagonal and 0.1436 for those distance 2.

11.5 For all reportings of results for each randomization, which trial characteristics affect the time from close to submission?

Time from close to submission is longer for randomizations with the following characteristics:

- Greater number of questions (NOQ) (p=0.0035), typical effect 7 months
- Not presented at a major meeting (PRESENTD) (p=0.0016), typical effect 1 year 8 months
- From a publication with no or a low impact factor (IMPACT) (p=0.0041), typical effect 1 year 5 months
- Results have smaller p-values associated with them (LOGPEST) (p<0.0001), typical effect 9 months
- Direction of results is not reported (POSNG) (p=0.0449), typical effect 1 year 9 months

This is based on all 129 observations. The mean and range of response variable are approximately 5 years 1 month (–6 months to 16 years 1 month). The correlation coefficients used in the working

correlation matrix were calculated as 0.5657 for those distance 1 from the leading diagonal and 0.0650 for those distance 2.

11.6 For all reportings of results for each randomization, which trial characteristics affect the time from receipt to publication?

Time from receipt to publication is longer for randomizations with the following characteristics:

Main effects:

- Limited international or single-country as opposed to truly international participation (INTERNL) (p<0.0001), typical effect 1 year 2 months
- Multi-centre participation (MULTIC) (p<0.0001), typical effect 4 months
- Published in a non-European journal (JGROUP) (p<0.0001), typical effect 5 months
- Published in a North American or European journal (JGROUP) (p<0.0001), typical effect 1 year 1 month

(Results published in non-US, non-European journals are published more quickly than those published in European journals, with those published in North American journals taking longest.)

- Longer duration of accrual period (DURRAN) (p<0.0001), typical effect 2 months
- The main questions as stated in the paper are answered in the paper (ANSWER) (p=0.0053), typical effect 2 months

Interaction:

- Truly international trials published in journals with a high impact factor also have a longer time to publication (INTERNL*IMPACT) (p=0.0005), typical effect 9 months

[The typical effect of the non-significant main term **lower** impact factor (IMPACT) is 1 month]. For journals with a non-zero impact factor of less than 8, international trials have a shorter time to publication. For those with an impact factor of greater than 8, time to publication is increased.

This is based on 121 out of 137 observations. The mean and range of the response variable are approximately 9 months (1 month to 2 years 4 months). The correlation coefficients used in the

working correlation matrix were calculated as -0.5996 for those distance 1 from the leading diagonal and 0.0805 for those distance 2.

11.7 For all reportings of results for each randomization, which trial characteristics affect the time from close to publication?

Time from close to publication of article is longer for randomizations with the following characteristics:

Main effects:

- Earlier close date of accrual period (NCLOSE) (p=0.0001), typical effect 10 months
- Not presented at a meeting (PRESENTD) (p=0.0014), typical effect 1 year
- A clear indication is given as to whether or not the main questions in the paper are answered in that paper (ANSWER) (p=0.0119), typical effect 10 months
- Reported in articles which mention a greater number of trials (NTREP) (p=0.0071) typical effect 4 months
- Direction of result is negative (POSNG) (p=0.0399), typical effect 1 year
- Results with smaller p-values (LOGPEST) (p<0.0001), typical effect 1 year

Interaction:

- Results clearly in favour of either the experimental (p=0.0001) or the standard (p=0.0061) arm and the associated p-values are larger, the former having a very slightly greater effect than the latter (POSNG*LOGPEST), typical effect 1 year 1 month in each case

[The typical effect of the non-significant main term POSNG (direction of results is positive) is 8 months.]

In the case of results in favour of the standard arm the effect of the interaction is in the opposite direction to that of the main term. A negative result only leads to a delay in publication if it is significant only at the p=0.05 level or if significance is not achieved. Otherwise the time to publication is shortened.

This is based on all 372 observations. The mean and range of the response variable are approximately 5 years 0 months (-2 years 8 months to 19 years 0 months). The correlation coefficients used in the working correlation matrix were calculated as 0.5526 for those distance 1 from the leading diagonal and 0.1911 for those distance 2.

It is interesting to note that 'direction of result is negative' (POSNG) had a *p*-value of 0.1253 when the independence model was used, and was retained only due to the significance of the interaction between statistical significance and direction of results.

11.8 A comparison of the results obtained using standard error estimates based on the observed information with those obtained using Fisher's (expected) information

All the results quoted are based on Fisher's (expected) information. Results based on the observed information were also produced. Although the *p*-values for some variables in some analyses varied considerably between the two methods, the final choice of variables, each with a *p*-value < 0.05 remained unchanged for five out of six of the analyses. The exception is the analysis of time from close to submission for all reportings of results, where the *p*-value for direction of results reported/not reported increased from $p=0.0449$ when expected information standard error estimates were used to $p=0.1109$ when the observed information standard error estimates were used, weakening the case for retaining this variable.

12 REPORTINGS OF RANDOMIZATIONS THAT ARE PUBLISHED BEFORE CLOSURE

12.1 Introduction

Section 3.11 described how although an article may report one randomization in detail, it may also mention several other studies briefly, and that provided it is clear that from the brief description that the latter are randomized, they too are eligible for inclusion in the analyses reported in this thesis. Sometimes open trials, and occasionally future trials, are mentioned, typically in the discussion. In such cases the dates of submission and publication are before the date of closure, leading to negative times from close to submission and publication respectively. All of the analyses reported throughout this thesis use all publication records (**'the whole set'**) (provided missing values do nor preclude this) including those where the 'time to' response variable is negative (**'negative responses'**), since a brief mention of an open or future randomization is sufficient for its identification. However open and future trials differ from closed trials (**'positive responses'**) in that since their data are not mature, there is necessarily a delay to when they can actually be used in meta-analyses.

This chapter investigates whether this group of mentions are atypical. Eight of the twelve 'time to' analyses included negative responses. These are the analyses of 'time from close to submission' and 'time from close to publication' for first mentions, first reportings of results, all mentions and all reportings of results. These are re-run, firstly on the set of negative responses alone, and then on the set of positive responses alone, to see if any of significant variables obtained with the whole set are no longer significant. In each case the set of variables found to be significant in the analysis using all available records was taken and a process of backwards elimination used in order to obtain a revised model. The remaining four 'time to' analyses are for the response variable 'time from receipt to publication', for which the response is always necessarily positive.

In the following descriptions of variables no longer found to be significant when only positive responses are used, the variables are listed in the order in which they were dropped from the model and in each case the *p*-value quoted is that at the point the variable was dropped.

12.2 Time from close to submission for first mentions

Of the 63 publication records used in the analysis of the whole set the time from close to submission was negative for 10 and positive for 53.

Analysis of negative responses

Two variables were found to be significant predictors of a longer time from close to submission. These are a shorter duration of randomization period (DURRAN) (p=0.0204) and if the trial comprises fewer randomizations (NRAND) (p=0.0427). These effects act in the same directions as in the analysis of the whole set of observations.

Analysis of positive responses

All variables significant in the analysis using the whole set of records remain so when the negative responses are removed, and all act in the same directions as in the analysis of the whole set.

12.3 Time from close to publication for first mentions

Of the 195 publication records used in the analysis of the whole set the time from close to publication was negative for 32 and positive for 163.

Analysis of negative responses

Three variables are significant predictors of a longer time from close to publication. These are, again, shorter duration of randomization period (DURRAN) (p=0.0098), and also if a clear indication is **not** given as to whether the main questions in the paper are answered in that paper (ANSWER) (p=0.0119) and if reported in articles which mention **fewer** trials (NTREP) (p=0.0255). In this analysis both ANSWER and NTREP act in the opposite direction to when used in the analysis of the whole set.

Analysis of positive responses

Three variables significant in the analysis of the whole set are now no longer so. These are: the duration of randomization period (DURRAN) (p=0.8791), country group of trialist; Europe and Other versus North America (CGROUP) (p=0.4018) and impact factor of journal (IMPACT) (p=0.0939). Note that whereas CGROUP was the least significant variable in the analysis of the whole set (p=0.0434) and IMPACT was moderately significant (p=0.0089), DURRAN was highly significant (p<0.0001).

12.4 Time from close to submission for first reportings of results

Of the 52 publication records used in the analysis of the whole set the time from close to submission was negative for 1 and positive for 51. Therefore no analysis was performed for negative responses.

Analysis of positive responses

Not surprisingly all variables significant in the analysis of the whole set remained so when the one negative response was removed.

12.5 Time from close to publication for first reportings of results

Of the 170 publication records used in the analysis of the whole set the time from close to publication was negative for 16 and positive for 154.

Analysis of negative responses

Only one variable was found to be a significant predictor of a longer time from close to publication; a shorter duration of randomization period (DURRAN) (p=0.0412).

Analysis of positive responses

The only variable significant in the analysis of the whole set and now not so is duration of randomization period (DURRAN) (p=0.1238).

12.6 Time from close to submission for all mentions

Of the 209 publication records used in the analysis of the whole set the time from close to submission was negative for 14 and positive for 195.

Analysis of negative responses

None of the variables significant in the analysis of the whole set remain so.

Analysis of positive responses

All variables remain in the model with the exception of duration of randomization period (DURRAN) (p=0.1197).

12.7 Time from close to publication for all mentions

Of the 581 publication records used in the analysis of the whole set the time from close to publication was negative for 46 and positive for 535.

Analysis of negative responses

Four variables remain significant predictors of a longer time to publication. Three of these are only of borderline or moderate significance and act in the opposite direction. These are that clinical significance **is** reported (CLNSG) (p=0.0440), a clear indication is **not** given as to whether the main questions in the paper are answered in that paper (ANSWER) (p=0.0418) and if reported in articles which mention **fewer** trials (NTREP) (p=0.0239). The fourth variable, shorter duration of randomization period (DURRAN) is highly significant (p=0.0002) and acts in the same direction as it does in the analysis of the whole set.

Analysis of positive responses

Five variables significant in the analysis of the whole set are no longer so. These are; duration of randomization period (DURRAN) (p=0.5137), number of randomizations mentioned in article (NRREP) (p=0.2168), impact factor of journal (IMPACT) (p=0.1386), published in full in a language other than English versus English (ENGLISH) (p=0.0698) and reported as a meeting

abstract versus as a journal article or book chapter (PUBTYPE) (p=0.0802). NRREP, IMPACT, ENGLISH and PUBTYPE were all of only moderate significance in the analysis of the whole set (p= 0.0362, 0.0128, 0.0349 and 0.0316 respectively). However, DURRAN was highly significant (p=0.0008). Of the variables remaining in the model, all act in the same direction as in the whole set analysis.

12.8 Time from close to submission for all reportings of results

Of the 129 publication records used in the analysis of the whole set the time from close to submission was negative for one and positive for 128. Therefore no analysis of negative responses was undertaken.

Analysis of positive responses

Not surprisingly all variables significant in the analysis of the whole set remained so when the one negative response was removed.

12.9 Time from close to publication for all reportings of results

Of the 129 publication records used in the analysis of the whole set the time from close to publication was negative for 20 and positive for 352.

Analysis of negative responses

Two variables remain significant predictors of a longer time to publication. These are; that an indication is given as to whether the main questions as stated in the article are answered in the article (ANSWER) (p<0.0001) and degree of statistical significance (LOGPEST) (p=0.0412), the latter acting in the opposite direction to its effect in the analysis of the whole set.

Analysis of positive responses

Only one variable is no longer significant, whether an indication is given as to whether the main questions as stated in the article are answered in the article (ANSWER) (p=0.0710).

12.10 Conclusions

The group of randomizations which are mentioned before closure are clearly atypical. Their removal generally leads to one or more variables becoming no longer significant. Two analyses of the whole set contained only one negative response, those of time from close to submission for the first reporting of results and for all results. In both cases its removal made minimal difference to the model.

Of the other six whole set analyses, five contained variable DURRAN. This variable appears to be acting as a surrogate for whether or not the response is negative or positive, most notably in the analyses of time from close to publication for first mentions and for all mentions. The set of all mentions was taken and a variable created to indicate those publication records with date of publication prior to that of closure. The correlation of this indicator with DURRAN was found to be moderately high, $\rho=0.32775$ ($p<0.0001$ using the 582/610 observations for which start and close date of accrual period and date of publication are known), supporting this theory.

13 WHICH FACTORS AFFECT THE THREE TIME PERIODS? – CONCLUSIONS

13.1 Introduction

This chapter begins by summarizing the findings of the twelve 'How long?' analyses, in tabular form. This facilitates a comparison of factors affecting a particular time period response, between the 'all records', 'all results', 'first mentions' and 'first results' analyses, or for a particular dataset a comparison of the factors affecting the three time periods analysed. This is followed by a written description of the key findings, with some suggestions as to why these may be important factors and implications, i.e. the possibility of introducing bias when identifying randomized comparisons for inclusion in meta-analyses. Finally, the findings of the three 'first results' analyses are compared with results of other similar studies identified from a literature search.

13.2 Table summarizing the findings from the twelve 'How long?' analyses

Table 13.1 summarises the significant variables in each of the twelve regressions.

Unchanging trial characteristics taken from the definitive record are given first, followed by results variables, followed by variables to do with the article, followed by interactions.

Key

ANSWER	main questions answered (Yes, No, X =not reported)
ARMS	number of arms
CGROUP	country group of trialists (America, Europe, Other)
CLNSG	clinical significance (Yes, No, Possibly, X=not reported)
DEVLPNG	conducted in a 'developing' country
DURRAN	duration of randomization
ENGLISH	published as an Abstract in English rather than in full
EQUIV	equivalence trial
FIRSTL	relapse/refractory as opposed to first-line treatment
FUNDG	funding source (Government, Charity)
IMPACT	impact factor
INTERNL	international participation (Yes, Limited, No)
JGROUP	country group of publisher (America, Europe, Other)
LOGAUTH	log_{10} (number of co-authors)
LOGPEST	spacing between log_e (typical *p*-value) and that of the non-significant/not reported category
LOGNTREP	log_{10} (number of trials reported)
MULTIC	multi-centre participation (Yes, Limited, No)
NCLOSE	close date of accrual period
NOIMPACT	no impact factor associated
NOQ	number of questions
NRAND	number of randomizations
NRREP	number of randomizations reported
NSTART	number of trials reported
POSNG	direction of results (+, - , Null, Opposite, X=not reported)
PRESENTD	presented at a major meeting
PUBTYPE	Journal, Book or Meeting abstract

n number of observations used in the final model out of total number for which the response variable is present

p number of variables in the final model including all indicator variables used to express a categorical variable and interactions

↑ explanatory variable for which an **increase**, in the case of a continuous variable, or belonging to that class, in the case of a categorical variable, corresponds to an **increase** in the response variable

↓ explanatory variable for which a **decrease**, the case of a continuous variable, or **not** belonging to that class, in the case of a categorical variable, corresponds to an **increase** in the response variable

↑↑, ↓↓ class having the greatest effect for a categorical variable with three categories

* interaction between two main effects

Dataset **Response variable**	**All mentions**	**All results**	**1st mentions**	**1st results**
Close to submission ↑ (TCLREC)	*n*=209/209 *p*=6 DURRAN ↓ DEVLPNG ↑ LOGPEST ↑ IMPACT ↓ LOGNTREP ↑ NRREP ↓	*n*=129/129 *p*=5 NOQ ↑ LOGPEST ↑ POSNG (X↑ vs. +,-,F,O) PRESENTD ↓ IMPACT ↓	*n*=63/63 *p*=4 NRAND ↓ DURRAN ↓ CGROUP (E,O ↑ vs. A) DEVLPNG ↑	*n*=52/52 *p*=7 NCLOSE ↓ MULTIC (Y,L ↑ vs. N) CGROUP (O ↑↑ vs. E ↑ vs A) POSNG (-↓↓vs+,F↓ vs. O,X) IMPACT ↑
Receipt to publication ↑ (TRECPUB)	*n*=195/218 *p*=6 NSTART ↑ DURRAN ↑ CGROUP(O,A↑ vs. E) JGROUP (A,E vs. O ↓) NRREP ↓ LOGAUTH ↑	*n*=121/137 *p*=8 DURRAN ↑ MULTIC (Y,L ↑ vs. N) INTERNL (Y↓ vs. L,N) ANSWER (Y↑ vs. N,X) JGROUP (A vs. E ↓vs. O↓↓) INTERNL Y* IMPACT↑	*n*=67/72 *p*=7 NRAND ↓ CLNSG (Y,N ↓ vs. P,X) JGROUP (O↓ vs. A,E) NRREP ↓ CGROUP O,E * DURRAN ↑	*n*=49/60 *p*=6 FUNDG (G vs. G+C↑) CLNSG (Y↓ vs. N↓↓vs. P,X) PRESENTD ↑ NTREP ↑ NRREP ↓
Close to publication ↑ (TCLPUB)	*n*=581/582 *p*=16 NSTART ↓ DURRAN ↓ MULTIC (Y ↑ vs. L,N) CGROUP (O,Avs.E ↓) DEVLPNG ↑ CLNSG (Y↑↑ vs.X↑ vs. N,P) ANSWER (Y,N ↑ vs. X) PRESENTD ↓ IMPACT ↓ PUBTYPE (M↓ vs. J,B) ENGLISH (A↓ vs. E) NTREP ↑ NRREP ↓ CGROUP E,O * MULTIC N,L↑	*n*=372/372 *p*=9 NCLOSE ↓ LOGPEST ↑ POSNG (-↑ vs. +,F,O,X) ANSWER (Y,N ↑ vs. X) PRESENTD ↓ NTREP ↑ LOGPEST * POSNG + ↓ LOGPEST * POSNG - ↓	*n*=195/195 *p*=9 NSTART ↓ DURRAN ↓ CGROUP (O ↑vs. A,E) DEVLPNG ↑ EQUIV ↓ ANSWER (Y,N ↑ vs. X) PRESENTD ↓ IMPACT ↓ NTREP ↑	*n*=170/170 *p*=10 DURRAN ↓ LOGPEST ↓ CLNSG (Y,N ↓ vs. X,P) POSNG (F ↓vs. +,-,O,X) IMPACT↓ NOIMPACT ↓ CGROUP E*IMPACT ↑ CGROUP O*POSNG F↑

Table 13.1 Summary of the findings of the twelve 'How long?' analyses

13.3 Implications for the identification of randomized comparisons for inclusion in meta-analyses

The output from the twelve regressions, using the final model in each case, is given in Appendices VIII ('first mentions' and 'first results' analyses) and XV and XVIII (analyses incorporating repeated measures), and the findings were summarized in Chapters 6 and 11, quoting the R^2 value for each independence model, and the *p*-value for each variable in all models, independence and those incorporating repeated measures. In this section the key findings are discussed, with some suggestions as to why these may be important factors. In Table 13.1 the significant variables are given in an order such that those to do with the randomization are given first, followed by results variables, and then those to do with publication. Terms included in interactions are mentioned last.

It can be seen that there are approximately two and a half to three times as much data for the 'time from close to publication' analyses than for those investigating the other two time periods. Date of submission/receipt of article is only available for a subset of particular journals and for no books or meeting abstracts.

Since both the number of observations (*n*) and the number of variables (*p*) used vary between the twelve analyses it is interesting to look at the ratio of *n* to *p*. In four analyses the number of observations is between 30 and 41 times that of the number of variables. These are all three of the 'all mentions' analyses plus the analysis of time from close to publication for all reportings of results. For five analyses this ratio is greater than 15 but less than 30. These are the analyses of time from close to submission and time from receipt to publication for all reportings of results, time from close to submission and time from close to publication for first mentions, and time from close to publication for first reportings of results. However, for the analyses of time from receipt to publication for both first mentions and first reportings of results the ratio is approximately 10, and for that of close to publication for first reportings of results the ratio is just over 7.

For the final models for all four analyses of time from close of randomization to submission and for three of the four analyses of time from close to publication, relevant observations were excluded

solely due to missing values for the response variable. In the analysis of time from close to publication for all mentions, only one observation was excluded due to a missing value for one of the chosen explanatory variables. However, for each of the four 'time from receipt to publication' analyses a considerable number of records, for which the response variable was present, had to be omitted due to missing values for one or more of the chosen explanatory variables. The numbers and proportions of observations omitted for this reason are as follows: 'all mentions' 23/218=11%, 'all results' 16/137=12%, 'first mentions' 5/72=7% and 'first results' 11/60=18%. This could indicate that the models developed for the time from receipt to publication analyses are less stable than those developed for analyses of the other two time periods.

Also, the date of submission/receipt used is that for the journal in which the article appears. In some cases an article was submitted to, and rejected by, other journals before finally being accepted. Since it is likely that the more prestigious journal is approached first, and that this is likely to have a high impact factor, it is to be expected that if impact factor is significant, then the time from close to submission (and to publication) will be longer for journals with no or a low impact factor.

It is not easy to predict the direction in which variables may affect the outcome. For example trialists may be keen to publish a randomization with a significant result. However, a significant result may not be achieved for several years (if ever), when the event rate is small. Similarly, a large multi-centre or international trial may be considered worthy of swift publication. However this may mean that the agreement of many co-authors or working party members must be obtained before a manuscript is submitted, which could slow down the process.

13.4 Conclusions for the first reporting of results of each randomization

Records are counted as reporting results if any indication of the type of result (e.g. survival), statistical significance, clinical significance or direction of results, is given, however brief.

Time from close to submission for the first publication of results can be considered an indication of how keen the trialists are to publish the results of their work. Factors associated with a longer time from close to submission are:

- Randomizations are from early trials
- Involving multi-centre participation
- Conducted in countries other than the US, with those conducted outside both North America and Europe taking still longer.
- Direction of results is opposite (both a positive and a negative result are reported) or not reported. Negative results are submitted fastest, followed by positive and null results.
- Published in journals with a high impact factor

A surprising finding is that results in favour of the standard or control arm, were **submitted** more quickly than all others, including those clearly in favour of the experimental arm. However this was not found to be significant in the analysis of time from close to **publication** for first reportings of results, and indeed when extended to the analysis of time from close to publication for all reportings of results, negative results were found to have a **longer** time to publication than all others. This indicates that although trialists may be keen to publish their negative findings, journal editors may be less so.

The fact that time from close to submission was longer for randomizations with multi-centre participation is to be expected, since the agreement of several people on details of the article to be submitted is likely to be necessary. The finding that time to submission was longer for articles published in journals with a high impact factor was surprising. Initially the reason for this was thought to be that results are often first reported as meeting abstracts, many of which do not have an impact factor, and that for analysis purposes where the impact factor is missing it has been set to zero. However this cannot be so since the date of submission/receipt is not available for any meeting abstracts.

Time from receipt of article to publication is an indication of how eager the journal is to publish the results of a randomization. Factors associated with a delay in time from receipt of article by the publisher to actual publication of first results are:

- Funded by charity as well as Government money
- No clear indication is given as to whether results are clinically significant or not. (Those clearly not clinically significant are published quickest, followed by those that clearly are clinically significant.)
- Presented at a major meeting
- Reported in articles that mention a greater number of trials
- Reported in articles that mention fewer randomizations

The time from close to publication, the sum of the above two time periods, for the first publication of results, is very important, since this is the delay between obtaining the results of a randomized comparison and making them widely available, through publication, to the clinicians who treat patients, and who may change their practices as a result. Factors associated with longer delay from close to publication are:

Main effects:

- Shorter duration of accrual period
- Results have larger *p*-values associated with them
- No clear indication is given as to whether the results are clinically significant or not
- Reported in a journal with an impact factor associated with it

Interactions:

- Conducted in a European country and published in a journal with a higher impact factor
- Conducted outside Europe and published in a journal with a lower impact factor
- Conducted outside North America and Europe and results are null
- Conducted in North America or Europe and results are not null

The findings to do with impact factor are interesting. As expected, the time from close to publication was longer when results were published in a journal with an impact factor associated with it. This is because the group of journals without an impact factor includes many meeting abstracts and all book chapters, and it is common for results of a randomization to be reported as a meeting abstract prior to publication as a full paper in a journal. For non-European trials, the lower the impact factor the longer the time to publication. This is also as expected since it is common for an article to be submitted to a prestigious journal, with a high impact factor, rejected and then re-submitted to a less prestigious journal, with a lower impact factor. However, surprisingly, it was found that for European trials, a longer time from close of randomization was associated with publication in a high impact factor journal. One possible explanation for this is that many of the higher impact factor journals are American (as described in Section 4.4), and there may be a tendency towards accepting articles which report results of American trials.

The findings to do with the nature of the results reported are important. The first indicates that results with smaller *p*-values associated with them are made available more quickly than those that are less striking. However this is countered by the second, which is that null results (where no clear benefit of either the experimental or the standard arm is shown) from North American and European trials are published more quickly than when the trial found in favour of one of the treatment arms.

A likely explanation for this is as follows: where results are highly statistically significantly in favour of a particular randomization arm, the trialists will be keen to publish. Where there is clearly no difference between the outcomes, there is no reason to delay publication. However, if the outcome in one trial arm is better than that in the other, and this result is almost statistically significant, the trialists will wish to delay publication in the hope that the *p*-value will go below 0.05.

An alternative explanation of these two findings is that this may be due to two differing attitudes of trialists. One group of trialists may value a null result and seek to publish it, going on to re-publish

the results after a longer follow-up period, by which time there may well be a sufficient number of events to achieve statistical significance. The other may only publish if, and when, statistical significance is reached.

It is of some concern that null results from randomizations conducted outside North America and Europe are taking a longer time to publish than others.

13.5 Conclusions for the first mention of each randomization

The first mention of a randomization often does not include any results. For example, a paper publishing detailed results of a trial may also refer briefly to a subsequent trial that is currently open or in its planning stages. However, provided it was clear that this new trial is randomized, a brief mention such as this would be sufficient for the purposes of those involved in identifying randomizations for inclusion in meta-analyses. The time periods 'close to submission' and 'receipt to publication' are less useful when looking at first mentions, as opposed to first publication of results, since briefly mentioning a new trial is not a motivation for either the trialists or the publisher to publish the report quickly.

The factors likely to increase time from close to submission for first mentions are as follows:

- Trial comprises fewer randomizations
- Shorter duration of accrual period
- Conducted outside North America
- Conducted in a 'developing' country

The factors likely to increase time from receipt to publication, for first mentions are:

Main effects:

- Trial comprises fewer randomizations
- No clear indication is given of whether results are clinically significant or not
- Published in a North American or European journal

- Reported in an article that mentions fewer randomizations

Interaction:

- For trials conducted outside North America, longer duration of accrual period

However, time from close to publication is very important, since this is the delay between the randomization closing and its availability for identification and inclusion in meta-analyses.
The factors likely to increase this delay are as follows:

- Earlier randomizations
- Shorter duration of accrual period
- Not an equivalence trial
- Conducted in a 'developing' country
- Conducted outside North America and Europe (of borderline significance, once 'developing'/ 'developed' country is taken into account)
- A clear indication is given as to whether the main questions in the paper are answered in that paper
- Not presented at a major meeting
- Reported in a journal with a low or no impact factor
- Reported in articles which mention a greater number of trials

A possible explanation as to why the time from close to publication is shorter for equivalence trials is that a longer time is needed to obtain enough events to demonstrate the superiority of one treatment arm over the other than to demonstrate no significant difference in outcome between treatment arms. This variable did not remain in the reduced set of significant predictors for the 'time from close to publication for first results' analysis, although the reason may be that there were other variables that had a greater effect.

Again, it is worth noting that the time to publication is longer for randomizations conducted in 'developing' countries and also in countries other than North America and Europe. It is important that these randomizations are not excluded from any meta-analyses for which they are eligible.

A possible reason for randomizations reported in articles which mention a greater number of trials leading to a longer time from close of randomization to publication is that the close dates of the randomizations are likely to span a wider time period and so include some earlier dates.

Note that since a brief mention of a new trial may occur when the trial is still open, or even before it is open for accrual, it is reasonable to expect some negative 'close to publication' time periods.

It is encouraging to note that the results of more recent randomizations are being reported more quickly than was the case with earlier trials, and also that they are first mentioned more quickly and so can be identified sooner after closure.

13.6 Conclusions for all reportings of results of each randomization

Of the three time periods, the most straightforward to interpret, when looking at all reportings of results, is 'time from receipt to publication'. The results of a randomization may be reported several times, for example shortly after closure, and after follow-up periods of various lengths. When looking at 'time from close to publication' (or 'from close to submission') a high value of the response variable may indicate that the results of a randomization were slow to be published because they were not considered important, or alternatively, that they were considered so important that they were re-analysed and updated many years after the randomization closed. A third possibility is that a randomization may not yield statistically significant results to begin with, but later when there have been more events, a level of statistical significance is achieved. Therefore it is not always possible to know whether the time period response variable is affected by the explanatory variables or whether it is the other way around. All that is known is that there is an association between the two.

Factors associated with a longer time from close to submission for all reportings of results are:

- Greater number of questions
- Results with smaller *p*-values associated with them

- Direction of results is not reported
- Not presented at a major meeting
- From a publication with a low or no impact factor

It is interesting to note that the two significant trial characteristics common to the 'first results' and 'all results' analyses have opposite effects. For the first results, the higher the impact factor of the journal the longer the time from close to submission. However, when including all reportings of results, a lower impact is associated with a longer time from close to submission. Since only journal articles are used in this analysis, due to date of submission/receipt of article being unknown for all meeting abstracts (and book chapters), the reason for this finding is unclear.

The second significant predictor of time from close to submission common to the two analyses is the direction of results. In the 'first results' analysis, negative results were submitted most quickly, followed by positive and null results, with opposite results (i.e. where one main result is in a positive direction and the other in a negative direction) and those where the direction of results is not reported, having a longer time to submission. When all reports of results are included, again those results where direction is not stated have a longer time to submission, but there is no significant difference in the time to submission for results where direction is stated.

Factors associated with a longer time from receipt to publication for all records containing results are:

Main effects:

- Longer duration of accrual period
- Multi-centre participation
- The main questions as stated in the paper are answered in the paper
- Published in a North American journal, followed by European journal, with journals from elsewhere being the quickest to publish.
- Limited international or single-country as opposed to truly international participation

Interaction:

- However, truly international trials published in a journals with a high impact factor also have a longer time to publication. For journals with a non-zero impact factor of less than 8, international trials have a shorter time to publication than others. For those with an impact factor greater than 8, the time to publication is increased. A possible explanation for this is that American journals may be more keen to publish American trials than those conducted elsewhere, and many of the higher impact journals are American (as was described in Section 4.4.). Also international trials are more likely to be conducted outside the US. Therefore a journal with a higher impact factor is likely to publish an international trial less quickly.

None of the trial characteristics found to be significant predictors of time from receipt to publication for first results were also significant when all reportings of results were included.

Factors associated with a longer time from close to publication, for all records reporting results are:
Main effects:

- Earlier randomization
- An indication is given as to whether the main questions in the paper are answered in that paper
- Not reported at a major scientific meeting
- Reported in articles which mention a greater number of trials
- Direction of results is negative
- Results with smaller *p*-values

Interaction:

- Results clearly in favour of either the experimental or the standard arm, and the associated *p*-values are larger, the former having a slightly greater effect than the latter.

In the case of results in favour of the standard arm the effect of the interaction is in the opposite direction of that of the main term. A negative result only leads to a delay in publication if it is

significant only at the p=0.05 level or if significance is not achieved. Otherwise the time to publication is shortened.

The significant predictors of time from close to publication for first results that were also significant when all results were included are statistical significance of results and direction of results. For the set of first reportings of results, as has already been discussed, results with larger p-values attached to them had a longer time from close to publication, as did non-null results, as might be expected. When all results are included in the analysis, the effect of statistical significance is similar, in that for clearly positive or negative results those with larger p-values associated with them have a longer time to publication.

13.7 Conclusions for all mentions of each randomization

Variables found to be associated with a longer time from close of randomization to submission are:

- Shorter duration of accrual period
- Conducted in a 'developing' country
- Results with smaller p-values associated with them
- Reported in a publication with a low or no impact factor
- Published in an article which mentions a **greater** number of trials
- Published in an article which mentions **fewer** randomizations

Two of the significant predictors of a longer time from close to submission for first mentions remained significant when all mentions were included. These are a shorter duration of accrual period and if the trial was conducted in a 'developing' country. These two factors are also significant predictors of time from close of randomization to actual publication for both first and all mentions. The importance of the latter has already been discussed. The reason for the former is unclear.

Factors associated with a longer time from submission/receipt to publication are:

- More recent trials

- Longer duration of accrual period
- Conducted outside Europe
- Published in a North American or European journal
- Reported in articles which mention fewer randomizations
- Larger number of co-authors

Three of the significant predictors of a longer time from receipt to publication in the 'first mentions' analysis are also significant in the 'all mentions' analysis. These are being published in an American or European journal, being reported in an article which mentions a fewer number of randomizations and, for trials conducted outside North America, a longer duration of accrual period.

Factors associated with a longer time from close of randomization to publication of article are:
Main effects:

- Earlier trials
- Shorter duration of accrual period
- Conducted in a 'developing' country
- Results clearly reported as clinically significant, followed by those where clinical significance is not reported. Results clearly not or only possibly clinically significant were published quickest.
- A clear indication is given as to whether the main questions in the paper are answered in that paper
- Not presented at a major meeting
- Published in full in the English language.
- Reported in articles which mention a **greater** number of trials
- Reported in articles which mention a **smaller** number of randomizations
- Reported in a publication with a low or no impact factor
- Reported in a journal article or book chapter as opposed to a meeting abstract

- Conducted outside Europe
- Participation of five or more centres

Interaction:

- However, for trials conducted outside North America, the participation of less than five centres also acts to delay publication

This is the only analysis where publication type, i.e. whether reported in a journal article, book chapter or meeting abstract was found to be significant. Since time to publication is shorter for reportings in meeting abstracts this supports the case that time-consuming hand-searching of meeting abstract books is a worthwhile process for trial identification. However, publication type was not significant in either the 'first mentions' or 'first results' analyses of time from close to publication.

All but one of the nine significant predictors of a longer time from close to publication for first mentions are also significant for the analysis using all mentions.

Seven of these affect the response in the same direction for the two analyses. These are: an earlier start date of the accrual period, a shorter duration of accrual period, trial conducted in a 'developing' country, if the article clearly states whether or not the main questions of the randomization are answered in the article, if the findings have not been already presented at a major scientific meeting, if the article is published in a journal with a lower impact factor, and if the article mentions a larger number of trials.

In addition, the country group of trialists is also significant in the two analyses. For the first mentions, randomizations conducted by trialists in North America or Europe had a longer time to publication than those run in other countries. In the 'all mentions' analysis randomizations from North American and other non-European trials were found to have a longer time to publication than those conducted in Europe. In addition, for trials conducted outside North America, the participation of less than five centres also leads to a delay in publication.

13.8 Studies of 'pipeline' bias identified from a literature search

As was described in Section 2.5, several other studies of 'pipeline bias' were identified. [Liebeskind et al (1999), Misakian and Bero (1998), Cheng et al (1998), Stern and Simes (1997), Handysides (1996), Dickersin and Min (1993), Chew (1991) and Ioannidis (1998)] These, plus this current study, are summarised as a table in Appendix XIX.

XVII

13.9 Differences between this project and the other studies being compared

Before comparing the findings of this project with those of the other studies, it is necessary to note three important differences:

(i) the definition of statistical significance

In this thesis two variables are used to specify statistical significance. LOGPEST is a continuous variable, the magnitude of the distance of log_e (a typical value of the *p*-value for the category to which the result belongs) from that for the 'non-significant/not reported' category. The advantage of the implementation of a continuous variable to represent statistical significance is that the degree of statistical significance of the results, as opposed to merely 'statistically significant versus non-significant' is built in. This was described in detail in Section 4.2.1. LOGPNR is used to specify whether or not any indication of the *p*-values associated with results is given in the article. Together the two variables specify the degree of statistical significance of the results, and also whether a value of zero for LOGPEST means 'non-significant' or 'not reported'. A third variable, POSNG, specifies the direction of the results, i.e. positive, negative, null (flat), opposite (one positive, the other negative) and not reported.

In all the analyses performed, where LOGPEST and POSNG were found to be either significant or non-significant, but only eliminated at $p<0.2$, the interaction between LOGPEST and the categories of POSNG were tried in the model, as well as the main effects.

In three of the other studies, Liebeskind et al (1999), Misakian and Bero (1998) and Ioannidis (1998), one variable is used in a two-dimensional capacity to specify both statistical significance and direction. In Cheng et al (1998) and Dickersin and Min (1993) different definitions are used again.

In this thesis statistical significance and direction of results are two distinct one-dimensional variables. The only 'two dimensional' results variable used is 'clinical significance'. This combines the direction of the result with the 'strength of the impression given by the text', in the opinion of the authors, regardless of any *p*-values reported. If the general tone of the paper indicated that the results were in favour of the experimental arm, or in the case of an equivalence trial, the experimental arm was at least as good as the standard arm, then the findings are classed as 'clinically significant'. If the impression given was that this was clearly not so, then the findings are classed as 'not clinically significant'. There are also 'possibly clinically significant' and 'not reported' categories. The definition of 'significance' as *"judged by investigator to be either statistically significant or 'of great importance'* " used by Dickersin and Min (1993) is similar to that of 'clinical significance' used here, in as much as it a measure of how important the trialists view their findings.

(ii) the response variable in question

This project analysed four datasets: 'all mentions', 'all reportings of results', 'first mentions' and 'first reportings of results'. It is the 'first reporting of results' that is the focus of the other studies found, and therefore compared here. In this project the main response variable analysed is the time from the close of randomization (i.e. of the accrual period) to first publication. The other two response variables are 'time from close of randomization to submission', and 'time from receipt by accepting journal to publication'.

In the other studies, time to publication is measured from varying starting points. Liebeskind et al (1999) and Ioannides (1998) used the start of the accrual period to publication, Misakian and Bero (1998) used the year funding began to year of publication, Stern and Simes (1997) used time from approval by ethics committee to publication.

(iii) the number of explanatory variables, in addition to statistical significance and direction of results, tried in the analysis

It is impossible to know all of the variables tried in the other studies. Variables may well have been tried in the model, found to be non-significant and not mentioned in the report. However, another possibility is that a larger number of variables were tried in this project, some of which were found to be more significant than those expressing statistical significance and direction of results, in which case statistical significance and direction were dropped from the model.

13.10 A comparison of the findings of this project with those of other studies

In the light of the above points, the findings of this thesis can be compared with those of other similar studies identified. The terminology used throughout this thesis will be used when referring to all studies, for consistency.

From the published studies there is a common theme: time to publication was shorter for statistically significant positive results than for statistically significant negative results, with non-significant results having the longest time to publication.

Lieberskind et al (1999) found that time to publication was shorter for statistically significant positive results than for statistically significant negative and null results. However this difference was not statistically significant (p=0.079). Misakian and Bero (1998) found that the time to publication was statistically significantly shorter for statistically significant results than for non-significant results (p=0.007). In this study the direction of results was not a factor. Stern and Simes (1997), similarly, reported that statistically significant positive results were published more quickly

than clearly non-significant results, with those of borderline significance having the longest time to publication. Ioannides (1998) found that time to publication was shorter for statistically significant positive results than for statistically significant negative results ($p<0.001$).

The findings of this thesis were similar to the above in that the more statistically significant the results, the shorter time to publication ($p=0.0153$). However, it was also found that null results were published more quickly than others ($p=0.0127$). As was discussed in Section 13.4, a possible explanation for these findings is as follows: where results are highly statistically significant in favour of a particular treatment arm, the trialists will be keen to publish as soon as possible. Where there is clearly no difference in the outcomes of the two arms, there is no reason to delay publication. However, if the outcome in one trial arm is better than that of the other, and this result is almost statistically significant, the trialist will wish to delay publication in the hope that statistical significance will be achieved. Ioannidis (1998) points out that '*long-protracted trials often had low event rates and failed to reach statistical significance, while trials that were terminated early had significant results'*. This thesis found that although generally null results were published faster than others, null results from trials conducted outside North America and Europe had a far longer time to publication ($p=0.0023$). Unlike Ioannidis' finding, this thesis found that a shorter duration of accrual is associated with a delayed first publication of results ($p=0.0002$). However this could be complicated by the fact that some randomizations are 'serial' as opposed to 'separate'. (This was discussed in Section 7.5.)

Prompted by the above comment by Ioannides, the correlation between duration of accrual period (DURRAN) and degree of statistical significance (LOGPEST), and that between 'decline in interest in the randomization', measured by 'time to accrue the second half of the patients minus time to accrue the first half' (WANE) and LOGPEST were obtained for first reportings of results. Both have low values and neither is statistically significant. In concordance with Ioannidis' finding, the correlation between DURRAN and LOGPEST is negative ($\rho= -0.11440$, $p=0.1374$ based on the 170/188 first reportings of results for which duration of accrual period is known, degree of

statistical significance being known for all 188.) i.e. a shorter accrual period is associated with a higher degree of statistical significance. Surprisingly, the correlation between WANE and LOGPEST is positive ρ=0.20378, i.e. a decline in interest is associated with a higher degree of statistical significance, although this is of only borderline statistical significance (p=0.0569) and based only on the 88/188 first reportings of results for which the former is known.

This thesis also found that if results were either clearly 'clinically significant' or 'not clinically significant' then the time to publication is shorter than if no clear impression is given.

As in this thesis, a number of the studies, Liebeskind et al (1999), Misakian and Bero (1998), Cheng et al (1998), Stern and Simes (1997) and Ioannides (1998), collected data on sample size. This is not mentioned as a significant predictor of time to publication in any of them, with the exception of Misakian and Bero (1998) and then only when univariate analysis was used and animal studies included (p=0.03). Similarly, in this project, number of patients accrued was not a significant predictor for any of the three time periods modelled.

Ioannides (1998) reports that randomizations with statistically significant positive results were submitted for publication significantly more rapidly after completion than were those with statistically negative results (p=0.001). This project found the opposite i.e. that randomizations with negative results were submitted for publication faster than those with either positive or null results, with opposite (i.e. one main result positive, the other negative) and those where the direction of the results is unspecified having the longest time to submission (p=0.0147 and p=0.0107 respectively). Of course, it could be argued that those results where direction is not specified are more likely to be negative, and that this could weaken the finding. It was also found that statistical significance was not a statistically significant predictor of time to submission. However, if some of the other significant variables had been omitted from the analysis, this may not have been the case.

Ioannides (1998) also found that the time from submission/receipt to publication was shorter for statistically significant positive trials than for statistically significant negative trials (p=0.04). This project did not find either statistical significance or direction of result to be statistically significant predictors of time from receipt to publication. However, it did find that results clearly reported as not 'clinically significant' are published faster than those that are, with those for which no impression of clinical significance is conveyed taking longest (p=0.0056 and p=0.0040 respectively), in contrast to Ioannides' findings.

Chew (1991), in the study of the eventual publication of papers rejected by one particular journal, the *American Journal of Roentgenology (AJR)*, found that the time from rejection by the *AJR* to publication elsewhere was longer for papers published in non-US journals, and that most of the journals of eventual publication had lower impact factors than the *AJR*. Although the results of this project cannot be compared directly, it is interesting to note that those articles eventually published in a journal with a high impact factor were found to have a longer time to submission (p=0.0242). For European trials, the time from close to publication was also longer (p=0.0062), but generally eventual publication in a high impact factor journal led to a shorter time to publication. The latter agrees with Chew's findings. Not surprisingly time from close to publication was shorter for articles in publications without an impact factor associated with them, which include some meeting abstract books and all book chapters as well as some journals.

13.11 Conclusions

In agreement with the other studies identified, it was found that a higher degree of statistical significance leads to a shorter time to publication. However, unlike the findings of most of the other studies, it was also found that the direction of results was not a significant predictor of time to publication, with the exception of null results which are generally published more quickly than others, except for null results from trials conducted outside North America and Europe which are published more slowly. In addition, an important factor was found to be the importance the trialists attach to their results, irrespective of the direction of those results. This thesis also found, in

contrast to the published studies, that randomizations which found in favour of the standard or control arm were submitted faster than all others, including those clearly in favour of the experimental arm, although they were not actually published faster.

14 OTHER INVESTIGATIONS I: WHICH TRIAL CHARACTERISTICS AFFECT (a) WHETHER A RANDOMIZATION IS EVER MENTIONED IN AN ARTICLE AND (b) WHETHER THE RESULTS OF THAT RANDOMIZATION ARE EVER REPORTED?

14.1 Introduction

As was discussed in 2.4 the two main questions addressed in this thesis are:

- Which factors affect *how long* the searching process should continue in order to identify all trials through publication?
- *How wide* does the search need to be?

The first of these has been thoroughly investigated and reported in this thesis. Important aspects of the second are discussed in this chapter and Chapter 16 but will be the subject of future research.

Issues addressed are:

- why some randomizations are never mentioned in any article
- why the results of some randomizations are never reported
- why some randomizations are mentioned in more articles than others
- why the results of some randomizations are published more frequently than those of others

A large proportion of the time spent on this project went into the collection and checking of the data. These same data, with some manipulation, could be used to answer other questions. Other aspects, not investigated here, could include, for example:

- which characteristics of randomizations result in publication in the most prestigious journals
- why some randomizations are published as abstracts, but never reach publication as full journal articles

14.2 Data used for the analyses

The 'definitive records' dataset was used for the analyses in this chapter. This means that only variables which remain unchanged over time have been included. Data specific to an individual publication, such as results at a point in time, have not been used. These analyses do not involve repeated measures since the 'definitive records' dataset comprises 243 records, one for each randomization, using an amalgamation of the data from all available sources.

The LOGISTIC procedure in SAS was used to perform logistic regression. The algorithm for the program and some SAS commands for this are given in Appendix XX.

Variables not included in the analyses

As with the 'How long?' analyses, variables with missing values for more than 25% of observations are not included in the analyses. The justification for this is that a variable available for a small proportion of the observations is unlikely to affect the response variable to any degree. Variables omitted for this reason are:

- Funding source (FUNDG) (34% missing)
- Method of randomization (in order of reliability: by central computer, by notification to a central office, or by a sealed envelope method) (RANDMETH) (81% missing)
- Timing of late randomization (done at the correct stage or done too early) (RANDTIME) (72% missing)
- Randomization design used (simple randomization, block randomization or minimization of imbalance) (RDESIGN) (95% missing)
- Whether or not any attempt was made to balance patient characteristics between randomization arms (BALANCED) (61% missing)
- Risk group eligibility (RISK) (41% missing)
- Whether the target number of patients was reached (TARGET) (85% missing)
- Whether the target number of centres was reached (CTARGET) (27% missing)

- Whether it took longer to accrue the second half of the patients than the first (WANE) (57% missing)

One variable was collected but not considered reliable enough for inclusion in the analyses:

- Whether the two main questions of the trial were ever answered in any publication (ANSEVER)

The reason for the unreliability of the variable is that the main questions (e.g. survival, disease-free survival, relapse in a particular site) were taken from the trial protocol where available and failing that from the articles themselves. However, when the main questions for the group of randomizations, for which there was information from both the trial protocol and the articles were investigated, it became clear that the main questions from the two sources often were discrepant. Therefore, if this variable had been included in the analyses, it may have introduced bias between the group of randomizations for which the trial protocol (see Appendix II) is available, and the group for which it is not.

Another variable not used is:

- Trial category (e.g. induction treatments, central nervous system prophylaxis, maintenance duration, bone marrow transplant) (TRCAT)

It was decided that, with eleven categories, it was impractical to use this variable. Instead, various combinations of the categories were used to form two new variables. These are:

- Treatment type (chemotherapy, radiotherapy, immunotherapy, transplant and antibiotic) (TXCHEMO)
- First-line treatment or relapse/refractory (FIRSTL)

Variables included in the analyses

A backwards elimination process is used for the initial stage of each analysis. As with the 'How long?' analyses, all variables having no, or very few missing values are put into the model. The regression is run and the least significant variable is then dropped. This process is then repeated

until a reduced set of significant variables (each with $p<0.05$) remains. Other variables can then be added to the model one-by-one to see if they, too, remain in at $p<0.05$.

Variables with no, or very few, missing values were included at the initial stage of the analysis. Such continuous variables are;

- Number of randomizations that make up the trial (NRAND) (0% missing)
- Number of randomization arms (ARMS) (0% missing)
- Start date of randomization (NSTART) (5% missing)

Categorical variables used in the initial stage of the analysis are:

- Age group eligibility (children or both children and adults) (AC) (3% missing)
- Country group of trialists (North America, Europe or Other) (CGROUP) (0% missing)
- 'Developing' country or 'developed' (DEVLPNG) (0% missing)
- Equivalence trial or not (EQUIV) (0% missing)
- First-line or relapse/refractory treatment (FIRSTL) (5% missing)
- Multi-centre, multi-centre(limited) or single-centre participation (MULTIC) (1% missing)
- Treatment type (a frequency table was produced, which to indicated that radiotherapy vs. immunotherapy vs. the rest should be tried) (TXCHEMO) (3% missing)

Continuous variables with too many missing values to be included in the initial regression are:

- Close date of randomization (NCLOSE) (10% missing)
- Duration of accrual period (DURRAN) (10% missing)
- log_{10}(number of patients accrued) (LOGSIZE) (13% missing)

A categorical variable with too many missing values to be included in the initial regression is:

- International, international-limited or single-country participation (INTERNL) (12% missing)

Another variable to be tried at a later stage is:

- Number of questions (NOQ) (0% missing)

This is very highly correlated with ARMS (ρ=0.94), and so is tried in place of ARMS at the point where ARMS is to be dropped as the least significant variable remaining in the model. If ARMS remains in the reduced set of significant variables, NOQ is tried in its place to see whether or not this improved the fit of the model. Similarly NCLOSE is highly correlated with NSTART (again ρ=0.97), and so a similar method is used.

Any variable, either included in the reduced set of significant variables at any stage, or eliminated at $p<0.2$, is tried again at the end of the process. Variables highly correlated with any variable in the model ($\rho>0.3$) are also tried in place of that variable to see if a better fit to the data can be found.

Interaction terms

Once the best fitting main effects model is found, interaction terms are tried.

Interactions which may have an effect are:

- Country group of trialist and duration of accrual period
- Country group of trialist and number of patients accrued
- North American trial with five or more centres participating
- Single-centre trial conducted outside North America and Europe
- Truly international trial, conducted in the US
- Single-country trial, conducted outside North America and Europe
- Conducted in a 'developing' country and number of patients accrued
- Truly international trial conducted in a 'developed' country
- Single-country participation in a trial conducted in a 'developing' country
- Multi-centre trial conducted in a 'developed' country
- Single-centre trial conducted in a 'developing' country

Interaction terms will only be tried if both main effects are either statistically significant or were eliminated from the model with $p<0.2$. Otherwise the interaction is unlikely to affect the response variable in question.

14.3 Which trial characteristics affect whether a randomization is ever mentioned in an article?

Selecting the model that best fits the data

Due to missing data for some variables, there were two possible models to choose from. The probability modelled is for the response 'mentioned'.

Model 1:

Explanatory variables:

- Start date of accrual period (NSTART) $p=0.0011$
- Relapse/refractory vs. first line treatment (FIRSTL) $p=0.0101$
- Conducted in a 'developing' country vs. 'developed' (DEVLPNG) $p=0.0055$
- Duration of accrual period (DURRAN) $p=0.0420$

n = 209/243 observations were used

Residual *SS*= -2 *log L* = 115.105

Residual *df* =205

Model 2:

Explanatory variables:

- Start date of accrual period (NSTART) $p=0.0038$
- Conducted in a 'developing' country vs. 'developed' (DEVLPNG) $p=0.0057$
- log_{10} (number of patients accrued) (LOGSIZE) $p=0.0003$

n = 200/243 observations were used

Residual *SS*= -2 *log L* = 69.863

Residual *df* =197

Since the two models in question are not nested and missing values mean that they may be using different observations, it was necessary to apply the methods described in Section 5.3.3 to select the model that best fits the data. Both models were run using (i) common observation only and (ii) using all 243 observations, but replacing all missing values for a variable by the mean. Both methods gave the same conclusion, that Model 2 has a smaller value for -2 *log L* than Model 1, and so is a better fit to the data.

Interpreting the results of the chosen model

The output for the chosen model is given as Table 14.1.

The LOGISTIC Procedure
Number of Observations 200

Response Profile

Ordered Value	MENTND	Total Frequency
1	1	186
2	0	14

Probability modeled is MENTND=1.

NOTE: 43 observations were deleted due to missing values for the response or explanatory variables.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	103.456	77.863
SC	106.754	91.057
-2 Log L	101.456	69.863

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Exp(Est)
Intercept	1	0.0103	2.0455	0.0000	0.9960	1.010
NSTART	1	-0.00054	0.000186	8.3799	0.0038	0.999
DEVLPNG 0	1	1.2853	0.4649	7.6445	0.0057	3.616
LOGSIZE	1	2.9060	0.8036	13.0782	0.0003	18.284

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits
NSTART	0.999	0.999 1.000
DEVLPNG 0 vs. 1	13.074	2.114 80.875
LOGSIZE	18.284	3.785 88.320

Note
DEVLPNG=0 denotes a 'developed' country
DEVLPNG=1 denotes a 'developing' country

Table 14.1 Output from the 'ever-mentioned/never mentioned' logistic regression analyses

14.3.1 Method for calculating an odds ratio (OR)

The general case

If p = the probability of an event occurring, given a particular characteristic

then $1 - p$ = the probability that it does not occur, given that characteristic.

and the odds it occurring, given that characteristic are $p / (1 - p)$

and $logit\ (p) = log\ [p / (1 - p)] = (\alpha + \beta_1 x_1 + \ldots + \beta_m x_m)$, for a model with m explanatory variables.

Similarly,

if p' = the probability of an event occurring, without that characteristic

then $1 - p'$ = the probability that it does not occur, without that characteristic

and the odds it occurring, without that characteristic are $p' / (1 - p')$

and $logit\ (p') = log\ [p' / (1 - p')] = (\alpha + \beta_1 x_1' + \ldots + \beta_m x_m')$, say.

Therefore, the odds ratio (OR) for the event occurring, characteristic present: not present is

$$OR = \frac{p / (1 - p)}{p' / (1 - p')}$$

$$log\ (OR) = log\ [p / (1 - p)] - log\ [p' / (1 - p')]$$

$$= (\alpha + \beta_1 x_1 + \ldots + \beta_m x_m) - (\alpha + \beta_1 x_1' + \ldots + \beta_m x_m')$$

If the characteristic that differs between the two groups is x_j and all other variables are the same

i.e. if $x_i' = x_i$ for all $i \neq j$, then

$$log\ (OR) = \beta_j (x_j - x_j')$$

So $$OR = exp\ [\beta_j (x_j - x_j')]$$

Application to these data

Let p = the probability that a randomization conducted in a developed country is **ever** mentioned in an article

and p' = the probability that a randomization conducted in a developing country is **ever** mentioned in an article

In this analysis the only categorical variable in the model is whether the trial was conducted in a 'developed' or a 'developing' country (DEVLPNG).

Let variable DEVLPNG = x_1 = 1 if randomization was conducted in a 'developed' country.

Let variable DEVLPNG = x_1' = -1 if randomization was conducted in a 'developing' country.

$$OR = exp\,[\beta_1 (x_1 - x_1')] = exp\ 2\beta$$

From Table 14.1 it can be seen that $\beta_1 = 1.2853$, $x_1 = 1$ and $x_1' = -1$ and that the odds ratio is 13.074.

To test the fit of the model:

$\chi^2 = 69.863$ with 196 *df*, $p = 0.9999$

i.e. the model is an excellent fit to the data.

Summary

Dependent variable:	Randomization ever mentioned/never mentioned, in a publication
Independent variables:	Conducted in a 'developed'/ 'developing' country
Continuous covariates:	Start date of accrual period
	log_{10} (number of patients accrued)

Fitted constants for odds of being mentioned

Overall constant	1.010	
'Developed'/'developing' country		$\chi^2 = 7.6445$, *df*=1, *p*-value= 0.0057
'Developed'	3.616	
'Developing'	0.277	
Start date of accrual period	0.999	$\chi^2 = 8.3799$, *df*=1, *p*-value = 0.0038
log_{10} (number of patients accrued)	18.284	$\chi^2 = 13.0782$, *df*=1, *p*-value = 0.0003

Therefore:

- The odds ratio for a randomization ever being mentioned in an article if conducted in a 'developed' country to if conducted in a 'developing' country is 13.074.

- The odds ratio of being mentioned for any start date of accrual period vs. the day before is 0.999 i.e. the odds that a randomization is ever mentioned is slightly greater for earlier than for more recent trials, as would be expected.
- The odds of ever being mentioned are greater for randomizations which accrued a greater number of patients. The odds ratio for a difference in log_{10} (number of patients accrued) of 1.0. i.e. for a tenfold increase in the number of patients accrued, is 18.284.

14.4 Which trial characteristics affect whether the results of a randomization are ever reported?

Interpreting the model

The output is given as Table 14.2.

Number of Observations 231

Response Profile

Ordered Value	RESPUB	Total Frequency
1	1	177
2	0	54

Probability modeled is RESPUB=1.

NOTE: 12 observations were deleted due to missing values for the response or explanatory variables.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	253.230	237.057
SC	256.672	247.384
-2 Log L	251.230	231.057

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Exp(Est)
Intercept	1	1.9220	0.6322	9.2413	0.0024	6.834
DEVLPNG 0	1	1.0273	0.3001	11.7136	0.0006	2.793
NSTART	1	-0.00023	0.000081	7.9034	0.0049	1.000

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
DEVLPNG 0 vs. 1	7.803	2.406	25.306
NSTART	1.000	1.000	1.000

Table 14.2 Output from the 'results ever reported/never reported' logistic regression analysis

The probability modelled is for the response 'results reported'.

Explanatory variables:

- Conducted in a 'developing' country vs. 'developed' (DEVLPNG) p=0.0006
- Start date of accrual period (NSTART) p=0.0049

n = 231/243 observations were used

Residual *SS*= -2 *log L* = 231.057

Residual *df* =228

From Table 14.2 it can be seen that the odds ratio for the results of a randomization ever being reported in an article for randomizations conducted in a 'developed' country versus those conducted in a 'developing' country is 7.803.

To test the fit of the model:

χ^2 = 231.057 with 228 *df*, p=0.4495

So the model is a fairly good fit to the data.

Summary

Dependent variable: Results of randomization ever reported / never reported

Independent variables: Conducted in a 'developed'/ 'developing' country

Continuous covariate: Start date of accrual period

Fitted constants for odds of being mentioned

Overall constant	6.834	
'Developed'/'developing' country		χ^2 = 11.7136, *df*=1, *p*-value = 0.0006
'Developed'	2.793	
'Developing'	0.358	
Start date of accrual period	1.000	χ^2 = 7.9034, *df*=1, *p*-value= 0.0049

Therefore:

- The odds ratio for a randomization ever being mentioned in an article if conducted in a 'developed' country to if conducted in a 'developing' country is 7.803.
- The odds ratio of being mentioned for any start date of accrual period vs. the day before is very slightly less than 1.000, so the odds that a randomization is ever mentioned is slightly greater for earlier than for more recent trials.

The model fits the data fairly well. However it contains only two explanatory variables. All other unchanging, 'definitive records' variables were tried and no others were found to be significant.

A possible explanation for this is that the nature of the results of the randomization probably has a great effect on whether they are published. Therefore new variables must be created to do with statistical significance, direction of results, clinical significance and whether the main questions of the randomization were ever answered. So far each reporting of results contains data specific to a particular publication. The next step would be to create fields to be added to the definitive record for each randomization. For example, statistical significance of the most statistically significant result ever obtained for each randomization.

In order to have this information for unpublished randomizations, it would be necessary to contact the trialists, and in order to avoid bias, it would be necessary to do the same for all published randomizations also. This would involve designing and sending out a questionnaire asking for at least the following information:

- The main questions being asked in each randomization e.g. survival, rate of relapse in a specific site etc.
- Whether the answers to the questions have ever been published
- Data on statistical significance, direction of results and clinical significance that can be used as a 'definitive result', for example *n* years after the randomization opened or closed
- If the randomization has remained unpublished, the reasons for this. Was it never analysed? Was it analysed, but its findings not considered worth reporting?

These data could then be added to the definitive record for each randomization and each new variable tried in the logistic regression, in order to obtain a better fitting model. It is intended that this should be undertaken in the future.

14.5 Conclusions

Earlier randomizations conducted in a 'developed' country are more likely both to have been mentioned in an article, and to have had their findings published, than those without these characteristics. If, in addition, a large number of patients was accrued, the likelihood of having been mentioned is even greater, although this latter trial characteristic does not increase the probability that the results of the randomization have been reported.

Although it may seem that earlier randomizations must necessarily have a greater chance of having been published, it should be remembered that all randomizations included in this project began prior to 1 January l988, and all articles relating to these, published prior to 1 January 2000 and identified by the cut-off date for analysis, 28 November 2000, allowing ample time for the accrual and follow-up periods and the publication process.

15 OTHER INVESTIGATIONS II: AN INVESTIGATION INTO WHICH TRIAL CHARACTERISTICS AFFECT (a) THE FREQUENCY OF MENTIONS OF A RANDOMIZATION AND (b) THE FREQUENCY OF REPORTING OF THE RESULTS OF A RANDOMIZATION?

15.1 Introduction

This chapter deals with the second 'How wide?' question, seeking to discover which trial characteristics affect frequency of publication.

15.2 Data used for the analyses

As with the analyses with response variable 'mentioned/not mentioned' and 'results reported/not reported', the 'definitive records' dataset is used for investigating which trial characteristics affect 'frequency of mentions'(NMENT) and 'frequency of reporting of results' (NRES). There is one record for each randomization, whether published or unpublished. Again, the only variables tried are those which do not change over time, which were described in detail in Section 14.2. The number of articles in which the randomization has been mentioned, and the number of articles in which its results have been reported are also attached to the definitive record.

The date at which the data to be analysed were frozen was 28 November 2000. Two new variables, STARTCUT and CLOSECUT were created. These are the time periods from the start of the accrual period for a randomization until the cut-off date of 28 November 2000 and from the close of the accrual period until the cut-off date respectively. The start date of the accrual period was chosen for use in the 'frequency of mentions' analysis so that mentions of a randomization before that randomization closed fall within the time period. For the 'frequency of reporting of results' analysis, the close date of the accrual period was used.

15.3 Which trial characteristics affect the frequency of mentions of a randomization?

Choosing an appropriate model

To check whether a Poisson regression model is suitable for the 'frequency of mentions' analysis, an ordinary linear regression was run with the response variable

$$\sqrt{(\text{NMENT/STARTCUT})} = \alpha + \sum_{j=1}^{p} \beta_j x_j$$

where x_j, $j=1 \ldots p$ are explanatory variables,

and α and β_j, $j=1 \ldots p$ are parameter estimates

$\sqrt{(\text{NMENT/STARTCUT})}$ is the variance stabilizing transformation. If the variable, NMENT, has a Poisson distribution, the plot of residuals against fitted values forms a random pattern, and that of residuals against normal scores an approximately straight line. If, however, the plot of residuals against fitted values produces a wedge-shaped pattern, this would indicate overdispersion, i.e. that the mean has a distribution with the variance increasing more rapidly than the mean, as the mean increases. In such a case, the assumption of the Poisson model,

$$\text{variance} = \text{mean} = \mu$$

is not satisfied.

Instead, the negative binomial model, with relationship between the mean and variance,

$$\text{variance} = \mu + k\mu^2, \qquad \text{where } k = \text{scale factor}$$

should be used.

The explanatory variables, used in this preliminary investigation, were any of the 'definitive records' variables significant in either the 'time from close to publication for all mentions' analysis or the 'ever mentioned/never mentioned' logistic regression. These are date of start of accrual period (NSTART), duration of accrual period (DURRAN), country group of trialists (CGROUP), whether or not the randomization was conducted in a 'developing' country (DEVLPNG), degree of multi-centre participation (MULTIC), log_{10}(number of patients accrued) LOGSIZE and whether first-line or relapse treatment (FIRSTL).

The diagnostic plots obtained are given as Figures 15.1 and 15.2.

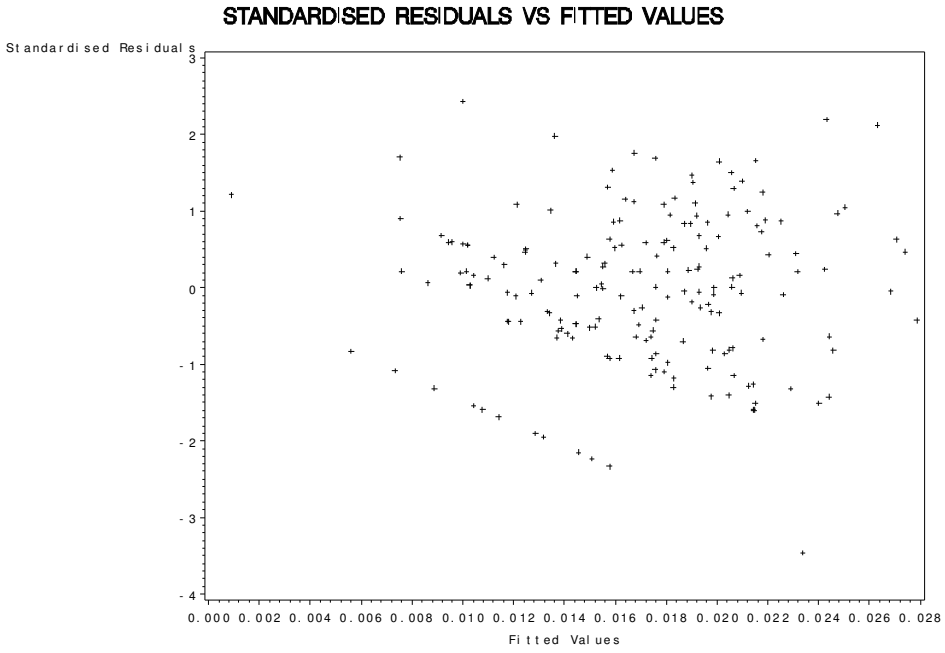


Figure 15.1 Modelling frequency of mentions using ordinary linear regression with the variance stabilizing transformation. Standardised residuals vs. fitted values

In Figure 15.1, the top half of the plot of standardised residuals against fitted values shows a random pattern, indicating that 'frequency of mentions' does have a Poisson distribution. The diagonal line in the lower half of the plot may be thought to indicate the opposite. However this is misleading for the reason that the response variable, 'frequency of mentions' (NMENT), is an integer, with lowest possible value zero.

For any observation y_i,

residual $\varepsilon_i = y_i$ – fitted value

$$= y_i - (\alpha_i + \sum_{j=1}^{p} \beta_{ij}\, x_{ij})$$

Therefore, for those observations for which $y_i = 0$, i.e. for those randomizations never published,

$\varepsilon_i =$ – fitted value

$$= -(\alpha_i + \sum_{j=1}^{p} \beta_{ij}\, x_{ij})$$

These observations form the diagonal line in the lower half of the plot. Since y_i cannot take a value lower than zero, it is not possible for any residuals to lie beneath this line. Therefore the lower half of the plot cannot be used to determine the suitability of the Poisson model.

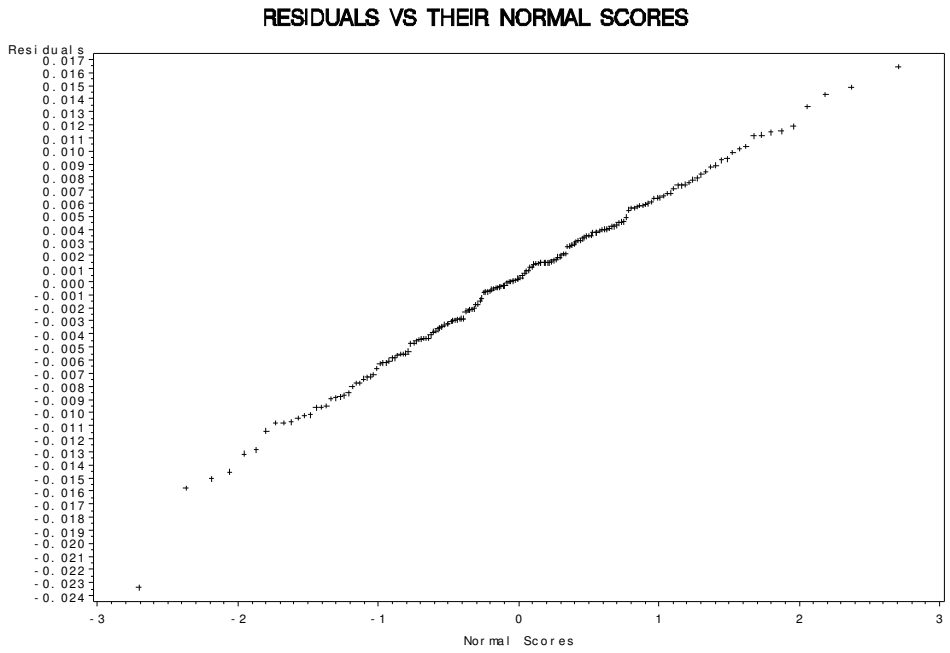


Figure 15.2 Modelling frequency of mentions using ordinary linear regression with the variance stabilizing transformation. Residuals vs. their normal scores

In Figure 15.2 the plot of residuals versus normal scores, with the exception of one observation (bottom left-hand corner), also suggests that the Poisson model may be appropriate.

Both the Poisson and negative binomial models are now applied, using a *log* link, and the results compared:

$$log\ (\text{NMENT/STARTCUT}) = \alpha + \sum_{j=1}^{p} \beta_j x_j$$

$$log\ \text{NMENT} = \alpha + \sum_{j=1}^{p} \beta_j x_j - log\ (\text{STARTCUT})$$

The same set of explanatory variables was used initially, with a backwards elimination process, dropping the least significant each time, until a reduced set of significant variables was obtained.

Other ‘definitive records’ variables were then added in, one by one, until the best-fitting model was found.

Interactions between country group of trialists (CGROUP) and whether multi-centre participation took place (MULTIC), whether there was international participation (INTERNL), and size of trial (LOGSIZE) were tried, as were those between whether conducted in a ‘developing’ country (DEVLPNG) with MULTIC, INTERNL and LOGSIZE respectively. None were found to be significant.

The algorithm for the program and some SAS commands used to perform the negative binomial and Poisson regressions are given in Appendix XX, and the output for the final models obtained are in Appendix XXI. The findings are summarized below:

Summary of findings

1. Using the negative binomial model

The following trial characteristics lead to a randomization being mentioned in a greater number of articles:

- Later start date of accrual period (NSTART) P(<0.0001)
- Conducted in a ‘developed’ country (DEVLPNG) (p=0.0028)
- Single-centre participation (MULTIC) (p=0.0006)
- Greater number of patients accrued (LOGSIZE) (p<0.0001)
- Equivalence trial (EQUIV) (p=0.0183)
- First-line treatment as opposed to treatment for relapse/refractory disease (FIRSTL) (p=0.0079)
- Conducted in North America (CGROUP) (p=0.0165)

Deviance=185.3164, 180 *df*

Deviance/ *df* = 1.0295

This is based on 188/243 of the randomizations.

The scale factor (negative binomial dispersion parameter) was estimated by maximum likelihood. k=0.0214, standard error of k=0.0363.

Diagnostic plots are given as Figures 15.3 and 15.4. The residuals used are the Pearson (χ^2) residuals for identifying poorly fitted observations.

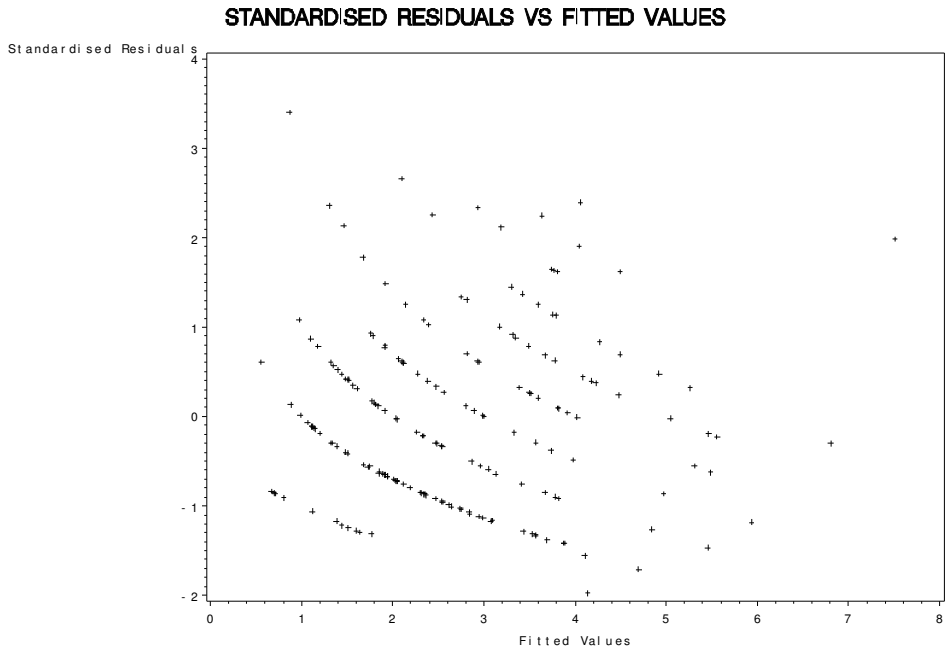


Figure 15.3 Modelling frequency of mentions using a negative binomial regression. Standardised residuals vs. fitted values

With the exception of one outlying observation (top left hand corner) Figure 15.3 shows a fairly random spread, indicating a constant variance. The series of parallel lines are due to the discrete nature of the response variable.

Outlying observation

The outlying observation, with standardised residual of 3.41, is record 349 from the 'definitive records' database, the randomization 1209. This is from a single randomization trial, ALL-REZ-BFM-85, (trial 1209), which was run by the BFM children group in Germany, for the short period April 1985-March 1986. It was a post-relapse randomization, which involved multi-centre participation and accrued 46 patients. The two treatment arms were high dose methotrexate infused over a short period versus intermediate dose methotrexate over a longer period, both

followed by folinic acid, and it is not thought to be an equivalence trial. It was mentioned in four articles, more frequently than might be expected.

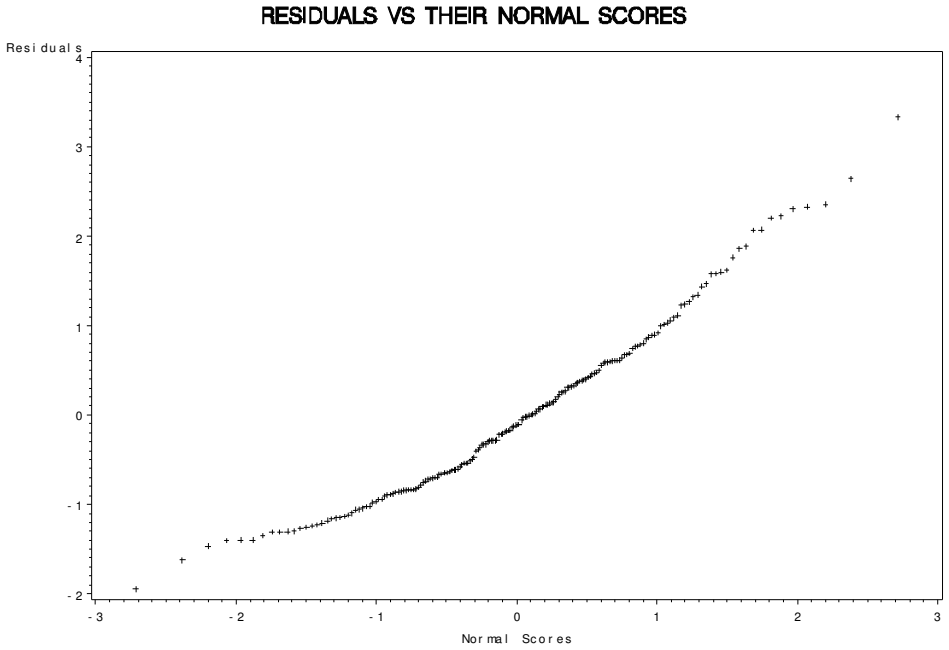


Figure 15.4 Modelling frequency of mentions using a negative binomial regression. Residuals vs. their Normal scores

Figure 15.4 shows a fairly straight line, indicating a normal distribution.

2. Using the Poisson model

Output from the Poisson regression is very similar.

The following trial characteristics lead to a randomization being mentioned in a greater number of articles:

- Later start date of accrual period (NSTART) ($p<0.0001$)
- Conducted in a 'developed' country (DEVLPNG) ($p=0.0025$)
- Single-centre participation (MULTIC) ($p=0.0004$)
- Greater number of patients accrued (LOGSIZE) ($p<0.0001$)
- Equivalence trial (EQUIV) ($p=0.0137$)
- First-line treatment as opposed to treatment for relapse/refractory disease (FIRSTL) ($p=0.0067$)
- Conducted in North America (CGROUP) ($p=0.0131$)

Deviance=196.2034, 180 *df*

Deviance/*df* =1.0900

This is based on 188/243 of the randomizations.

The diagnostic plots given as Figures 15.5 and 15.6.

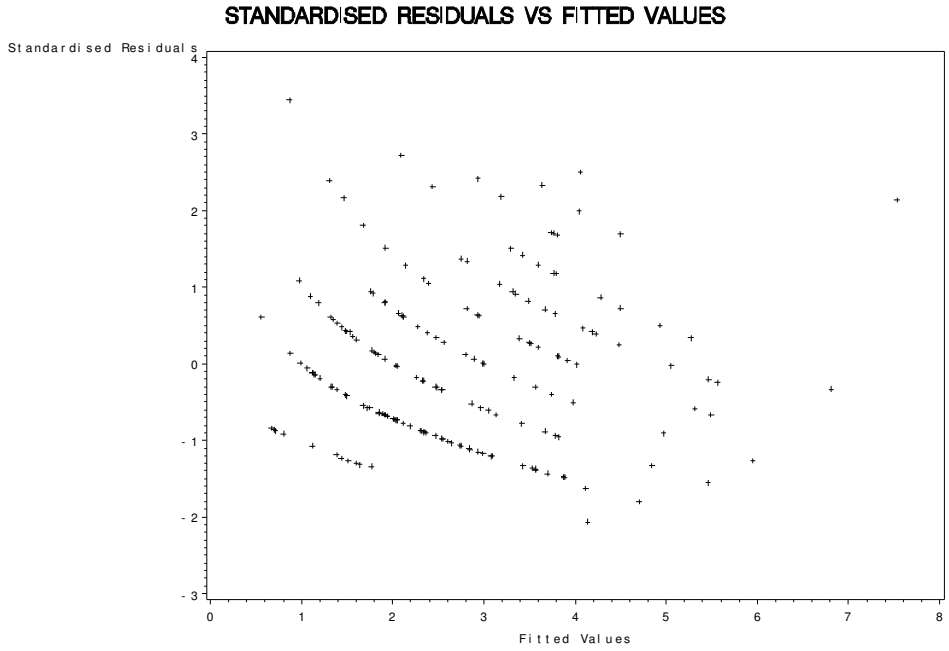


Figure 15.5 Modelling frequency of mentions using a Poisson regression. Standardised residuals vs. fitted values

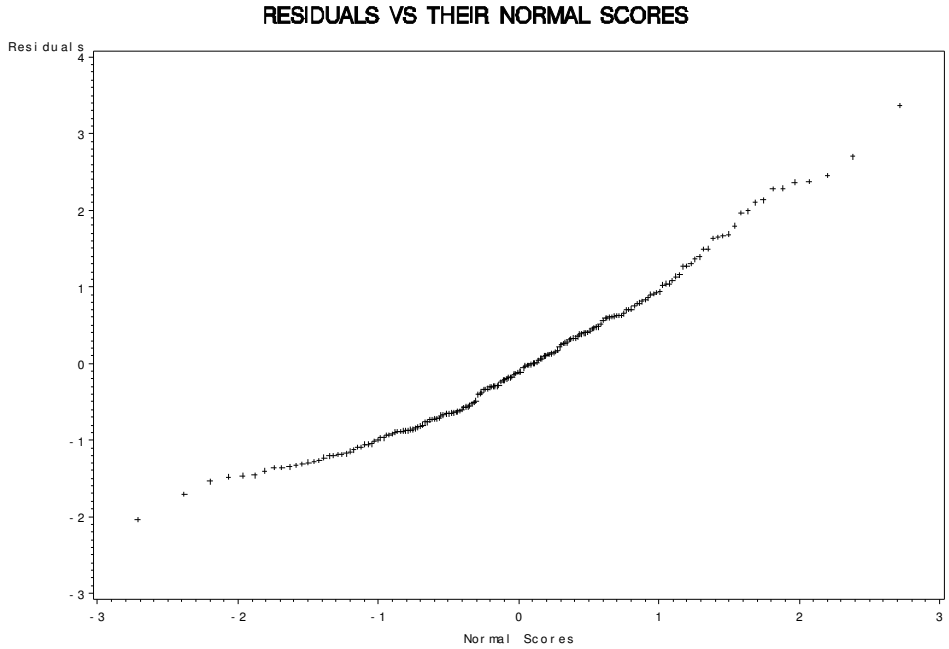


Figure 15.6 Modelling frequency of mentions using a Poisson regression. Residuals vs. their normal scores

Clearly, the diagnostic plots from the Poisson regression (Figures 15.5 and 15.6) are very similar to those from the negative binomial model (Figures 15.3 and 15.4). The outlier in Figures 15.3 and 15.5 is the same observation, and has already been discussed.

Conclusion

The same variables were found to be significant using either model, and the *p*-values associated with them did not differ much, but were slightly smaller with the negative binomial model than with the Poisson model. The deviance was slightly smaller when the former was used (185.3164 as opposed to 196.2034 for the same number of degrees of freedom 180).

The value of *k*, the scale factor in the negative binomial regression is small (0.0214). The largest number of mentions for any randomization is 13. The randomization in question is 3107, SJCRH X, which was discussed in Section 1.4. From Figure 15.3 it can be seen that the largest fitted value is approximately 7.5, that for randomization 3107.

Since variance = $\mu + k\mu^2$, the greatest value that the variance can take is approximately $7.5 + 0.0214 \times 7.5^2 = 8.7$, which is not much greater than the mean.

For these reasons it is immaterial which of the two models is used.

15.4 Which trial characteristics affect the frequency of reporting of the results of a randomization?

Choosing an appropriate model

Again the suitability of the Poisson regression model was tested by use of the variance stabilizing transformation,

$$\sqrt{(\text{NRES/CLOSECUT})} = \alpha + \sum_{j=1}^{p} \beta_j x_j$$

where x_j, j=1 … p are explanatory variables

and α and β_j, j=1 … p are parameter estimates.

The only ‘definitive records’ explanatory variable in the ‘time from close to publication’ analysis for all reportings of results was date of close of accrual period (NCLOSE). Date of start of accrual period (NSTART), which is highly correlated with NCLOSE, and whether or not conducted in a ‘developing’ country (DEVLPNG) were the only significant variables in the ‘results published/not published’ logistic regression analysis. For the purpose of this preliminary investigation, NCLOSE and DEVLPNG were used, plus other variables considered likely to affect the response, i.e. those tried in the ‘frequency of mentions’ preliminary investigation. Since NSTART is highly correlated with NCLOSE, the former was not included.

The diagnostic plots obtained are given as Figures 15.7 and 15.8.

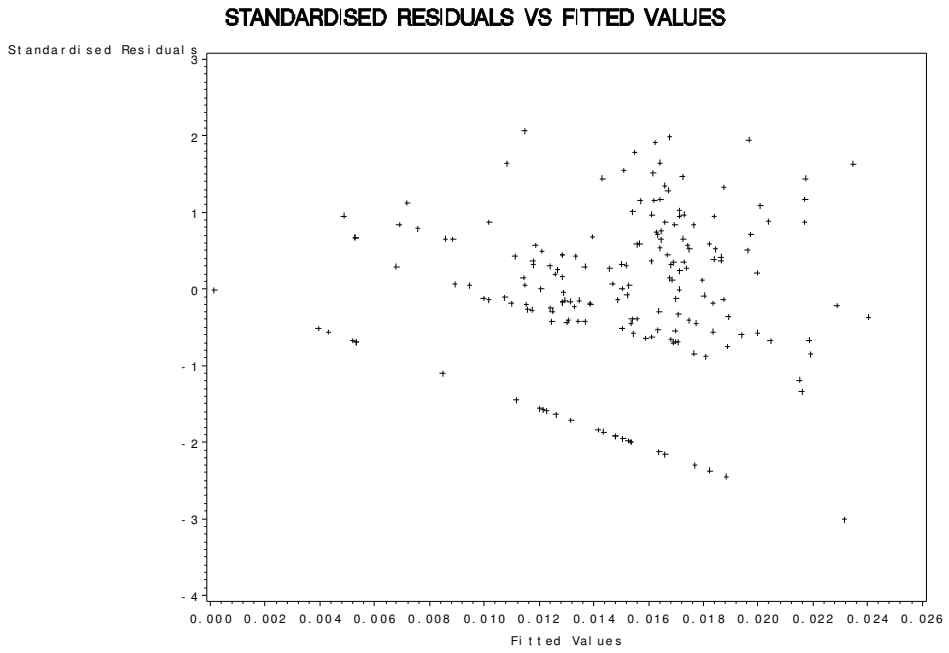


Figure 15.7 Modelling frequency of reporting of results using ordinary linear regression with the variance stabilizing transformation. Standardised residuals vs. their fitted values

For the same reason as with the preliminary investigation for the ‘frequency of mentions’ analysis, the lower half of the plot of standardised residuals versus fitted values (Figure 15.7) should not be used to assess whether the distribution is wedge-shaped. The upper half indicates that it is not, and that a Poisson model is therefore suitable. The same negative residuals are responsible for the kink in an otherwise fairly straight line in the plot of residuals versus their normal scores (Figure 15.8).

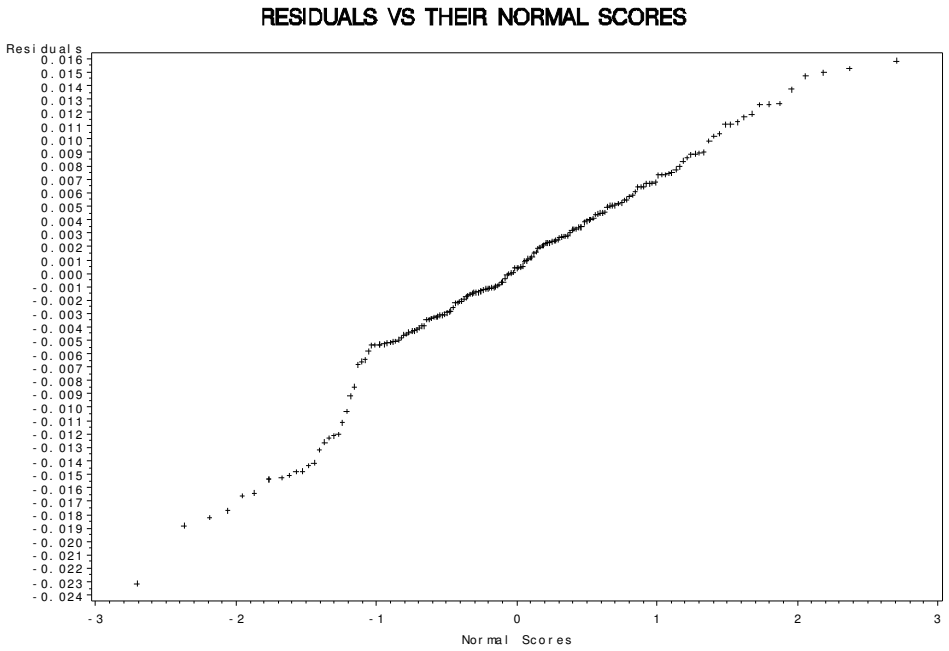


Figure 15.8 Modelling frequency of reporting of results using ordinary linear regression with the variance stabilizing transformation. Residuals vs. their normal scores

Again, both the negative binomial and Poisson regressions were run. For this analysis the results obtained in the two ways were identical. When the former was used, the dispersion parameter was estimated to be zero. This should be treated with caution since the model did not converge. However, since the results were identical when the Poisson model was used, and there were no convergence problems with this latter model, the results are confirmed.

Summary of findings

The following trial characteristics lead to the results of a randomization being reported in a greater number of articles:

- Later close date of accrual period (NCLOSE) ($p<0.0001$)
- Conducted in a 'developed' country (DEVLPNG) ($p=0.0004$)
- Single-centre participation (MULTIC) ($p<0.0001$)
- Number of patients accrued (LOGSIZE) ($p<0.0006$)
- Treatment type: immunotherapy or radiotherapy as opposed to chemotherapy or antibiotic (TXCHEMO) ($p=0.0031$)

Deviance=193.6045 182 *df*

Deviance/*df*=1.0638

This is based on 188/243 of the randomizations.

The diagnostic plots are given as Figures 15.9 and 15.10.

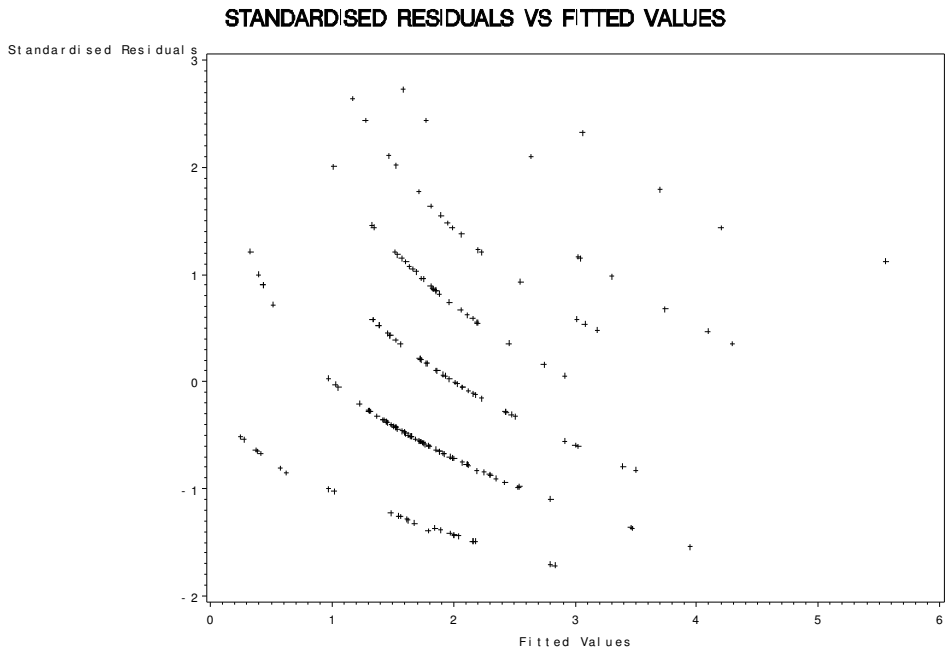


Figure 15.9 Modelling frequency of reporting of results using either the negative binomial or Poisson regression. Standardised residuals vs. fitted values

Again, the plot of standardised residuals versus fitted values (Figure 15.9) gives a random spread, and the plot of residuals versus their normal scores (Figure 15.10) is fairly straight, indicating the model is a good fit to the data.

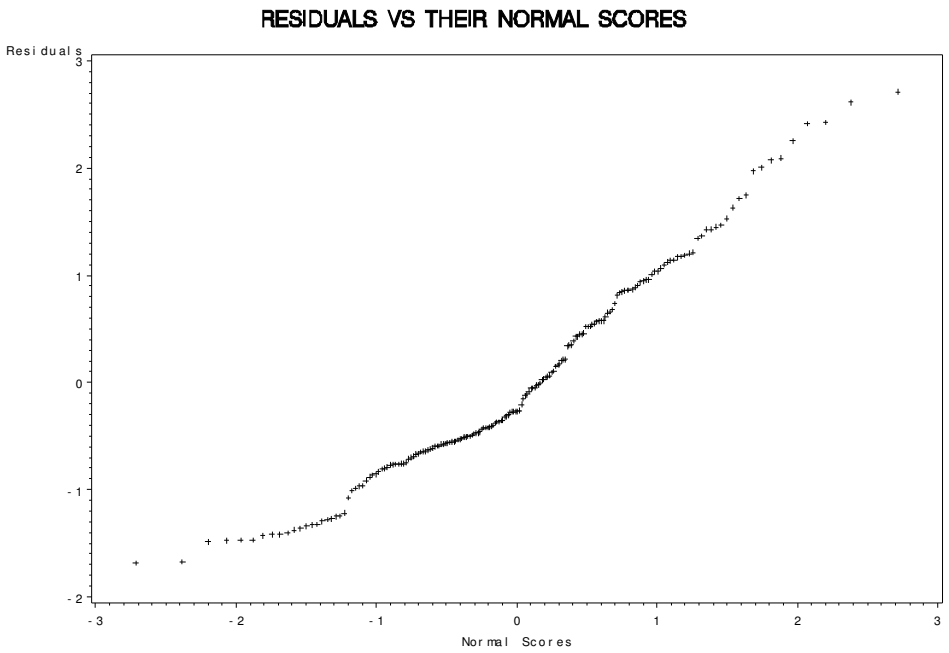


Figure 15.10 Modelling frequency of reporting of results using either the negative binomial or Poisson regression. Residuals vs. their normal scores

15.5 Conclusions

Several of the trial characteristics associated with a higher frequency of mentions are also associated with a higher frequency of reporting of results. These are, as may be expected, randomizations from trials conducted in 'developed' countries and those accruing a larger number of patients. Surprisingly, in both analyses, single-centre, as opposed to multi-centre, participation also leads to a greater frequency of publication. A possible explanation for this is that there are likely to be fewer co-authors of reports of results from single-centre randomizations and so the agreement of fewer people would be necessary in order for articles to be submitted. Later start date of accrual period was significant in the 'frequency of mentions' analysis, as was later close date of accrual period in the 'frequency of reporting results' analysis. These variables are very highly correlated (ρ=0.97), and both imply, encouragingly, that more recent randomizations are published more widely than earlier randomizations.

In addition, randomizations between treatments for first-line therapy tend to be mentioned in more articles than those for relapse/refractory disease, as do randomizations conducted in North America

and equivalence trials. However these factors do not affect the frequency of reporting results. Results from immunotherapy and radiotherapy trials were reported more frequently than those from chemotherapy or antibiotic trials, although there was no difference in the number of times mentioned.

16 UPDATE

16.1 Introduction

This short chapter summarises new data found since the analyses were run.

16.2 Data

In recent years, attempts have been made to update the list of randomized trials for the overview (meta-analysis), but these attempts have not yet been as comprehensive or systematic as the earlier searching which formed the basis for the list of trials used in this thesis.

The last comprehensive search of MEDLINE and EMBASE was done on 6 September 1999. After then, the processing of publications on two treatment questions only was continuously updated. These were central nervous system treatment and duration of therapy, which were topics for overviews and collaborative meetings. Searches were done targeting these specific types of treatment only, but if other trials were found by chance, they too were added to the database. On 28 November 2000 the Cancer Overviews Group decided that there would be no systematic update of the ALL database within the next 12 months. Therefore all analyses reported here have used the data frozen at that point.

The trial lists were checked for the last time on 1 June 2002 and showed that this fairly limited searching has not found any additional trials or randomizations starting prior to 1 January 1988. However, seven additional articles, published before 1 January 2000, relating to randomizations already identified had been found. Also, it had become apparent that one randomization had been registered twice, with two different ID numbers and under two different groups of trialists.

A summary of the seven new articles found:

1. Schrappe et al (1998) published in *Klinische Padiatrie*, is a report by the BFM group (Group 12), describing four of their trials: ALL-BFM- 81, ALL-BFM-83, ALL-BFM-86 and ALL-BFM-90.The last two digits of the trial name indicate the year the trial opened. The latter is ineligible for

inclusion in this study since the entry period started after 1 January 1988. The report demonstrates the efficiency of CNS-directed chemotherapy with or without reduced cranial radiotherapy in preventing CNS relapse. It is published in full in English with the abstract in German also. (The other six articles are published only in English.) ALL-BFM-81 (1204) contains two randomizations; maintenance duration (1204) and CNS prophylaxis; radiotherapy versus chemotherapy (intravenous methotrexate) (1205). ALL-BFM-83 (1206) consists of three randomizations; maintenance duration (1206), late/second reinduction block versus control (1207), and two doses of radiotherapy for CNS prophylaxis (1208). ALL-BFM-86 (1211) has one randomization (1211); intensification block versus not during maintenance treatment. 1211 had been reported in one other article, the others in several.

2. Freeman et al (1997) in *Medical and Pediatric Oncology*, describes the US Cancer and Leukemia Group B (CALGB) trial CLB 7611 (1416). This trial has a single randomization (1416) for CNS prophylaxis treatment; cranial irradiation versus chemotherapy in the form of intravenous methotrexate. Again this randomization, open between 1976 and 1979, had already been reported many times.

3. Miller et al (1981) is technically a book reference, although *Haematology and Blood Transfusion*, in which the article is published, is similar to a journal. This reports trial CCG-141 (1612) by the Children's Cancer Group (CCG) (Group 16). This has two randomizations, between two induction and consolidation treatments (1612) and between three maintenance treatments (1615). The maintenance treatments are, after three years, to stop treatment, to have a further 4-week block of treatment or to continue maintenance for another two years. Again several articles reporting each randomization had already been found.

4. Miller (1981) was published in the book *Cancer: Achievements, Challenges and Prospects for the 1980's*. Again the trial reported is CCG-141.

5. Gelber et al (1993), published in *Cancer*, is entitled 'Central Nervous System Treatment in Childhood Acute Lymphoblastic Leukemia'. It reports four trials, each comprising a single randomization, run by the Dana Farber Cancer Institute (DFCI), (Group 17). DFCI 73001 (1701) compares three induction treatments, DFCI 77001 (1702) compares two consolidation treatments, DFCI 80001 (1703) is between two intensification blocks and DFCI 81001(1704) compares a pre-induction treatment against a control. In each case the first two digits of the trial name indicate the year the accrual period began. Again all four randomizations have been published in other articles, 1703 in only one other.

6. Koizumi et al (1997) reports some of the randomizations belonging to four trials conducted by The Japanese Children's Cancer and Leukaemia Group (Group 26) in *Medical and Pediatric Oncology*. In each case the first two digits of the trial name indicate the year the trial began. JCCLSG S-811 (2601) compares two maintenance treatments for standard risk patients and JCCLSG H-811 (2602) two maintenance treatments for high risk patients (not reported in this article). JCCLSG I-841 (2603) contains two randomizations; 2603 is between two induction and maintenance regimens for intermediate risk patients (not reported in this article), one of which includes cranial irradiation, and 2604 is a two-arm induction randomization for low-risk patients. JCCLSG L-874 (2606) comprises three randomizations. 2606 is a CNS prophylaxis randomization for low risk patients, between cranial irradiation and chemotherapy (methotrexate + leucovorin rescue) and JCCLSG I-874 (2607) is a similar randomization for intermediate risk patients. JCCLSG H-874 (2608) is between two intensification treatments for high risk patients (not reported in this article). The fourth trial JCCLSG-911 is ineligible for inclusion since it opened after 1 January 1988. Again, other articles for all randomizations have already been found, although only one other each for randomizations 2602 and 2605.

7. Nishimura (1993) mentions four trials run by the Tokyo Children's Cancer Study Group, Japan (Group 127) in *Cancer Chemotherapy: Challenges for the Future*. Two are eligible for inclusion in these data. Again the first two digits specify the year the trial began. The first of these is TCCSG

84-11 (12701), which has two randomizations; TCCSG L84-11 SR (12701) and TCCSG L84-11 HR (12702), CNS prophylaxis randomizations of chemotherapy with intravenous and intrathecal methotrexate versus control for standard and high risk patients respectively. The other trial is TCLSG L81-10 (12703) for standard risk patients, which randomizes between two doses of cranial irradiation (12703) and between two consolidation and randomization treatments. The other two trials reported; TCCSG L-89-12 and L92-13 began too late to be eligible for inclusion. Two other articles reporting randomization 12701 have been found, and one other for the other three randomizations of interest.

To summarise; all the randomizations reported in the seven newly found articles have also been reported elsewhere. All seven articles are concerned, at least in part, with one or both of the two questions addressed for the last ALL Collaborative Overview, namely CNS therapy and duration of maintenance treatment. Four articles (1,2,6 and 7) mention randomizations between CNS prophylaxis treatments. Article 5 reports trials that do not randomize between CNS treatments, but nonetheless focuses on CNS treatments. Articles 1, 3 and 4 mention randomizations between maintenance treatments of differing duration.

The randomization which has been included twice is from the POG7623/SWOG7623/AlinC trial (2902 and 3218). It is under the Pediatric Group (POG) (Group 29) as randomization 2903, between two combined induction, consolidation and maintenance treatments, one of which involves cranial irradiation. It is also under the Southwest Oncology Group (SWOG) as a trial with a single randomization between two CNS treatments; cranial irradiation plus chemotherapy with methotrexate versus more methotrexate. Instead of two randomizations reported in one and two articles respectively, this should be considered as one randomization reported in three articles, publications 177 [van Eys et al (1989b)], 251 [Wells et al (1983)] and 252 [Whitt et al (1984)].

Since 257 articles have been included in the data analysed, the small amount of additional data (less than 3%) should not affect the results reported.

17 SUMMARY OF THE MAIN CONCLUSIONS

17.1 Introduction

This final chapter summarises the main findings of this thesis i.e. the chief factors leading to publication bias and the importance of these to meta-analysts.

17.2 The 'How long?' analyses

Of the twelve 'How long' analyses reported here, the most informative are those of time from close to publication for the first reporting of results and for the first mention of each randomization.

The first is important since this is the delay between the close of the randomization and making the results widely available through publication to clinicians treating patients, who may change their practices as a result of what they read. This analysis showed that results with smaller *p*-values associated with them are first published more quickly than those with larger *p*-values. This is in concordance with other similar studies conducted. It was also found that null results (where the trial found in favour of neither treatment arm) are generally first published faster than those in favour of either the experimental arm or the standard (control) arm, but that null results from trials conducted outside North America and Europe had a longer time to first reporting of results. It is pleasing to note that this project did not find direction of results (i.e. whether in favour of the experimental or the standard arm) to be a significant predictor of time to first reporting of results. This was not the case with most of the other studies. It was also found that if the trialists indicated, in an article, that that they considered results of their randomization were either clearly in favour of the experimental treatment arm, or that this was clearly not so, the time from close to publication was shorter than if no clear impression was given, irrespective of the *p*-values associated with the results. This demonstrates that the importance the trialists attach to their results and the motivation of the trialists to publish also influence time to publication.

Surprisingly, and in contrast to the findings of the published studies, it was found that results in favour of the standard or control arm, were submitted more quickly than all others, including those

clearly in favour of the experimental arm. Since this was not found to be significant in the analysis of time from close to publication for first reportings of results, this may indicate that although trialists may be keen to publish their negative findings, journal editors may be less so.

The second analysis showed that neither statistical significance nor direction of results are significant predictors of the time from close to publication, for the first mention of a randomization. This should be reassuring for those involved in identifying randomizations for inclusion in meta-analyses. This analysis also found that time from close to first mention was longer for randomizations conducted by 'developing' countries.

When the analyses of time from close to publication (and from close to submission) were re-run omitting records relating to reportings of randomizations prior to closure, typically the model changed significantly. It appears that duration of randomization is acting as a surrogate for whether or not the randomization is mentioned prior to closure. It is intended that this interesting finding will be investgated fully in future.

17.3 The 'How wide? analyses

Turning to the 'How wide?' analyses; neither statistical significance, direction of results nor any other variable associated with a particular publication is used in these analyses, only those variables unchanging over time have been included. It was found that the results of randomizations conducted in 'developing' countries were less likely to have ever been reported and that these trials, and smaller trials, were less likely to have ever been mentioned in an article. These characteristics, together with those of multicentre participation and recent occurrence, are also associated with both a lower frequency of mentions, and also a lower frequency of reporting of results.

17.4 Overall conclusion

The overall conclusion, therefore, is that there is evidence of 'pipeline bias' in the reporting of results in that highly statistically significant results are published faster than others, but this is not a problem for those wishing only to identify randomized trials for inclusion in meta-analyses. The geographic location of randomized trials was, however, found to be an important predictor of whether a randomization was ever mentioned in an article, frequency of mentions and of the time to publication. It may be worth going to extra lengths to track down the smaller trials conducted in countries outside North America and Europe in order to avoid selection bias when identifying randomized trials for inclusion in meta-analyses.

REFERENCES

I Articles used as a data source

Note: The publication ID used in this project is given after the reference.

Abromowitch M, Ochs J, Pui C-H, Fairclough D, Murphy SB, Rivera GK, SJCRH (1988a) Efficacy of High-Dose Methotrexate in Childhood Acute Lymphoblastic Leukemia: Analysis by Contemporary Risk Classification *Blood* **71 no 4 (Apr)** 866-869 225

Abromowitch M, Ochs J, Pui C-H, Kalwinsky D, Rivera GK, Fairclough D, Look T, Hustu HO, Murphy SB, Evans WE, Dahl GV, Bowman WP, SJCRH (1988b) High-Dose Methotrexate Improves Clinical Outcome in Children with Acute Lymphoblastic Leukemia: St Jude Total Therapy Study X *Medical and Pediatric Oncology* **16 (5)** 297-303 224

Aur RJA, Simone JV, Hustu HO, Verzosa MS, SJCRH (1972) A Comparative Study of Central Nervous System Irradiation and Intensive Chemotherapy early in Remission of Childhood Acute Lymphoblastic Leukemia *Cancer* **29** 381-391 213

Aur RJA, Hustu HO, Verzosa MS, Wood A, Simone JV (1973a) Comparison of Two Methods of Preventing Central Nervous System Leukemia *Blood* **42 no 3 (Sep)** 349-357 216

Aur R, Verzosa M, Hustu O, Simone J, SJCRH (1973b) Single or multiple agent chemotherapy during remission of childhood lymphocytic leukemia *Proceedings of the Annual Meeting of the American Association Cancer Research* **14** 19 a74 217

Aur RJA, Simone JV, Hustu HO, Verzosa MS, Pinkel D, SJCRH (1974) Cessation of therapy during complete remission of childhood acute lymphocytic leukemia *The New England Journal of Medicine* **291** 1230-1234 239

Aur RJA, Simone JV, Verzosa MS, Hustu HO, Barker LF, Pinkel DP, Rivera G, Dahl GV, Wood A, Stagner S, Mason C (1978) Childhood Acute Lymphocytic Leukemia. Study VIII. *Cancer* **42** 2123-2134 218

Benoit Y, Boilletot A, Francotte N, Hoyoux C, Margueritte G, Philippe N, Souillet G, Thyss A, Solbu G, Suciu S, Otten J (Academisch Ziekenhuis, GENT) (1988) Treatment of Medium and High-Risk (M-HR) Acute Lymphoblastic Leukemia (ALL) with or without Cranial Radiotherapy (RXT) Results of a Randomized Trial (EORTC nr 58832) *Medical and Pediatric Oncology* **16** 388 142

Berry DH, Pullen J, George S, Vietti TJ, Sullivan MP, Fernbach D (1975) Comparison of Prednisolone, Vincristine, Methotrexate, and 6-Mercaptopurine vs Vincristine and Prednisolone Induction Therapy in Childhood Acute Leukemia *Cancer* **36** 98-102 245

Berry DH, Fernbach DJ, Herson J, Pullen J, Sullivan MP, Vietti TJ (1980) Comparison of Prednisolone, Vincristine, Methotrexate and 6-Mercaptopurine vs 6-mercaptopurine and Prednisolone Maintenance Therapy in Childhood Acute Leukemia: A Southwest Oncology Group Study *Cancer* **46** 1098-1103 246

Bleyer WA, Coccia PF, Sather HN, Level C, Lukens J, Niebrugge DJ, Siegel S, Littman PS, Leikin SL, Miller DR, Chard Jr RL, Hammond GD and the Childrens Cancer Study Group (1983a) Reduction in Central Nervous System Leukaemia with a Pharmacokinetically Derived Intrathecal Methotrxate Dosage Regimen *Journal of Clinical Oncology* **1 no 5 (May)** 317-325 97

Bleyer WA, CCSG (1984) Intrathecal Methotrexate versus Central Nervous System Leukemia *Cancer Drug Delivery* **1 no 2** 157-167 110

Bleyer WA (1987) Some Clinical Features of Childhood Leukemia of Special Significance for Molecular Biology *Recent Advances in Leukemia and Lymphoma* Gale RP, Golde DW (eds.) pp 465-479 Alan R Liss Inc (Wiley Liss): New York 102

Bleyer WA (1989) Remaining Problems in the Staging and Treatment of Childhood Lymphoblastic Leukemia *American Journal of Pediatric Hematology/Oncology* **11 (4)** 371-379 90

Bleyer WA (1990) Central Nervous System Leukemia *William Dameshek & Frederick Gunz's Leukemia 4th edition* Gunz FW, Henderson ES (eds.) Chapter 35 PP 865-911 Grune and Stratton (subsidiary of Harcourt Brace Jovanovich Publishers): New York 57

Bleyer WA, Fallacollita J, Robison L, Balsom W, Meadows A, Heyn R, Sitarz A, Oretega J, Miller D, Constine L, Nesbit M, Sather H, Hammond D (1990) Influence of Age, Sex and Concurrent Intrathecal Methotrexate Therapy on Intellectual Function after Cranial Irradiation During Childhood: A report from the CCSG *Pediatric Hematology and Oncology* **7 (4)** 329-338 114

Bleyer WA, Sather HN, Nickerson HJ, Coccia PF, Finklestein JZ, Miller DR, Littman PS, Lukens JN, Siegel SE, Hammond GD (1991) Monthly Pulses of Vincristine and Prednisolone Prevent Bone Marrow and Testicular relapse in Low-Risk Childhood Acute Lymphoblastic Leukemia: A Report of the CCG-161 Study by the Childrens Cancer Study Group *Journal of Clinical Oncology* **9 no 6 (June)** 1012-1021 100

Boguslawska-Jaworska J, Raus Z, (1981) Chemoimmunotherapy with Levamisole During Maintenance Therapy in Children with Acute Lymphoblastic Leukemia *Archivum Immunologiae et Therapiae Experimentalis* **29** 719-723 253

Borella L, SJCRH (1973) Irradiation of CNS in Leukemia *British Medical Journal* **3** 456 212

Borowitz MJ, Shuster JJ, Civin CI, Carroll AJ, Look T, Behm FG, Land VJ, Pullen J, Crist WM (1990) Prognostic Significance of CD34 Expression in Childhood B-Precursor Acute Lymphocytic Leukemia: A Pediatric Oncology Group Study *Journal of Clinical Oncology* **8 no 8 (August)** 1389-1398 191

Bowman WP, Ochs J, Pui C-H, Kalwinsky DK, Abromowitch M, Aur RJA, Rivera G, Simone JV, SJCRH (1984) New Directions of St Jude Total-Therapy Protocols: Study X for Standard-Risk Childhood Acute Lymphoblastic Leukemia *Leukemia Research: Advances in Cell Biology and Treatment* Murphy SB, Gilbert JR (eds.) pp 203-211 Elsevier Science Publishing Co. Inc.: Oxford 220

Brandalise S, Odone V, Pereira W, Andrea M, Zanichelli M, Aranga V, ALL Brazilian Group, State University of Campinas, Sao Paulo, Brazil (1993) Treatment results of three consecutive Brazilian co-operative childhood ALL protocols: GBTLI-80, GBTLI-82 and -85 *Leukemia* **7 suppl 2** S142-5 270

Brecher ML, Berger P, Freeman AI, Krischer J, Boyett J, Glicksman AS, Foreman E, Harris M, Jones B, Cohen ME, Duffner PK, Rowland JH, Huang Y-P, Batnitzky S (1985) Computerised Tomography Scan Findings in Children With Acute Lympohocytic Leukemia Treated With Three Different Methods of Central Nervous System Prophylaxis *Cancer* **56** 2430-2433 60

Brouwers P, Moss H, Reaman G, McGuire T, Trupin E, Libow J, Tarnowski K, Bleyer W, Feusner J, Ruymann F, Miser J, Hammond D, Poplack D, NCI and CCSG (1988) Central Nervous System Preventative Therapy with Systemic High Dose Methotrxate versus Cranial Radiation and Intrathecal Methotrexate: Comparison of Effects of Treatment on Academic Achievement in Children with Acute Lymphoblastic Leukemia *Proceedings of the American Society of Clinical Oncology (ASCO)* **7** 176 a*678 171

Buchanan GR, Boyett JM, Rivera G (1988) Intensive Continuation Therapy for Patients with Acute Lymphoblastic Leukemia (ALL) in Second Marrow Remission: A Pediatric Oncology Group (POG) Study *Proceedings of American Society of Clinical Oncology (ASCO)* **7 (Mar 1988)** 188 a727 197

Buhrer C, Henze G, Hofmann J, Reiter A, Schellong G, Riehm H (1990) Central Nervous System Relapse Prevention in 1165 Standard-Risk Children with Acute Lymphoblastic Leukemia in Five BFM Trials *Haematology and Blood Transfusion* Buchner, Schellong, Hiddemann, Ritter (eds.) **33** pp 500-503 Springer-Verlag: Berlin, Heidelberg 47

Campbell AC, Hersey P, MacLennan ICM, Kay HEM, Pike MC and theMedical Research Council's Working Party on Leukaemia in Childhood. (1973) Immunosuppressive Consequences of Radiotherapy and Chemotherapy in Patients with Acute Lymphoblastic Leukaemia *British Medical Journal* **2** 385-388 4

Carden PA, Mitchell SL, Waters KD, Tiedemann K, Ekert H from the Oncology Pharmacy & Dept of Clinical Haematology & Oncology, Royal Children's Hosp, Parkville, Australia (1990) Prevention of Cyclophosphamide/Cytarabine-Induced Emesis with Ondansetron in Children with Leukemia *Journal of Clinical Oncology* **8 no 9 (Sep)** 1531-1535 269

Cherlow J, Steinherz P, Sather H, Gaynon P, Grossman N, Kersey J, Maurer H, Brenneman J, Trigg M, Hammond D, Childrens Cancer Study Group (1990) The Role of Radiation Therapy in the Management of Acute Lymphoblastic Leukemia with Lymphomatous Presentation (ALL/LP) *International Journal of Radiation Oncolog, Biology Physics* **19 supp 1** 176 (a100) 92

Cherlow J, Steinherz PG, Sather HN, Gaynon PS, Grossman NJ, Kersey JH, Johnstone HS, Breneman JC, Trigg ME, Hammond GD (1993) The Role of Radiation Therapy in the Treatment of Acute Lymphoblastic Leukemia with Lymphomatous Presentation: A Report from the Childrens Cancer Group *International Journal of Radiation Oncolog,y Biology Physics* **27 no 5** 1001-1009 93

Chessells JM, Durrant J, Hardy RM, Richards S (1986) Medical Research Council Leukaemia Trial - UKALL V: An Attempt to Reduce the Immunosupressive Effects of Therapy in Childhood Acute Lymphoblastic Leukaemia. Report to the Council by the Working Party on Leukaemia in Childhood. *Journal of Clinical Oncology* **4** 1758-1764 9

Chessells JM, Leiper AD, Tiedemann K, Hardisty RM, Richards S (1987) Oral methotrexate is as effective as intramuscular in maintenance therapy of acute lymphoblastic leukaemia *Archives of Disease in Childhood* **62** 172-176 152

Chessells JM, Cox TCS, Kendall B, Cavanagh NPC, Jannoun L, Richards S (1990) Neurotoxicity in lymphoblastic leukaemia: comparison of oral and intramuscular methotrexate and two doses of radiation *Archives of Disease in Childhood* **65** 416-422 13

Chessells JM, Bailey CC, Richards S (1991) MRC UKALL X: The UK National Protocol for Acute Lymphoblastic Leukaemia (ALL) 1985-90 *Haematologica* **76 supp 4** 104 (N407) 11

Chessells JM, Bailey CC, Richards S on behalf of the Medical Research Council Working Party on Childhood Leukaemia (1992a) MRC UKALL X. The UK protocol for childhood ALL: 1985-1990 *Leukemia* **6 Supp 2** 157-161 10

Chessells JM, Bailey C, Wheeler K, Richards SM (1992b) Bone marrow transplantation for high-risk childhood lymphoblastic leukaemia in first remission: experience in MRC UKALL X. Report from the Medical Research Council Working Party on Childhood Leukaemia. *Lancet* **340** 565-68 23

Chessells JM, Bailey C, Richards SM for the Medical Research Council Working Party on Childhood Leukaemia (1995) Intensification of treatment and survival in all children with lymphoblastic leukaemia: results of UK Medical Research Council trial UKALL X *Lancet* **345** 143-48 24

Children's Cancer and Leukemia Study Group (1989) Treatment of Childhood Acute Lymphoblastic Leukemia: Randomized Trials of Protocols CCLSG-L841 and I841 (Phase III Study) *Japanese Journal of Clinical Haematology* **30** 967-974 164

Clayton PE, Morris-Jones PH, Shalet SM, Price DA (1988) Growth in Children treated for Acute Lymphobastic Leukaemia *Lancet* **1** 460-2 20

Coccia PF, Bleyer WA, Siegel SE, Lukens JN, Gross S, Miller DR, Littman PF, Sather HN. Hammond GD, Children's Cancer Study Group (1984) Development and Preliminary Findings of Children's Cancer Study Group Protocols (161,162 and 163) for Low-, Average- and High-Risk Acute Lymphoblastic Leukemia in Children *Leukemia Research: Advances in Cell Biology and Treatment* Murphy SB, Gilbert JR (eds.) pp 241-250 Elsevier Science Publishing Co. Inc.: Oxford 101

Coccia PF, Bleyer WA, Lukens JN, Siegel SE, Sather H, Hammond D, CCRG (1989) Early response to induction therapy predicts early and late relapse in childhood acute lymphoblastic leukemia (ALL) *Blood* **74 (7) suppl** 227a 103

Crist W, Boyett J, Frankel L, Iyer R, van Eys J, Sullivan M, Harris M, Ravindranath Y, Pullen J (1988a) Extended triple intrathecal chemotherapy (TIC) is highly effective and superior to brief TIC and intermediate-dose MTX (IDM) as CNS prophylaxis in non-T, non-B ALL. A Pediatric Oncology Group (POG) study *Medical and Pediatric Oncology: InternationalSociety of Pediatric Oncology (SIOP) abstract* **16** 387 184

Crist W, Boyett J, Jackson J, Vietti T, Borowitz M, Chauvenet A, Winick N, Ragab A, Mahoney D, Head D, Iyer R, Wagner H, Pullen J (1989) Prognostic Importance of the Pre-B-Cell Immunophenotype and Other Presenting Features in B-Lineage Childhood Acute Lymphoblastic Leukemia: A Pediatric Oncology Group Study *Blood* **74 no 4 (Sept)** 1252-1259 186

Crist WM, Carroll AJ, Shuster JJ, Behm FG, Whitehead M, Vietti TJ, Look AT, Mahoney D, Ragab A, Pullen DJ, Land VJ (1990a) Poor Prognosis of Children with Pre-B Acute Lymphoblastic Leukemia is associated with the t(1;19)(q23;p13): A Pediatric Oncology Group Study *Blood* **76 no 1** 117-122 193

Crist W, Carroll A, Shuster J, Jackson J, Head D, Borowitz M, Behm F, Link M, Steuber P, Ragab A, Hirt A, Brock B, Land V, Pullen J, St Jude Children's Research Hospital (1990b) Philadelphia Chromosome Positive Childhood Acute Lymphoblastic Leukemia: Clinical and Cytogenetic Characteristics and Treatment Outcome. A Pediatric Oncology Group Study *Blood* **76 no 3 (Aug 1)** 489-494 204

Crist W, Shuster J, Steuber P, Pullen J, Wagner H, Laver J, Borowitz M, Look A, Link M, Frankel L, Amylon M, Land V, POG (1991) Results of Recent Trials of the Pediatric Oncology Group (POG) for Children with Acute Lymphoblastic Leukemia (ALL) *Haematologica* **76 supp no 4** 104 a N.408 189

Crist W, Shuster J, Look T, Borowitz M, Behm F, Bowman P, Frankel L, Pullen J, Krance R, Steuber P, Camitta B, Amylan M, Link M, Land V for the Pediatric Oncology Group (1992) Current results of studies of immunophenotype-, age- and leukocyte - based therapy for children with acute lymphoblastic leukemia *Leukemia* **6 supp 2** 162-165 192

Dolan G, Lilleyman JS, Richards SM on behalf of Leukaemia in Childhood Working Party of the Medical Research Council (1989) Prognostic importance of myelosuppression during maintenance treatment of lymphoblastic leukaemia *Archives of Disease in Childhood* **64** 1231-1234 285

Dritschilo A, Cassady JR, Camitta B, Jaffe N, Furman L, Traggis D (1976) The Role of Irradiation in Central Nervous System Treatment and Prophylaxis for Acute Lymphoblastic Leukemia *Cancer* **37** 2729-2735 125

Duttera MJ, Bleyer WA, Pomeroy TC, Leventhal CM, Leventhal BG, Pediatric Oncology Branch and Radiation Branch, National Cancer Institute (1973) Irradiation, Methotrexate Toxicity, and the Treatment of Meningeal Leukemia *Lancet* **2** 703-707 203

Eden OB, Lilleyman J, Shaw MP, Richards S, Peto J (1987) Medical Research Council Childhood Leukaemia Trial VIII Compared with Trials II -VII: Lessons for Future Management *Haematology and Blood Transfusion* Buchner, Schellong, Hiddermann, Urbanitz, Ritter (eds.) **30** pp 448-455 Springer-Verlag: Berlin, Heidelberg 19

Eden OB, Lilleyman JS, Richards S on behalf of the Medical Research Council Working Party on Leukaemia in Childhood (1990) Testicular irradiation in childhood lymphoblastic leukaemia *British Journal of Haematology* **75** 496-498 17

Eden OB, Lilleyman JS, Richards S, Peto J (1991) Results of Medical Research Council Childhood Leukaemia Trial UKALL VIII (Report to the Medical Research Council on behalf of the Working Party on Leukaemia in Childhood) *British Journal of Haematology* **78** 187-196 21

Eden OB, Shaw MP, Lilleyman JS, Richards S (1992) Children under two years treated according to the Medical Research Council UKALL VIII study and trial 1980-1984 (on behalf of the Medical Research Council Working Party on Leukaemia in Childhood). *British Journal of Cancer* **66 Suppl. XVIII** 558-562 22

Falletta JM, Shuster JJ, Crist WM, Pullen DJ, Borowitz MJ, Wharam M, Patterson R, Foreman E, Vietti TJ (1992) Different Patterns of Relapse Associated with Three Intensive Treatment Regimens for Pediatric E-Rosette Positive T-Cell Leukemia: a Pediatric Oncology Group Study *Leukemia* **6 no 6** 541-546 183

Fernbach DJ, George SL, Sutow WW, Ragab AH, Lane DM, Haggard ME, Lonsdale D, SWOG (1975) Long-Term Results of Reinforcement Therapy in Children with Acute Leukemia *Cancer* **36** 1552-1559 250

Freeman AI, Weinberg VE, Brecher ML, Jones B, Glicksman AS for CALGB (1981) Comparison of internmediate Dose Methotrexate (IDM) with Cranial Radiation (CRT) in Children with Acute Lymphocytic Leukaemia (ALL) *Proceedings of the American Association for Cancer Research (AACR) and the American Society of Clinical Oncology (ASCO)* **22** 486 C-599 280

Freeman AI, Pleuss H, Hananian J, Burgert EO, Gilchrist GS, Necheles T, Harris M, Kung F, Patterson RB, Maurer H, Leventhal B, Chevalier L, Forman E, Holland JF (1983) Comparison of Intermediate-Dose Methotrexate with Cranial Irradiation for the Post-Induction Treatment of Acute Lymphocytic Leukaemia in Children *New England Journal of Medicine* **30** 477-484 63

Freeman A, Boyett J, Glicksman A, Holland JF (1992) Comparison of Intermediate dose Methotrexate (IDM) with Cranial Irradiation (CRT) for Children with Acute Lymphocytic Leukemia (ALL). A long-term follow-up for CALGB. *28th Annual Meeting of the American Society of Clinical Oncology (ASCO). May 17-19,1992 Program/Procedings.* **11** 285 (abstract 951) 72

Frei E III, Karon M, Levin RH, Freireich EJ, Taylor RJ, Hananian J, Selawry O, Holland JF, Hoogstraten B, Wolman IJ, Abir E, Sawitsky A, Lee S, Mills SD, Burgert EO Jr, Spur CL, Patterson RB, Ebaugh FG, James GW III, Moon JH (1965) The Effectiveness of Combinations of Antileukemic Agents in Inducing and Maintaining Remission in Children with Acute Leukemia *Blood* **26 no 5 (November)** 642-656 50

Gaynon PS, Steinherz PG, Reaman GH, Bleyer WA, Sather H, Hammond GD (1987) Strategies for the Treatment of Children with Acute Lymphoblastic Leukemia and Unfavourable Presenting Features *Haematology and Blood Transfusion* Buchner, Schellong, Hiddemann, Urbanitz, Ritter (eds.) **30** 167-172 Springer-Verlag: Berlin, Heidelberg 89

Gaynon PS, Bleyer WA, Albo VC, Grossman NJ, Novak LT, Reaman GH, Steinherz PG, Ablin AR, Finklestein JZ, Littman PS, Pyesmany AF, Sather HN, Hammond GD (1988) Intensive Therapy for Children with Acute Lymphoblastic Leukaemia and Unfavourable Presenting Features. Early Conclusions of Study CCG-106 by the Childrens Cancer Study Group *Lancet* **ii** 921-924 87

Gaynon P, Steinherz P, Bleyer A, Ablin A, Albo V. Finklestein J, Grossman N, Novak L, Reaman G, Pyesmony A, Sather H, Hammond GD, Childrens Cancer Study Group (CCG) (1989) Event Free Survival (EFS) and Morbidity of Three Regimens (R) for Children (ch) with Acute Lymphoblastic Leukemia (ALL) and Infavourable Presenting Features (UPF) *Proceedings of the American Society of Clinical Oncology (ASCO)* **8** 212 a823 88

Gaynon PS, Steinherz PG, Bleyer WA, Ablin AR, Albo VC, Finklestein JZ, Grossman NJ, Novak LJ, Pyesmony AF, Reaman GH, Chappell RJ, Sather HN, Hammond GD (1993) Improved Therapy for Children with Acute Lymphoblastic Leukemia and Unfavorable Presenting Features: A Follow-up Report of the Childrens Cancer Group Study CCG-106 *Jouralof Clinical Oncology* **11 no 11 (Nov)** 2234-2242 86

Gilbert SJ, Shuster JJ, Land VJ, Wharam MD, Thomas PR, Nitschke R, Pinkel D, Vietti TJ (1991) Remission Induction and Continuation Therapy in Children with their first relapse of acute lymphoid leukemia. A Pediatric Oncology Group Study *Cancer* **67(1)** 37-42 a34 205

Gilchrist G, Tubergen D, Cocchia P, Novak L, O'Brien R, Waskerwitz M, Sather H, Bleyer A, Hammond D for Childrens Cancer Study Group (CCSG) Pasadena, California 91101 (1990) Cerebrospinal fluid blasts (CSFB) on cytocentrifuge (CC) do not predict for Central Nervous System Leukemia (CNSL) in children with Intermediate Risk Acute Lymphoblastic Leukemia (ALL) *26th Annual Meeting of the American Society of Clinical Oncology (ASCO) May 20-22, 1990. Program/Proceedings, Washington DC* **9** a 840 79

Giralt J, Ortega JJ, Olive T, Verges R, Forio I, Salvader L, (Hospital General Vall d'Hebron, Univ of Barcelona, Spain) (1992) Long-Term Neuropsychologic Sequelae of Childhood Leukemia: Comparison of

Two CNS Prophylactic Regimens *International Journal of Radiation Oncology Biology Physics* **45 no 1** 49-53 259

Glidewell OJ, Holland JF (1973) Clinical Trials of the Acute Leukemia Group B in Acute Lymphoblastic Leukaemia of Childhood *Unifying Concepts of Leukaemia. Bibl.haemat.* Dutcher RM, Chieco-Bianchi L (eds) **no 39** pp 1053-1067 Karger: Basel 51

Graham-Pole J on behalf of Pediatric Oncology Group (1989) Treating Acute Lymphoblastic Leukemi after Relapse: Bone Marrow Transplantation or not? *Lancet* **2** 1517-1518 188

Halazun JF, Wagner HR, Gaeta JF, Sinks LF, Acute Leukaemia Group B, Roswell Park Memorial Institute & Children's Hospital of Buffalo, Buffalo, New York (1974) Daunorubicin Cardiac Toxicity in Children with Acute Lymphocytic Leukaemia *Cancer* **33** 545-554 282

Hamre MR, Robison LL, Nesbit ME, Sather HN, Meadows AT, Ortega JA, d'Angio GJ, Hammond GD (1987) Effects of Radiatio on Ovarian Function in Long-Term Survivors of Childhood Acute Lymphoblastic Leukaemia: A Report from the Childrens Cancer Study Group *Journal of Clinical Oncology* **5 no 11 (Nov)** 1759-1765 112

Henderson ES, Hoelzer D, and Freeman AI (1990) Chapter 20: The treatment of Acute Lymphoblastic Leukemia *Leukemia* (5th edition) Henderson ES, Lister TA (eds.) pp 443-484 WB Saunders: Philadelphia 105

Henze G, Langermann H-J, Bramswig J, Breu H, Gadner H, Schellong G, Welte K, Riehm H (1981) Ergebnisse der Studie BFM 76/79 zur Behandlung der akuten lymphoblastischen Leukamie bei Kindern und Jugendlichen *Klinische Padiatrie* **193** 145-154 33

Henze G, Langhermann H-J, Fengler R, Brandeis M, Evers KG, Gadner H, Hinderfeld L, Jobke A, Kornhuber B, Lampert F, Lasson U, Ludwig R, Muller-Weihrich S,Neidhardt M, Nessler G, Niethammer, Rister M, Ritter J, Schaaff A, Schellong G, Stollmann B, Treuner J, Wahlen W, Weiner P, Wehinger H, Riehm H (1982) Acute Lymphoblastic Leukemia Therapy Study BFM 79/81 in Children and Adolescents: intensified Reinduction Therapy for Patients with Different Risk for Relapse (Therapie Studie BFM 79/81 zur Behandlung der akuten ly,phoblastischen Leukamie bei Kindern und Jugenlichen: Intensivierte Reinduktionstherapie fur Patientengruppen mit Unterschiedlichern Rezidivrisiko) *Klinische Padiarie* **194** 195-203 34

Henze G, Buchmann S, Fengler R, Hartmann R (1987) The BFM relapse studies in Childhood ALL: Concepts of Two Multicenter trials and results after 2 1/2 years. *Haematology and Blood Transfusion* Buchner, Schellong, Hiddeman, Urbanitz, Ritter (eds.) **30:Acute Leukaemias** pp 147-155 Springer-Verlag: Berlin, Heidelberg 44

Henze G, Fengler R, Hartmann R, Niethammer D, Schellong G, Riehm H (1990a) BFM Group Treatment Results in Relapsed Childhood Acute Lymphoblastic Leukaemia *Haematology and Blood Transfusion* Buchner, Schellong, Hiddeman, Ritter (eds.)**33** pp 619-626 Springer-Verlag: Berlin, Heidelberg 43

Henze G, Fengler R, Reiter A, Ritter J, Riehm H (1990b) Impact of Early Intensive Reinduction Therapy on Event-Free Survival in Children with Low-Risk Acute Lymphoblastic Leukaemia *Haematology and Blood Transfusion* Buchner, Schellong, Hiddeman, Ritter (eds.) **33** pp 483-488 Springer-Verlag: Berlin, Heidelberg 41

Henze G, Fengler R, Hartmann R, Kornhuber B, Janka-Schaub G, Niethammer D, Riehm H (1991) Six year experience with a comprehensive approach to the treatment of recurrent childhood acute lymphoblastic leukaemia (ALL-REZ BFM 85) A relapse study of the BFM group. *Blood* **78 (5)** 1166-72 40

Heyn RM, Joo P, Karon M, Nesbit M, Shore N, Breslow N, Weiner J, Reed A, Hammond D (1975) BCG in the Treatment of Acute Lymphoblastic Leukemia *Blood* **46 no 3 (Sept)** 431-442 106

Hill JM, Kornblith AB, Boyett JM, Freeman A, Holland JF, Glicksman AS, Lenherr B, Mao L, Bricker M, Dubowy R et al (1994) Long-term function of childhood acute lymphocytic leukaemia (ALL) survivors treated by CNS prophylaxis *Proceedings of Annual Meeting of the American Society of Clinical Oncology (AACR)* **13** 23 64

Hitchcock-Bryan S, Gelber R, Cassady JR, Sallan SE (1986) The Impact of Induction Anthracycline on Long-Term Failure-Free Survival in Childhood Acute Lymphoblastic Leukemia *Medical and Pediatric Oncology* **14** 211-215 122

Holland JF, Glidewell O for Acute Leukaemia Group B (1972a) Chemotherapy of Acute Lymphocytic Leukaemia of Childhood *Cancer* **30** 1480-1487 52

Holland JF, Glidewell O (1972b) Oncologists' Reply: Survival Expectancy in Acute Lymphocytic Leukemia *New England Journal of Medicine* **287 no 15** 769-777 55

Holland JF (1978) Acute Lymphocytic Leukemia and Lymphomas: Status of Chemotherapy *Immunotherapy of Cancer: Present Status of Trials in Man* Terry WD, Windhorst D (eds.) pp 441-450 Raven Press: New York 53

Hughes WT, Kuhn S, Chaudhary S, Feldman S, Verzosa M, Aur RJA, Pratt C, George SL, SJCRH (1977) Successful Chemoprophylaxis for Pneumocystis Carinii Pneumonitis *New England Journal of Medicine* **297 no26** 1419-1426 241

Hughes WT, Rivera GK, Schell MJ, Thornton D, Lott L, SJCRH (1987) Successful Intermittent Chemoprophylaxis for Pneumocystis Carinii Pneumonitis *New England Journal of Medicine* **316 no 26** 1627-32 237

Hustu HO, Aur RJA, Verzosa MS, Simone JV, Pinkel D (1973) Prevention of Central Nervous System Leukemia by Irradiation *Cancer* **32 (3)** 585-597 214

Hvizdala E, Berry DH, Chen T, Dyment PG, Kim TH, Steuber CP, Sullivan MP (1984) Impact of the Timing of Triple Intrathecal Therapy on Remission Induction in Childhood Acute Lymphoblastic Leukemia: A Pediatric Oncology Group Study. *Medical and Pediatric Oncology* **12** 173-177 176

Janka-Schaub GE, Winkler K, Jurgens H, Goebel U, Gutjahr P, Spaar H-J (1986) Intermediate-dose methotrexate in the treatment of childhood acute lymphocytic leukaemia: lack of benefit during maintenance therapy following intensive induction therapy. *European Journal of Pediatrics* **145** 14-17 154

Janka GE, Winkler K, Juergens H, Goebel U for the COALL Study Groups. (1987) Early Intensification Therapy in High-Risk Childhood Acute Lymphoblastic Leukemia: Lack of Benefit from High-Dose Methotrexate *Haematology and Blood Transfusion* Buchner, Schellong, Hiddemann, Urbanitz, Ritter (eds.) **30** pp 456-460 Springer-Verlag: Berlin, Heidelberg 155

Janka-Schaub GE, Winkler K, Gobel U, Graubner U, Gutjahr P, Hass RJ, Jurgens H, Spaar J, for COALL Study Group (1988) Rapidly Rotating Combination Chemotherapy in Childhood Acute Lymphoblastic Leukemia: Preliminary Results of a Randomized Comparison with Conventional Treatment *Leukemia* **2 no 12** 735-785 158

Janka-Schaub GE, Goebel U, Graubner U, Hass RJ, Juergens H, Spaar HJ, Winkler K, for the COALL Study Group (1990) Improved Prognosis for Childhood Acute Lymphocytic Leukemia with Very High White Blood Cell Count (>100/nl) with Rotation of Non-Cross-Resistant Drug Combinatons *Haematology and Blood Transfusion* Buchner, Schellong, Hiddemann, Ritter (eds.) **33** pp 489-493 Springer-Verlag: Berlin, Heidelberg 159

Janka GE, Goebel U, Graubner U, Juergens H, Haas R, Winkler U for the COALL Study Group (1991) Randomized Comparison of Rotational Chemotherapy (CT) in high risk (HR) Childhood ALL *ECCO6 (Sixth European Conference on Clinical Oncology and Cancer Nursing)/ European Journal of Cancer* **27 supp 2** s247 a1512 156

Janka-Schaub GE, Harms D, Goebel U, Graubner U, Gutjahr P, Haas RJ, Juergens H, Spaar HJ, Winkler K for the COALL Study Group (1996) Randomized comparison of rotational chemotherapy in high-risk acute lymphoblastic leukaemia of childhood - follow-up after 9 years *European Journal of Pediatrics* **155** 640-648 157

Jones B, Hilland JF, Morrison AR, Lee SL, Sinks LF, Cuttner J, Rausen A, Kung F, Pluss HJ, Haerani FI, Patterson LB, Blom J, Burgert Jr EO, Moon JH, Chevalier L, Sawitsky A, Albala MM, Forcier J, Falkson G, Glidewell O (Acute Leukemia Group B) (1971) Daunorubicin (NSC 82151) in the Treatment of Advanced Childhood Lymphoblastic Leukaemia *Cancer Research* **31** 84-90 288

Jones B, Cuttner J, Levy RN, Patterson RB, Kung F, Pleuss HJ, Falkson G, Treat CL, Haurani F, Burgert EO Jr, Rosner F, Carey RW, Lukens J, Blom J, Degnan TJ, Wohl H, Glidewell O, Holland JF (Acute Leukaemia Co-operative Study Group B) (1972) Daunorubicin vs Dnr + Pred vs Dnr + Vinc + Pred in Advanced Childhood Lymphoblastic Leukemia *Cancer Chemotherapy Reports Part 1* **56 no 6 Dec 1972** 729-737 71

Jones B, Holland JF, Glidewell O, Jacquillat C, Weil M, Pochedly C, Sinks L, Chevalier L, Maurer HM, Koch K, Falkson G, Patterson R, Seligman B, Sartorius J, Kung F, Haurani F, Stuart M, Burgert EO, Ruymann F, Sawitsky A, Forman E, Pluess H, Truman J, Hakami N (1977) Optimal Use of L-Asparaginase (NSC-109229) in Acute Lymphocytic Leukaemia *Medical and Pediatric Oncology* **3** 387-400 56

Jones B, Freeman AI, Shuster JJ, Jacquillat C, Weil M, Pochedly C, Sinks L, Chevalier L, Maurer HM, Koch K, Falkson G, Patterson R, Seligman B, Sartorius J, Kung F, Haurani F, Stuart M, Burgert EO, Ruymann F, Sawitsky A, Forman K, Pluess H, Truman J, Hakami N, Glidewell O, Glicksman AS, Holland JF (members of Cancer and Leukemia Group B) (1991) Lower Incidence of Meningeal Leukaemia when Prednisolone is Replaced by Dexamethasone in the Treatment of Acute Lymphoblastic Leukemia *Medical and Pediatric Oncology* **19** 269-275 61

Koizumi S, Fujimoto T, Takeda T, Yatabe M, Ytsumi J, Mimaya J, Ninomiya T, Yanai M and the Japanese Children's Cancer and Leukemia Study Group (1988a) Comparison of Intermittent or Continuous Methotrexate Plus 6-Mercaptopurine in Regimens for Standard-Risk Acute Lymphoblastic Leukemia in Childhood (JCCLSG-S811) *Cancer* **61 no 7** 1292-1300 161

Koizumi S, Fujimoto T, The Japanese Children's Cancer and Leukemia Study Group (1988b) Comparison of Intermittent or Continuous Methotrexate plus 6-Mercaptopurine in Regimens for Standard-Risk Acute Lymphoblastic Leukemia in Childhood (CCLSG-S811) *Medical and Pediatric Oncology* **16** 396 163

Koizumi S, Fujimoto T, Sasaki K and the Children's Cancer and Leukemia Study Group, Japan (1989) Late intensification therapy with high dose methotrexate in standard-risk acute lymphoblastic leukemia in childhood *Proceedings of the American Association for Cancer Research (AACR)* **30** 266 a1060 162

Koizumi S, Fujimoto T, Sasaki U, Takeda T, Utsumi J, Mimaya J, Ohta S, Ninomiya T, Takaue Y and the Children's Cancer and Leukemia Study Group (1991) Retrospective analysis of late intensification therapy with high-dose methotrexate for standard-risk acute lymphoblastic leukemia in childhood (CCLSG-S811 study) *International Journal of Haematology* **54** 307-313 160

Koizumi S, Fujimoto T, Sasaki K, Kawai S, Takeda T, Japan (1992) Efficacy of Intermediate Methotrexate (MTX) and 6-Mercaptopurine (6MP) in Maintenance Chemo-therapy for Childhood Acute Lymphoblastic Leukemia (ALL): A report from the Japanese Children's Cancer and Leukemia Study Group (CCCSG) *28th Annual Meeting of the American Society of Clinical Oncology (ASCO) May 17-19 1992. Program/proceedings.* **11** 286 a952 166

Koizumi S, Fujimoto T (1994) Improvement in treatment of childhood acute lymphoblastic leukemia: a 10-year study by The Children's Cancer and Leukemia Study Group *International Journal of Haematology* **59** 99-112 83

Komp D, Falletta J, Ragab A, Humphrey GB, SWOG (1975) Is cranial radiation necessary for CNS prophylaxis in ALL of childhood? *Proceedings of American Society of Clinical Oncology (ASCO)* **16** a1046 243

Komp DM, George SL, Falletta J, Land VJ, Starling KA, Humphrey GB, Lowman J (1976) Cyclophosphamide-Asparaginase-Vincristine-Prednisolone Induction Therapy in Childhood Acute Lymphocytic and Non Lymphocytic Leukemia *Cancer* **37** 1243-1247 242

Komp DM, Fernandez CH, Falletta JM, Ragab AH, Humphrey GB, Pullen J, Moon T, Shuster J (1982) CNS Prophylaxis in Acute Lymphoblastic Leukemia. Comparison of Two Methods. A Southwest Oncology Group Study. *Cancer* **50** 1031-1036 244

Kretschmar C, Bernal S, Chen LB, Sallan SE, DFCI (1984) Measurements of leukemic cell kill in untreated childhood acute lymphoblastic leukemia (ALL) after single agent methotrexate (MTX) treatment *Proceedings of the American Association for Cancer Research (AACR)* **190** a752 124

Krischer J, Land VJ, Civin CI, Ragab AH, Mahoney DH, Frankel LS (1984) Evaluation of AMSA in children with Acute Leukemia. A Pediatric Oncology Group Study. *Cancer* **54** 207-210 200

Kung F, Nyhan WL, Cuttner J, Falkson G, Lanzkowsky P, Del Duca V, Nawabi IU, Koch K, Pluess H, Freeman A, Burgert EO, Leone LA, Ruymann F, Patterson RB, Degnan T, Hakami N, Pajak TF, Holland J (1978) Vincristine, Prednisolone and L-Asparaginase in the Induction of Remission in Children with Acute Lymphoblastic Leukaemia following Relapse *Cancer* **41** 428-434 59

Land VJ, Thomas PRM, Boyett JM, Glicksman AS, Culbert S, Castleberry RP, Berry DH, Vats T, Humphrey GB (1985) Comparison of Maintenance Treatment Regimens for First Central Nervous System Relapse in Children with Acute Lymphocytic Leukemia. A Pediatric Oncology Group Study *Cancer* **56** 81-87 198

Land VJ, Pullen DJ, Shuster JJ, Alvarado C, Amylon M, Harris MB (1989) Continuing Improvement of Outcome in Childhood non-T, non-B Acute Lymphoblastic Leukemia (NTNB-ALL): Pediatric Oncology Group (POG) experience in the 1980's. (poster session II) *Blood* **74 suppl 80a** a292 190

Land VJ, Shuster JJ, Crist WM, Ravindranath Y, Harris MB, Krance RA, Pinkel D, Pullen DJ (1994) Comparison of Two Schedules of Intermediate-Dose Methotrexate and Cytarabine Consolidation Therapy for Childhood B-Precursor Cell Acute Lymphoblastic Leukaemia: A Paediatric Onology Group Study *Journal of Clinical Oncology* **12 no 9 (Sept)** 1939-1945 279

Lange BJ, Blatt J, Sather HN, Meadows AT (1996) Randomized Comparison of Moderate-Dose Methotrexate Infusions to Oral Methotrxate in Children with Intermedite Risk Acute Lymphoblastic Leukemia: A Childrens Cancer Group Study *Medical and Pediatric Oncology* **27** 15-20 95

Leverger G, Democq F, Freycon F, Bancillon A, Lepage E, Bordigoni P, Bernaudin F, Schaison G, for the French ALL cooperative group (1988) Treatment of Childhood Low Risk Lymphoblastic Leukemias (ALL): Preliminary Results of the French ALL Study Group FRALLE83 *Medical and Pediatric Oncology* **16** 389 145

Lilleyman JS, Richards S, Rankin A (1985) Medical Research Council leukaemia trial, UKALL VII A report to the Council by the Working Party on Leukaemia in Childhood. *Archives of Disease in Childhood* **60** 1050-1054 14

Lilleyman JS, Hann IM, Stevens RF, Eden OB, Richards SM on behalf of the Medical Research Council's Working Party on Leukaemia in Childhood (1986) French American British (FAB) morphological classification of childhood lymphoblastic leukaemia and its clinical importance. *Journal of Clinical Pathology* **39** 998-1002 8

Littman P, Coccia P, Bleyer WA, Lukens J, Siegel S, Miller D, Sather H, Hammond D (1987) Central Nervous System (CNS) Prophylaxis in Children with Low Risk Acute Lymphoblastic Leukemia (ALL) *International Journal of Radiation Oncology Biology Physics* **13 (10)** 1443-1449 99

MacLean WE Jr, Noll RB, Stehbens JA, Kaleita TA, Schwartz E, Whitt JK, Hammond GD for the Childrens Cancer Group (1995) Neuropsychological Effects of Cranial Irradiation in Yound Children with Acute Lymphoblastic Leukemia 9 months after Diagnosis *Archives of Neurology* **52** 156-160 85

MacLennan ICM, Kay HEM, Festenstin M, Smith PG (1976) Analysis and Treatment in Childhood Leukaemia II. Timing and the Toxicity of Combined 6-Mercaptopurine and Methotrexate Maintenance Therapy. The Medical Research Council's Working Party on Leukaemia in Childhood. *British Journal of Haematology* **33** 179-188 12

Maekawa T, Miura T, Takau F, Fujioka S, Kawato M, Adachi Y, Miura Y, Mizoguchi H, Nomura T, Ohashi T, Onozawa Y, Shishido H, Toyama K (1987) Effect of alevamisole on the duration of First Remission and Survival in Adult Patients with Acute Leukaemia *Acta Therapeutica* **13** 425-437 153

Mahmoud HH, Rivera GK, Hancock ML, Krance RA, Kun LE, Behm FG, Ribeiro RC, Sandlund JT, Crist WM, Pui C-H, SJCRH (1993) Low leukocyte counts with blast cells in cerebrospinal fluid of children with newly diagnosed acute lymphoblastic leukemia *New England Journal of Medicine* **329 no 5** 314-319 236

Maschmeyer G, Daenen S, de Pauw BE, Vries-Hospers H de, Dekker AW, Donnelly JP, Gaus W, Haralambie E, Kern W, Konrad H, Link H, Sizoo W, van der Waaij D, von Eiff M, Wendt F (1990) Prevention of Infection in Acute Leukemia *Haematology and Blood Transfusion* Buchner, Schellong, Hiddemann, Ritter (eds.) **33** pp 525-530 Springer-Verlag: Berlin, Heidelberg 141

Mathe G, Amiel JL, Schearzenberg L, Schneider M, Cattan A, Schlumberger JR, Hayat M, de Vassal F, Institute of Cancerology and Immunogenetics, Hopital Paul-Brousse, & Dept of Haematology, Institute Gustave Roussy, 94-Villejuif, France (1969) (Preliminary Communications) Active Immunotherapy for Acute Lymphoblastic Leukaemia *The Lancet* **I** 697-699 263

Mathe G, Institut de Cancerologie et d'Immunogenetique, Hopital Paul-Brousse, 94 Villejuif, France. (1970) Immunological Treatment of Leukaemias *British Medical Journal* **4** 487-488 264

Mathe G, Amiel JL, Schearzenberg L, Schneider M, Cattan A, Schlumberger JR, Hayat M, de Vassal F, Institut de Cancerologie et d'Immunogenetique Unite Fred Siguier de l'Hopital Paul Brousse & Servie d'Hematologie de l'Institut Gustave-Roussy, Villejuif (1977a) Original Papers/Memoires: Follow-up of the first (1962) Pilot Study on Active Immunotherapy of Acute Lymphoid Leukaemia: A Critical Discussion *Biomedicine* **26** 29-35 265

Mathe G, de Vassal F, Schwarzenberg L, Delgado M, Pena-Angulo J, Belpomme D, Pouillart P, Machover D, Misret JL, Pico JL, Jasmin C, Hayat M, Schneider M, Cattan A, Amiel JL, Musset M, Rosenfeld C, Institut de Cancerologie et d'Immunogenetique (INSERM), Unite Fred-Siguier de l'Hopital Paul-Brousse et Service d'Hematologie de l'Institut Gustave-Roussy, 94800-Villejuif, France. (1977b) Results in Children of Acute Lymphoid Leukemia Protocol ICIG-ALL9 Consisting of Chemotherapy for Only Nine Months followed by Active Immunotherapy. Comparison with the results of more prolonged chemotherapy protocols. Recognition of two groups of acute lymphoid leukaemias from prognostic parameters *Adjuvant Therapy of Cancer* Salmon SE, Jones SE (eds.) pp 357-372 Elsevier/North-Holland Biomedical Press: Amsterdam 261

Mathe G, de Vassal F, Misset JL, Ribaud P, Hayat M, Schwarzenberg L, Rosenfeld C, Jasmin C, Machover D, Delgado M, Gil MA, Gouveia J, Pico JL, Belpomme D, Musset M, Gremy F, Institut de Cancerologie et d'Immunogenetique, Hopital Paul-Brousse, 94800-Villejuif & Departements de Biophysique et de Biomathematiques, CHU Pitie-Salpetriere, 75013-Paris, France. (1979) Comparison of Sequential Chemotherapy-Immunotherapy Protocols in Acute Lymphoid Leukemia: Correlation of the Results Accoding to the Length of Pre-Immunotherapy-Chemotherapy *Adjuvant Therapy of Cancer* Salmon SE, Jones SE (eds.) pp 191-198 Elsevier/North-Holland Biomedical Press: Amsterdam 262

Mauer AM, Simone JV, SJCRH (1976) The current status of the treatment of childhood acute lymphoblastic leukemia *Cancer Treatment Reviews* **3** 17-41 215

Meadows A, Lange B, Blatt J, Sather H, Hammond D, CCSG (1991) Moderate Dose Methotrexate (MTX) does not improve the outcome of Children with Intermediate Risk Acute Lymphoblastic Leukaemia (ALL): A Childrens Cancer Study Group (CCSG) Limited Institution Trial *Proceedings of the American Society of Clinical Oncology (ASCO)* **10** 239 94

Medical Research Council's Working Party on Leukaemia in Childhood (1971) Treatment of Acute Lymphoblastic Leukaemia. Comparison of Immunotherapy (BCG), Intermittent Methotrexate, and No Therapy after a Five-Month Intensive Cytotoxic Regimen (Concord Trial). Preliminary Report to the Medical Research Council by the Leukaemia Committee and the Working Party on Leukaemia in Childhood. *British Medical Journal* **4** 189-194 277

Medical Research Council's Working Party on Leukaemia in Childhood (1973) Treatment of Acute Lymphoblastic Leukaemia: Effect of 'Prophylactic' Therapy against Central Nervous Sysytem Leukaemia. Report to the Medical Research Council by the Leukaemia Committee and the Working Party on Leukaemia in Childhood *British Medical Journal* **2** 381-384 5

Medical Research Council's Working Party on Leukaemia in Childhood (report prepared by MacLennan ICM, Kay HEM, Festenstein M, Smith PG) (1975) Analysis of Treatment in Childhood Leukaemia. I-Predisposition to Methotrexate-induced Neutropenia after Craniospinal Irradiation: Report to the Medical Research Council of the Working Party on Leukaemia in Childhood *British Medical Journal* **3** 563-566 1

Medical Research Council's Working Party on Leukaemia in Childhood (report prepared by Hardisty RM, Kay HEM, Peto J) (1977) Treatment of acute lymphoblastic leukaemia: effect of variation in length of treatment on duration of remission. Report to the Medical Research Council by the Working Party on Leukaemia in Childhood *British Medical Journal* **2** 495-497 2

Medical Research Council's Working Party on Leukaemia in Childhood (report prepared by Hardisty RM, Kay HEM,Peto J) (1978) Effects of varying radiation schedule, cyclophosphamide treatment, and duration of treatment in acute lymphoblastic leukaemia. Report to the Medical Research Council by the Working Party on Leukaemia in Childhood. *British Medical Journal* **2** 787-791 3

Medical Research Council's Working Party on Leukaemia in Childhood (1982a) The Treatment of Acute Lymphoblastic Leukaemia (ALL) in Childhood, UKALL III: The Effects of Added Cytosine Arabinoside and/or Asparaginase, and a Comparison of Continuous or Discontinuous Mercaptopurine in Regimens for Standard Risk ALL *Medical and Pediatric Oncology* **10 (5)** 501-510 16

Medical Research Council's Working Party on Leukaemia in Childhood (1982b) Duration of Chemotherapy in Childhood Acute Lymphoblastic Leukaemia *Medical and Pediatric Oncology* **10 (5)** 511-520 15

Miller DR, Leikin S, Albo V, Sather H, Karon M, Hammond D (1983) Prognostic Factors and Therapy in Acute Lymphoblastic Leukemia of Childhood: CCG-141 *Cancer* **51** 1041-1049 108

Miller D, Leiken S, Albo V, Sather H, Hammond D, Childrens Cancer Study Group (CCSG) (1986) Duration of Therapy in Childhood Acute Lymphoblastic Leukemia *Proceedings of the American Society of Clinical Oncology (ASCO)* **5** 156 a608 98

Miller DR, Leikin SL, Albo VC, Sather H, Hammond GD (1989) Three versus Five Years of Maintenance Therapy Are Equivalent in Childhood Acute Lymphoblastic Leukemia: A report from the Childrens Cancer Study Group *Journal of Clinical Oncology* **7 no 3 (March)** 316-325 77

Miller DR, Leikin SL, Albo VC, Palmer NF, Sather HN, Hammond GD (1990) The Prognostic Value of Testicular Biopsy in Childhood Acute Lymphoblastic Leukemia: A report from the Childrens Cancer Study Group. *Journal of Clinical Oncology* **8 no 1 (January)** 57-66 96

Miser JS, Roloff J, Blatt J, Reaman GH, Krailo MD, Hammond GD (1992) Lack of significant activity of 2'-deoxycoformycin alone or in combination with adenine arabinoside in relapsed childhood acute lymphoblastic leukemia. A randomized phase II trial from the CCSG *American Journal of Clinical Oncology* **15 (6)** 490-3 120

Moss HA, Nannis ED, Poplack DG (1981) The Effects of Prophylactic Treatment of the Central Nervous System on the Intellectual Functioning of Children with Acute Lymphoblastic Leukemia *American Journal of Medicine* **71** 47-52 168

Mulhern RK, Fairclough D, Ochs J, SJCRH (1991) A Prospective Comparison of Neuropsychologic Performance of Children Surviving Leukemia Who received 18-Gy, 24-Gy, or No Cranial Irradiation *Journal of Clinical Oncology* **9 no 8 (Aug)** 1348-1356 227

Nakadate H, Hatae Y, Takeda T, Koizumi S-C, Fujimoto T, The Japanese Children's Cancer and Leukemia Study Group (JCCLSG) (1992) Treatment of Pediatric Acute Lymphoblastic Leukemia: The Results of JCCLSG IR-874 Study *Medical and Pediatric Oncology: International Society of Pediatric Oncology (SIOP) XXIV Meeting abstract* **20** 408 a P69 165

Nesbit M Jr, Gilchrist G, Beatty E Jr, Hartmann J, Krivit W for CCSG A (1973) Increased Survival or cure following cyclic-sequential therapy of acute lymphoblastic and undifferentiated leukemia (ALL/AUL) of Childhood *Proceedings of the American Association for Cancer Research (AACR), abstracts 1973* **46** a181 107

Nesbit M, Chard C, Evans A, Karon M, Hammond GD for CCSG (1979) Evaluation of intramuscular versus intravenous administration of L-asparaginase in childhood leukemia *American Journal of Pediatric Hematology/Oncology* **1 no 1 (Spring)** 9-13 104

Nesbit Jr ME, Robison LL, Ortega JA, Sather HN, Donaldson M, Hammond D (1980) Testicular Relapse in Childhood Acute Lymphoblastic Leukemia: Association with Pretreatment Patient Characteristics and Treatment. A Report for Childrens Cancer Study Group *Cancer* **45** 2009-2016 115

Nesbit Jr ME, Robison LL, Littman PS, Sather HN, Ortega J, d'Angio GJ, Hammond GD (1981a) Presymptomatic Central Nervous Sysytem Therapy in Previously Untreated Childhood Acute Lymphoblastic Leukaemia: Comparison of 1800 Rad and 2400 Rad. A Report for Children's Cancer Study Group *Lancet* **I** 461-466 113

Nesbit ME, Sather HN, Ortega J, d'Angio G, Robison LL, Donaldson M, Hammond GD (1981b) Effect of Isolated Central Nervous System Leukaemia on Bone Marrow Remission and Survival in Childhood Acute Lymphoblastic Leukaemia. A report for Children's Cancer Study Group. *Lancet* **1** 1386-89 119

Nesbit ME, Sather H, Robison LL, Donaldson M, Littman P, Ortega JA, Hammond GD (1982a) Sanctuary Therapy: A Randomized Trial of 724 Children with Previously Untreated Acute Lymphoblastic Leukemia. A Report from Childrens Cancer Study Group. *Cancer Research* **42 (2)** 674-680 116

Nesbit M, Sather H, Robison L, Ortega J, Donaldson M, Hammond D, Childrens Cancer Study Group, Los Angeles, CA 90031 (1982b) The Duration of Chemotherapy for Childhood Acute Lymphoblastic Leukemia (ALL): A Randomized Study of 316 Patients *Proceedings of the American Society of Clinical Oncology (ASCO)* **124 (or 1/-)** a C480 80

Nesbit ME Jr, Sather HN, Robison LL, Ortega JA, Hammond GD and Childrens' Cancer Study Group (1983) Randomized Study of 3 Years vs 5 Years of Chemotherapy in Childhood Acute Lymphoblastic Leukemia *Journal of Clinical Oncology* **1 no 5 (May)** 308-316 75

Nesbit ME Jr, D'Angio GJ, Sather HN, Robison LL, Ortega JA, Hammond D. Reply: Freeman AI. (1984) Post-induction treatment of Childhood Acute Lymphoblastic Leukaemia *New England Journal of Medicine* **310 no 4** 262-263 62

Niemeyer CM, Gelber RD, Blattner SR, Tarbell NJ, Donahue K, Sallan SE, DFCI Institute (1987) Importance of Early Intensive Therapy in Childhood Acute Lymphoblastic Leukemia (ALL) *Blood* **70 suppl 235a** a797 129

Niemeyer CM, Gelber RD, Tarbell NJ, Donnelly M, Clavell LA, Blattner SR, Donahue K, Cohen HJ, Sallan SE (1991a) Low-Dose versus High-Dose Methotrexate During Remission Induction in Chilhood Acute Lymphoblastic Leukemia (Protocol 81-01 Update) *Blood* **78 no 10** 2514-2519 128

Niemeyer CM Reiter A, Riehm H, Donnelly M, Gelber RD, Sallan SE (1991b) Comparative Results of Two Intensive Treatment Programs for Childhood Acute Lymphoblastic Leukaemia: The Berlin-Frankfurt-Munster and Dana-Farber Cancer Institute Protocols *Annals of Oncology* **2 (10)** 745-9 38

Ochs JJ, Parvey LS, Whitaker JN, Bowman WP, Chien L, Campbell M, Coburn T, SJCRH (1983) Serial Cranial Computed-Tomography Scans in Children with Leukemia Given Two Different Forms of Central Nervous System Therapy *Journal of Clinical Oncology* **1 no 12 (Dec)** 793-798 221

Ochs J, Parvey LS, Mulhern R, SJCRH (1986) Prospective Study of Central Nervous System Changes in Children with Acute Lymphoblastic Leukemia Receiving Two Different Methods of Central Nervous System Prophylaxis *Neurotoxicity* **7(2)** 217-226 223

Ochs J, Mulhern R, Fairclough D, Parvey L, Ch'ien L, Mauer A, Simone J, SJCRH (1989) Prospective evaluation of Central Nevous System (CNS) Changes in Children with Acute Lymphoblastic Leukemia (ALL) treated with Prophylactic Cranial Irradiation (RT) or IV Methotrexate (MTX) *Proceedings of the American Society of Clinical Oncology (ASCO)* **8** 212 a824 231

Ochs J, Mulhern R, Fairclough D, Parvey L, Whitaker J, Ch'ien L, Mauer A, Simone J, SJCRH (1991a) Comparison of Neuropsychologic Functioning and Clinical Indicators of Neurotoxicity in Long-Term Survivors of Childhood Leukemia Given Cranial Radiation or Parenteral Methotrexate. A Prospective Study. *Journal of Clinical Oncology* **9 no 1 (Jan)** 145-151 229

Ochs J, Rodman J, Abromowitch M, Kavanagh R, Harris M, Yalowich J, Rivera GK (1991b) A Phase II Study of Combined Methotrexate and Teniposide Infusions Prior to Reinduction Therapy in Relapsed Childhood Acute Lymphoblastic Leukemia: A Pediatric Oncology Group Study *Journal of Clinical Oncology* **9 no 1 (Jan)** 139-144 201

Ortega JJ on behalf of Pethema, Spanish Cooperation Group, Barcelona, Spain. (1985) Intensive Induction-Consolidation Therapy in Childhood and Adult Acute Lymphoblastic Leukemia (Protocol LAL 17/84) *Neoplasia* **3** 101-107 260

Ortega Aramburu JJ, Javier G, Montagut JM, Toran N, Ciudad Sanitaria "Valle de Hebron", Universidad Autonoma, Barcelona (1985) Tratamiento de las leucemias agudas linfoblasticas de alto ybajo riesgo en el nino, con dos modalidades de terapeutica preventiva sobre SNC (Protocolo Pethema 7/78) (Treatment of Standard and High-Risk Acute Lymphoblastic Leukemia in Children with Two Modalities of CNS Prophylaxis (Protocol Pethema Lal 7/78)) *Anales Espanoles de Pediatria* **23** 417-430 257

Ortega JJ, Javier G, Olive T (1987a) Treatment of Standard and High-Risk Childhood Acute Lymphoblastic Leukemia with Two CNS Prophylaxis Regimens *Haematology and Blood Transfusion (Acute Leukemias)* Buchner, Schellong, Hiddemann, Urbanitz, Ritter (eds.) **30** pp 483-492 Springer-Verlag: Berlin, Heidelberg 258

Ortega JA, Nesbit ME, Sather HN, Robison LL, D'Argio G Jr, Hammond GD (1987b) Long-Term Evaluation of a CNS Prophylaxis Trial - Treatment Comparisons and Outcome After CNS Relapse in Childhood ALL: A Report from the Childrens Cancer study Group *Journal of Clinical Oncology* **5 no 10 (Oct)** 1646-1654 118

Ortega JJ, Javier G, Olive T y PETHEMA (1988) Treatment of high-risk acute lymphoblastic leukemia with protocols PETHEMA LAL 7/78 and 17/84 *Anales Espanoles de Pediatria* **29 (34)** 72-83 287

Ortega J J (1998) Spanish Acute Lymphoblastic Leukemia Trials *International Journal of Pediatric Hematology/Oncology* **5 (2-4)** 163-176 286

Otten J, Boilletot A, Philippe N, Munzer M, Robert A, Manel AM, Hoyoux C, Thyss A, Benoit Y, Marguerite G, Souillet F, Solbu G, Suciu S, for EORTC Children's Leukemia Cooperative Group. (1988a) Treatment of Children with Medium and High Risk (M-HR) Acute Lymphoblastic Leukemia (ALL) with or without Radiotherapy (RXT) to the Brain: EORTC NR 58832 Randomized Trial *Proceedings of the American Association of Clinucal Oncology (ASCO)* **7** 175 a676 143

Otten J, Stryckmans P, Benoit Y, Chantraine JM, Gyselinck J, Hainaut H, Mauru SR, Solbu G, Suciu S (1988b) Comparison of Chemotherapy with Immunotherapy as Maintenance Treatment in Acute Lymphoblastic Leukemia (ALL): A long term re-evaluation (EORTC randomized trial nr 58741) *Medical and Pediatric Oncology: International Society of Pediatric Oncology (SIOP) abstracts* **16** 396 135

Paolucci G, Masea MD, Vecchi V, Marsoni S, Zurlo MG (1989) Treating Childhood Acute Lymphoblastic Leukaemia (ALL): Summary of Ten Years' Experience in Italy. ALL Steering Committee of the Associazione Italiana Ematologica Oncologia Pediatrica (AIEOP) *Medical & Pediatric Oncology* **17** 83-91 30

Peto J, Eden OB, Lilleyman J, Richards S (1986) Improvement in treatment for children with Acute Lymphoblastic Leukaemia. The Medical Research Council UKALL Trials, 1972-84. Report to the Council by the Working Party on Leukaemia in Childhood *Lancet* **1** 408-411 18

Peylon-Ramu N, Poplack DG, Pizzo PA, Adornato BT, di Chiro G (1978) Abnormal CT scans of the Brain in Asymptomatic Children with Acute Lymphocytic Leukemia after Prophylactic Treatment of the Central Nervous Sysytem with Radiation and Intrathecal Chemotherapy *New England Journal of Medicine* **298 no 15** 815-818 169

Pinkel D, Hernandez K, Borella L, Holton C, Aur R, Samoy G, Pratt C (1971) Drug dosage and remission duration in Childhood Lymphocytic Leukemia *Cancer* **27** 247-256 207

Pinkel D, Hustu HO, Aur RJA, Smith K, Borella LD, Simone J (1977) Radiotherapy in Leukemia and Lymphoma of Children *Cancer* **39** 817-824 211

Pinkel D (1979) The Ninth Annual David Karnofsky Lecture. Treatment of Acute Lymphoblastic Leukemia *Cancer* **43** 1128-37 219

Poplack DG, Graw RG, Pomeroy TC, Henderson ES, Leventhal BG, National Cancer Institute (1975) Chemotherapy (CT) vs Chemotherapy and Immunotherapy (CT+IMT) in Childhood Acute Lymphatic Leukemia (ALL) *Proceedings of the American Association for Cancer Research (AACR) and American Society for Clinical Oncology (ASCO)* **16** a1038 167

Poplack DG, Reaman GH, Bleyer WA, Miser J, Feusner J, Wesley R, Hammond D, Paediatric Branch, NCI, CCSG. (1984) Central Nervous System (CNS) Preventative Therapy with High Dose Methotrexate (HD MTX) in Acute Lymphoblastic Leukemia (ALL): A preliminary report *20th Meeting of the American Society of Clinical Oncology (ASCO) May 6-8, 1984 Proceedings* **3** 204 aC-797* 170

Poplack DG (1987) NCI77-02/CCG191 *1987 International Workshop on ALL. Assessment of Progress and Future Directions 28-30 October 1987* 24-25 172

Pui C-H, Aur RJA, Bowman WP, Dahl GV, Dodge RK, George SL, Ochs J, Kalwinsky DK, Abromowitch M, Hustu HO, Simone JV, SJCRH (1984) Failure of Late Intensification Therapy to Improve a Poor Result in Childhood Lymphoblastic Leukemia *Cancer Research* **44** 3593-3598 238

Pui C-H, Dahl GV, Bowman WP, Rao BN, Abromowitch M, Ochs J, Rivera G, SJCRH (1985) Elective Testicular Biopsy During Chemotherapy for Childhood Leukaemia is of no Clinical Value *Lancet* **2** 410-412 222

Pui C-H, Behm FG, Raimondi SC, Dodge RK, George SL, Rivera GK, Mirro J, Kalwinsky DK, Dahl GV, Murphy SB, Crist WM, Williams DL, SJCRH (1989) Secondary Acute Myeloid Leukemia in Children treated for Acute Lymphoblastic Leukemia *New England Journal of Medicine* **321 no 3** 136-142 226

Pui C-H, Ribeiro RC, Hancock ML, Rivera GK, Evans WE, Raimondi SC, Head DR, Behm FG, Mahmoud MH, Sandlund JT, Crist WM, SJCRH (1991) Acute Myeloid Leukemia in Children Treated with Epipodophyllotoxins for Acute Lymphoblastic Leukemia *New England Journal of Medicine* **325 no 24** 1682-1687 235

Pui C-H, Simone JV, Hancock ML, Evans WE, Williams DL, Bowman WP, Dahl GV, Dodge RK, Ochs J, Abromowitch M, Rivera GK, SJCRH. (1992) Impact of Three Methods of Treatment Intensification on Acute Lymphoblastic Leukemia in Children: Long-Term Results of St Jude Total Therapy Study X. *Leukemia* **6 no 2 (Feb)** 150-157 230

Pullen DJ, Sullivan MP, Falletta JM, Boyett JM, Humphrey GB, Starling KA, Land VJ, Dyment PG, Vats T, Duncan MH (1982) Modified LSA2-L2 Treatment in 53 Children with E-Rosette-Positive T-Cell Leukemia: Results and Prognostic Factors (A Pediatric Oncology Group Study) *Blood* **60 no 5 (Nov)** 1159-1168 174

Pullen J, Boyett J, Frankel L, Iyer R, van Eys J, Crist W, Harris M, Ravindranath Y, Sullivan M (1988) Extended triple intrathecal (TIT) chemotherapy provides effective Central Nervous System (CNS) prophylaxis for both good and poor prognosis patients with non-T, non-B acute lymphoblastic leukemia (ALL); substitution of Intermediate Dose Methotrexate (IDM) for TIT after consolidation provides less effective protection for the CNS. A Pediatric Oncology Group (POG) study *Proceedings of the American Society of Clinical Oncology (ASCO)* **7** 176 a681 185

Pullen J, Boyett J, Shuster J, Crist W, Land V, Frankel L, Iyer R, Backstrom L, van Eys J, Harris M, Ravindranath Y, Sullivan M (1993) Extended Triple Intrathecal Chemotherapy Trial for Prevention of CNS Relapse in Good-Risk and Poor-Risk Patients with B-Progenitor Acute Lymphoblastic Leukemia: A Pediatric Oncology Group Study *Journal of Clinical Oncology* **11 no 5 (May)** 839-849 187

Ragab AH, Boyett JM, Frankel L, Falletta J (1986) Rubidazone in the Treatment of Recurrent Acute Leukemia in Children. A Pediatric Oncology Group Study. *Cancer* **57** 1461-1463 199

Rausen AR, Glidewell O, Holland JF, Ohnuma T, Sinks L, Freeman A, Cuttner J, Patterson RB, Falkson G, Kung FH, Levy RN (1979) Superiority of L-Asparaginase combination of chemotherapy in advanced acute lymphocytic leukemia of childhood. Randomized comparative trial of combination versus solo therapy. *Cancer Clinical Trials* **2 Summer** 137-144 69

Reiter A, Schrappe M, Ludwig W-D, Hiddemann W, Sauter S, Henze G, Zimmermann M, Lampert F, Havers W, Niethammer D, Odenwald E, Ritter J, Mann G, Welte K, Gadner H, Riehm H (1994) Chemotherapy in 998 Unselected Childhood Acute Lymphoblastic Leukemia Patients. Results and Conclusions of the Multicenter trial ALL-BFM 86 *Blood* **84 no 9 (November 1)** 3122-3133 45

Riccheri C, Dibar E, Aversa L, Sackmann Muriel F and members of GATLA (1994) A Comparison of Two Different Regimen fro Continuation Chemotherapy for Treatment of Childhood High Risk ALL *Proceedings of the American Society of Clinical Oncology (ASCO)* **13** 320 a1055 150

Richards S, Burrett J, Hann I, Chessells J, Hill F, Bailey C for the Medical Research Council Working Party on Childhood Leukaemia. (1998) Improved survival with early intensification: combined results from the Medical Research Council childhood ALL randomized trials, UKALL X and UKALL XI *Leukemia* **12 (7)** 1031-1036 283

Riehm H, Gadner H, Henze G, Kornhuber B, Langermann H-J, Muller-Weihrich S, Schellong G (1984) Acute Lymphoblastic Leukemia: Treatment Results in Three BFM Studies (1970-1981). *Leukemia Research: Advances in Cell Biology and Treatment* Murphy SB, Gilbert JR (eds) pp 51-263 Elsevier Science Publishing Co. Inc.: Oxford 37

Riehm H, Feickert H-J, Schrappe M, Henze G, Schellong G for BFM Study Group (1987) Therapy Results in five ALL-BFM Studies since 1970: Implications of Risk Factors for Prognosis *Haematology and Blood Transfusion* Buchner, Schellong, Hiddemann, Urbanitz, Ritter (eds.) **30 Acute Leukemias** pp 139-146 Springer-Verlag: Berlin, Heidelberg 46

Riehm H, Gadner H, Henze G, Kornhuber B, Lampert F, Niethammer D, Reiter A, Schellong G (1990) Results and Significance of Six Randomized Trials in Four Consecutive ALL-BFM Studies *Haematology and Blood Transfusion* **33** 439-450 32

Riehm H, Schrappe M, Dordelmann M, Welte K, Reiter A, Department of Pediatric Hematology & Oncology, Hannover Medical School 30623, Hannover, Germany (1996) Update Results of ALL-BFM Protocols *British Journal of Haematology* **93 suppl 2** 145 A556 281

Ritter J, Creutzig U, Reiter A, Riehm H, Schellong G (1990) Childhood Leukemia: Cooperative Berlin-Frankfurt-Munster trials in the Federal Republic of Germany 21st Symposium of The Gesellschaft zur

Bekamfung der Krebskrankheiten Nordrhein-Westfalen (GBK), Dusseldorf, June 1989 *Journal of Cancer Research and Clinical Oncology* **116** 100-103 48

Rivera G, Pui C-H, Mirro J, Raimondi S, Look A, Crist W, SJCRH (1990) Rotational Combination Chemotherapy for Childhood Acute Lymphocytic Leukemia (ALL): St Jude Study XI *26th Annual Meeting of the American Society of Clinical Oncology (ASCO). May 20-22, 1990. Program/Proceedings, Washington DC.* **9** A844 232

Rivera GK, Pui CH, Hancock M, Crist WM, SJCRH (1991a) Update of St Jude Study XI for Childhood Acute Lymphoblastic Leukemia (ALL) *Haematologica* **76 supp no 4** 104 a.N406 234

Rivera GK, Raimondi SC, Hancock ML, Behm FG, Pui C-H, Abromowitch M, Mirro J Jr, Ochs JS, Look AT, Williams DL, Murphy SB, Dahl GV, Kalwinsky DK, Evans WE, Kun LE, Simone JV, Crist WM, SJCRH (1991b) Improved outcome in childhood acute lymphoblastic leukemia with reinforced early treatment and rotational combination chemotherapy *Lancet* **337 no 8733** 61-66 233

Robison LL, Sather HN, Coccia PF, Nesbit ME, Hammond GD (1980) Assessment of the interrelationship of prognostic factors in childhood acute lymphoblastic leukemia. A report from Childrens Cancer Study Group. *American Journal of Pediatric Hematology/Oncology* **2 no 1** 5-13 111

Robison LL, Nesbit Jr ME, Sather HN, Meadows AT, Ortega JA, Hammond GD from CCSG (1984) Factors associated with IQ scores in long term survivors of childhood acute lymphoblastic leukemia *American Journal of Pediatric Hematology/Oncology* **6 no 2** 115-121 117

Rossi MR, Masera G, Zurlo MG, Amadori S, Mandelli F, Bagnulo S, Carli M, Zanesco L, Dini G, Guazzelli C, Madon E, Nespoli L, Paolucci G, Pession A, Tamaro P (1986) Randomized multicentric Italian study on two treatment regimens for marrow relapse in childhood acute lymphoblastic leukemia *Pediatric Hematology and Oncology* **3 (1)** 1-9 31

Rowland JH, Glidewell OJ, Sibley RF, Holland JC, Tull R, Berman A, Brecher ML, Harris M, Glicksman AS, Forman E, Jones B, Cohen ME, Duffner PK, Freeman AI for Cancer and Leukaemia Group B. (1984) Effects of Different Forms of Central Nervous System Prophylaxis on Neuropsychologic Function in Childhood Leukaemia. *Journal of Clinical Oncology* **2 no 12 (December)** 1327-1335 54

Rubic H, Benoit Y, Behar C, Lutz P, Maurus R, Philippe N, Plouvier E, Robert A, Sauveur E, Solbu G, Suciu S, Otten J (1988) Treatment of Standard Risk (SR) Acute Lymphoblastic Leukemia (ALL) with the BFM Protocol with or without cyclophosphamide: Preliminary Results of a Randomized Trial (EORTC nr 58831) *Medical and Pediatric Oncology: International Society for Pediatric Oncolocy (SIOP) extract* **16** 389 139

Sackmann-Muriel F, Svarch E, Eppinger-Helft M, Braier JL, Pavlovsky S, Guman L, Vergara B, Ponzinibbio C, Failace R, Guray GE, Bugnard E, Ojeola FG, de Bellis R, de Sijvarger SR, Saslavsky (1978) Evaluation of Intensification and maintenance Programs in the Treatment of Acute Lymphoblastic Leukemia *Cancer* **42 (4)** 1730-1740 148

Sackmann-Muriel F, Pavlovsky S, Lastin F, Bustelo P, Jiminez E, Svarch E, Hospital de Pediatria S AMIC Prof Dr J P Garrahan, Servicio de Hematooncologica Combate de los Pozos 1881, Rm # 3309, (1245) Buenos Aires Argentina (1998) Latin American Trials in Childhood Acute Lymphoblastic Leukemia. GATLA/GLATHEM Report of Results from 1967 through 1994. *International Journal of Pediatric Hematology/Oncology* **5 (2-4)** 177-185 284

Sallan SE, Camitta BM, Frei E III, Furman L, Leavitt P, Bishop Y, Jaffee N, DFCI (1977) Clinical and Cytokinetic Aspects of Remission Induction of Childhood Acute Lymphoblastic Leukemia (ALL): Addition of an Anthracycline to Vincristine and Prednisolone *Medical and Pediatric Oncology* **3** 281-287 121

Sallan SE, Hitchcock-Bryan S, Gelber R, Cassady JR, Frei III E, Nathan DG (1983) Influence of Intensive Asparaginase in the Treatment of Childhood Non-T-cell Acute Lymphoblastic Leukemia *Cancer Research* **43** 5601-5607 123

Sallan SE, Gelber RD, Kimball V, Donnelly M, Cohen HJ (1990) More is better! Update of Dana-Farber Cancer Institute/Children's Hospital Childhood Acute Lymphoblastic Leukemia Trials *Haematology and Blood Transfusion* Buchner, Schellong, Hiddemann, Ritter (eds.) **33** pp 459-466 Springer-Verlag: Berlin, Heidelberg 126

Schaison G, Leverger G, Bancillon A, Marty M, Olive D, Cornu G, Griscelli C, Lemerle S, Haronsseau JL, Bonnet M, Freycon F, Dufillot D, Demeocq M, Bauters F, Lamagnere JP, Taboureau O. (1987) Intermediate Risk Childhood Acute Lymphoblastic Leukemias: Amsacrine + Cytosine Arabinoside versus Intermediate-Dose Methotrexate for Consolidation, and 6-Mercaptopurine + Methotrexate + Vincristine versus Monthly Pulses for Maintenance *Haematology and Blood Transfusion* Buchner, Schellong, Hiddemann, Urbanitz, Ritter (eds.) **30** pp 461-465 Springer-Verlag: Berlin, Heidelberg 146

Schaison G, Madelaine I, Baruchel A, Bellanger P, Faure P, Leverger G, Stettler E (1990a) Diffusion of High Dose Methotrexate (HD MTX) into Cerebro Spinal Fluid (CSF) of Leukemic Patients. Comparative Study of 3g/m^2 versus 8g/m^2 *Proc Am Soc Clin Oncol (26th Annual Meeting of the American Society of Clinical Oncology May 20-22,1990. Program/Proceedings.* **9** 220 a852 147

Schaison G, Olive D, Leverger G, Vannier JP, de Lumley L, Bancillon A, Cornu G (1990b) Treatment of Acute Lymphoblastic Leukemia: Protocol Fralle 83-85 *Haematology and Blood Transfusion* **33** 467-472 144

Schrappe M, Beck J, Brandeis WE, Feickert H-J, Gadner H, Graf N, Havers W, Henze G, Jobke A, Kornhuber B, Kuhl J, Lampert F, Muller-Weihrich S, Niethammer D, Reiter A, Rister M, Ritter J, Schellong G, Tausch W, Weinel P, Riehm H (1987) Treatment of acute lymphoblastic leukaemia in young age: Results of multicenter study ALL-BFM 81. (Die Behandlung der akuten lymphoblastischen Leukamie im Kinder - und Jugendalter: Ergebnisse der multizentrischen Therapiestudie ALL-BFM 81. *Klinische Padiatrie* **199 (3 Mai)** 133-150 35

Shuster JJ, Holland JF for Cancer and Leukemia Group B. (1984) Lower Incidence of Meningeal Leukaemia when Dexamethasone is substituted for Prednisolone in the Treatment of Acute Lymphocytic Leukemia - A Late Follow-up. *20th Annual Meeting of the American Society of Clinical Oncology (ASCO) May 6-8,1984. Proceedings. Toronto, Ontario, Canada. Scientific Proceedings* **3** 191 C-744 58

Shuster JJ, Falletta JM, Pullen DJ, Crist WM, humphrey GB, Dowell BL, Wharam MD, Borowitz MJ (1990) Prognostic Factors in Childhood T-Cell Acute Lymphoblastic Leukemia: A Pediatric Oncology Group Study *Blood* **75 no1** 166-173 182

Shuster J, Look T, Crist W, Borowitz M, Carrol A, Pullen J, Steuber P, Land V (1991) Clinical Features and Tumor Cell DNA Index Improve Prediction of Treatment Outcome in Pediatric B-Progenitor Acute Lymphoblastic Leukemia (ALL): A Pediatric Oncology Group (POG) Study *27th Annual Meeting of the American Society of Clinical Oncology (ASCO) May 19-21, 1991 Program/Proceedings, Houston TX* **10** 234 194

Simone JV, Chief of Hematology, SJCRH (1974) Acute Lymphocytic Leukemia in Childhood *Seminars in Haematology* **11 no 1 (Jan)** 25-39 209

Simone JV, Aur RJA, Hustu HO, Verzosa M, Pinkel D (1975) Combined Modality Therapy of Acute Lymphocytic Leukemia *Cancer* **35** 25-35 210

Smith SD, Sadowitz PD, Shuster J, Rivera GK, Pediatric Oncology Group (1990) Treatment of late bone marrow (LBM) relapses in children with acute lymphoblastic leukemia (ALL): A pediatric oncology group (POG) study *Proceedings of the 81st Annual Meeting of the American Association for Cancer Research (AACR). May 23-26, 1990 Washington DC* **31** a1193 195

Southwest Oncology Group (participants included Bickers JN, Gehan EA, Freireich EJ, Coltman Jr CA, Wilson HE, Hewlett JS, Stuckey WJ, van Slyck EJ (1974) Cytarabine for Acute Leukemia in Adults. Effect of Schedule on Therapeutic Response *Archives of Internal Medicine* **133** 251-259 247

Stark B, Cohen IJ, Kaplinsky C, Tamari H, Yaniv I, Zaizov R, Sambun Center for Pediatric Hematology/Oncology, Beilinson Medical Center, Sackler Faculty of Medicine, Tel Aviv University, Israel (1992) Successful Central Nervous System Prophylaxis without Radiotherapy in Childhood Acute Lymphoblastic Leukemia (ALL), Israel National Studies (INS '84, '89) *Proceedings of the American Society of Clinical Oncology (ASCO)* **11** 283 a942 272

Stark B, Abramov A, Attias D, Balin A, Burstein Y, Chividalli G, Ramu N, Rechavi G, Sharon R, Sthoeger D, Zaizov R, Pediatric Hematology Oncology, Schneider Children's Medical Center of Israel, Petach Tikvah, Israel (1996) Extended Intrathecal Triple Therapy for Preventing CNS Relapse in non very high risk Childhood Acute Lymphoblastic Leukemia treated with the Israeli Studies INS-84 and -89 (modified BFM-86) *Proceedings of the American Society of Clinical Oncology (ASCO)* **15** a1087 273

Steinherz P, Gaynon P, Grossman N, Kersey J, Kranstuber S, Maurer H, Sather H, Trigg M, Cherlow J, Hammond D, Childrens Cancer Study Group (1988) Radiotherapy to Extra-abdominal Bulk Disease and/or for Central Nervous System Prophylaxis with Intensive Chemotherapy of Acute Lymphoblastic Leukemia and Lymphomatous Presenting Features *Proceedings of the American Association for Cancer Research (AACR)* **29** 213 (a845) 91

Steinherz PG, Redner A, Steinherz L, Meyers P, Tan C, Heller G (1993) Development of a New Intensive Therapy for Acute Lymphoblastic Leukemia in Childhood at Increased Risk of Early Relapse. The Memorial Sloan-Kettering-New York II Protocol *Cancer* **72 no 10** 3120-3130 266

Steinherz PG, Gaynon PS, Breneman JC, Cherlow JM, Grossman NJ, Kersey JH, Johnstone HS, Sather HN, Trigg ME, Uckun FM, Bleyer WA (1998) Treatment of Patients with Acute Lymphoblastic Leukaemia with Bulky Extramedullary Disease and T-Cell Phenotype or Other Poor Prognostic Features. Randomized Controlled Trial from the Childrens Cancer Group *Cancer* **82 no 3** 600-612 276

Stryckmans P, Marie JP, Debusscher L, Bury J, Peetermans M, Andrien JM, Solbu G, Suciu S, Fiere D, Cauchie Ch, Zittoun R (Intr. By A Chanana). The EORTC Study Group. (1985) Chemotherapy of Adult Acute Lymphoblastic Leukemia (ALL). Protocol 58791. A Randomized Study. *Blood* **66 suppl 1** 209a a721 137

Stryckmans P, Marie JP, Debusscher L, Bury J, Peetermans M, Andrieu JM, Solbu G, Suciu S, Fiere D, Cauchie Ch, Zittoun R (1986) Chemotherapy of Residual Disease in Adult Acute Lymphoblastic Leukemia (ALL). A randomized study of the EORTC Leukemia Lymphoma Study Group *Leukaemia Research (Scientific Program and Abstracts)* **10** 112 136

Stryckmans P, Marie JP, Suciu S, Solbu G, Debusscher L, Bury J, Peetermans M, Andrieu JM, Fiere D, Cauchie C, van Camp B, Zittoun R (1987) Therapy for Adolescent and Adult Acute Lymphoblastic Leukemia: Randomization of Induction and Consolidation Therapies (Preliminary Results of EORTC Study 58791) *Haematology and Blood Transfusion* Buchner, Schellong, Hiddemann, Urbanitz, Ritter (eds.) **30** pp 130-136 Springer-Verlag: Berlin, Heidelberg 138

Sullivan MP, Chen T, Dyment PG, Hvizdala E, Steuber CP (1982) Equivalence of Intrathecal Chemotherapy and Radiotherapy as Central Nervous System Prophylaxis in Children with Acute Lymphatic Leukemia: A Pediatric Oncology Group Study *Blood* **60 no 4 (Oct)** 948-958 175

Tarbell N, Waber D, Cohen H, Gelber R, Dalton V, Donnelly M, Sallan S, Dana-Farber Cancer Institute (1991) Hyperfractionated Cranial Irradiation (HCI) in Childhood Acute Lymphoblastic Leukemia (ALL): Rationale and Preliminary Results *Proceedings of the American Society of Clinical Oncology (ASCO)* **10 (Mar)** 239 131

Tokyo Children's Cancer Study Group (ALL Committee): Tsuchida M, Akatsuka J-I, Bessho F, Chihara H, Hayashi Y, Hoshi Y, Hosoya R, Furukawa T, Ikuta K, Inara I, Ishikawa A, Ishimoto K, Ho K, Kaneko M, Kaneko T, Katoi S-I, Komiyama J, Matsuyamo S, Nagao T, Naikazawa S, Nishihara K, Ohira M, Okimoto Y, Ohkawa Y, Ohtsuki H, Sato T, Shibuya A, Shitara T, Sugita K-I, Taguchi N, Torigoe K, Tsukada M, Tsukimoto I, Tsunematsu Y, Wada E, Yamada K, Yamada K, Yamamoto K, Yamamoto M, Yata J-I, Nishimura K, Saito T (1991) Treatment of AcuteLymphoblastic Leukemia in the Tokyo Children's Cancer Study Group - Preliminary Results of L84-11 Protocol *Acta Paediatrica Japonica* **33 no 4** 522-532 267

Tsukada M, Komiyama A, Nakazawa S, Tsuchida M, Nishihira H, Shitara T, Ohira M, Tsunematsu Y, Yamamoto K, Hoshi Y, Yamada K, Hosoya Y, Sato T, Bessho F, Tsukimoto I, Yamamoto M, Ikuta K, Saito T, Nishimura K & the Tokyo Children Cancer Study Group (1993) Treatment of standard risk acute lymphoblastic leukemia in children with the Tokyo Children Cancer Study Group (TCCSG) L84-11 protocol in Japan *International Journal of Hematology* **57** 1-7 268

Tsurusawa M, Katano N, Yamamoto Y, Hirota T, Koizumi S, Watanabe A, Takeda T, Hatae Y, Yatabe M, Mimaya J, Gushihen T, Nishi U, Anami K, Kikuta A, Kanegane H, Asami K, Nishikawa K, Sekine I, Kawano Y, Iwai A, Furuyama T, Ijichi O, Miyake M, Mugishima H, Ota S, Fujimoto T, for the Children's Cancer and Leukemia Study Group (1999) Improvement in CNS Protective Treatment in Non-High-Risk Childhood Acute Lymphoblastic Leukemia: Report From the Japanese Children's Cancer and Leukemia Study Group *Medical and Pediatric Oncology* **32** 259-266 76

Tubergen D, Gilchrist G, Sather H, Coccia P, Novak L, O'Brien R, Waskerwitz M, Hammond D. For Children's Cancer Study Group, Pasadena, California, 91101. (1988) Intrathecal methotrexate (IT MTX) provides adequate Central Nervous System (CNS) therapy in Acute Lymphoblastic Leukemia patients with Intermediate Risk Features and an age of less than 10 years *Proceedings of the American Society of Clinical Oncology (ASCO) Leukemia (adult and pediatric)* **7** 178 abstract 688 81

Tubergen D, Gilchrist G, Cocchia P, Novak L, O'Brien R, Waskerwitz M, Sather H, Bleyer A, Hammond D. For Children's Cancer Study Group, Pasadena, California, 91101 (1990) The Role of Intensified Chemotherapy in Intermediate Risk Acute Lymphoblastic Leukaemia (ALL) of Childhood CCG-105. *26th Annual Meeting of the American Society of Clinical Oncology May 20-22, 1990. Program/Proceedings Washington DC* **9** no 835 78

Tubergen DG, Gilchrist GS, O'Brien RT, Coccia PF, Sather HN, Waskerwitz MJ, Hammond GD (1993a) Prevention of CNS Disease in Intermediate-Risk Acute Lymphoblastic Leukemia: Comparison of Cranial Radiation and Intrathecal Methotrexate and the Importance of Systemic Therapy: A Childrens Cancer Group Report *Journal of Clinical Oncology* **11 no 3 (March)** 520-526 82

Tubergen DG, Gilchrist GS, O'Brien RT, Coccia PF, Sather HN, Waskerwitz MJ, Hammond GD (1993b) Improved Outcome with Delayed Intensification for Children with Acute Lymphoblastic Leukemia and Intermediate Presenting Features: A Childrens Cancer Group Phase III Study *Journal of Clinical Oncology* **11 no 3 (March)** 527-537 84

van der Does-van den Berg A, van Wrins ER, de Koning J, Rammeloo JA, Solbu G, Suciu S, van Zanen GE (1987) Addition of Rubidomycin to Induction Treatment with Vincristine, Prednisolone and L-asparaginase in Standard-Risk Childhood Acute Lymphoblastic Leukemia (study ALL V). A Report on Behalf of the Dutch Childhood Leukemia Study Group *Haematology and Blood Transfusion* Buchner, Schellong, Hiddemann, Urbanitz, Ritter (eds.) **30** pp 444-447 Springer-Verlag: Berlin, Heidelberg 134

van der Does-van den Berg A, van Wering ER, Suciu S, Solbu G, Rammeloo JA, Koning J de, En GE van Zanen (1988) Resultaten van behandling van kinderen met acute lumfatische leukemie (ALL) volgens SNWLK protocol ALL V *Tijdschr Kindergeneesk* **56 no 2** 61-66 132

van der Does-van den Berg A, van Wering ER, Suciu S, Solbu G, van'tVeer MB, Rammeloo JA, de Koning J, van Zanen GE (1989) Effectiveness of Rubidomycin in Induction Therapy with Vincristine, Prednisolone, and L-Asparaginase for Standard Risk Childhood Acute Lymphoblastic Leukemia: Results of a Dutch Phase III Study (ALL V). A Report on Behalf of the Dutch Childhood Leukemia Study Group (DCLSG) *American Journal of Pediatric Hematology/Oncology* **11 (2)** 125-133 133

van Eys J, Berry DM, Crist W, Doering EJ, Fernbach DJ, Pullen J, Shuster J (1987) Effect of Tremithoprim/Sulfamethoxazole Prophylaxis on Outcome of Childhood Lymphocytic Leukemia. A Pediatric Oncology Group Study *Cancer* **59 no1** 19-23 181

van Eys J, Berry D, Crist W, Doering E, Fernbach D, Pullen J, Shuster J, Wharam M (1989a) A Comparison of Two Regimens for High-Risk Acute Lymphocytic Leukemia in Childhood. A Pediatric Oncology Group Study *Cancer* **63 no 1** 23-29 179

van Eys J, Berry D, Crist W, Doering E, Fernbach D, Pullen J, Shuster J (1989b) Treatment Intensity and Outcome for Children with Acute Lymphocytic Leukemia of Standard Risk. A Pediatric Oncology Group Study *Cancer* **63 no 8** 1466-1471 177

Vats T, Buchanan G, Mehta P, Ragab A, Hvizdale E, Nitschke R, Link M, Beardsley GP, Maybee D, Krischer J (1992) A study of toxicity and comparative therapeutic efficacy of vindesine-prednisolone vs vincristine-prednisolone in children with acute lymphoblastic leukemia in relapse. A Pediatric Oncology Group Study *Investigational New Drugs* **10** 231-234 206

Waber DP, Tarbell NJ, Kahn CM, Gelber RD, Sallan SE (1992) The Relationship of Sex and Treatment Modality to Neuropsychologic Outcome in Childhood Acute Lymphoblastic Leukemia *Journal of Clinical Oncology* **10 no 5 (May)** 810-817 127

Waber DP, Tarbell NJ, Fairclough D, Atmore K, Castro R, Isquith P, Lussier F, Romero I, Carpenter PJ, Schiller M, Sallan SE (1995) Cognitive sequelae of Treatment in Childhood Acute Lymphoblastic Leukemia: Cranial Radiation Requires an Accomplice *Journal of Clinical Oncology* **13 no 10 (Oct)** 2490-2496 130

Wells RJ, Foster MB, d'Ercole AJ, McMillan CW, Dept. Paed, Univ N Carolina (1983) The Impact of Cranial Irradiation on the Growth of Children with Acute Lymphocytic Leukemia *American Journal of Diseases of Children* **137** 37-39 251

Whitt JK, Wells RJ, Lauria MM, Wilhelm CL, McMillan CW (1984) Cranial Radiation in Childhood Acute Lymphocytic Leukemia. Neuropsychologic Sequelae *American Journal of Diseases of Children* **138** 730-736 252

Williams KS, Ochs J, Williams JM, Mulhern RK (1991) Parental Report of Everyday Cognitive Abilities Among Children Treated for Acute Lymphoblastic Leukemia *Journal of Pediatric Psychology* **16 no1** 13-26 228

Willoughby MLN (1976) Treatment of overt meningeal leukaemia in children: results of second MRC meningeal leukaemia trial *British Medical Journal* **1** 864-867 28

Willoughby MLN (1983) Chapter 13 Treatment of Overt CNS Leukaemia Mastrangelo R, Poplack DG, and Riccardi R (eds.) pp 113-122 *Central Nervous System Leukaemia* Martinus Nijhoff Publishers: Boston 29

Wofford MM, Smith SD, Shuster JJ, Johnson W, Buchanan GR, Wharam MD, Ritchey AK, Rosen D, Haggard ME, Golembe BL, Rivera GK (1992) Treatment of Occult or Late Overt Testicular Relapse in Children with Acute Lymphoblastic Leukemia: A Pediatric Oncology Group Study *Journal of Clinical Oncology* **10 no 4 (April)** 624-630 196

Wolfrom C, Hartmann R, Fengler R, Bruhmuller S, Ingwersen A, Henze G (1993) Randomized Comparison of 36-Hour Intermediate-Dose versus 4-Hour High-Dose Methotrexate Infusions for Remission induction in Relapsed Childhood acute Lymphoblastic Leukaemia *Journal of Clinical Oncology* **11 no 5 (May)** 827-833 42

Zintl F, Plenert W, Malke H, Univ of Jena (1987) Results of Acute Lmphoblastic Leukemia Therapy in Childhood with a Modified BFM Protocol in a Multicenter Study in the German Democratic Republic *Haematology and Blood Transfusion* Buchner, Schellong, Hiddemann, Urbanitz, Ritter (eds.) **30** pp 471-479 Springer-Verlag: Berlin, Heidelberg 254

Zintl F, Malke H, Reimann M, Dorffel W, Domula M, Eggers G, Exadaktylos P, Kotte W, Krause I, Kunert W, Mittler U, Mobius D, Reddemann H, Weinmann G, Weissbach G, GDR Hematology and Oncology Working Group, Department of Pediatrics, University of Jena (1990) Results of Acute Lymphoblastic Leukemia Therapy in Childhood: GDR Experiences 1981-1987 *Haematology and Blood Transfusion* Buchner, Schellong, Hiddemann, Urbanitz, Ritter (eds.) **Acute Leukemias II 33** pp 478-482 Springer-Verlag: Berlin, Heidelberg 255

Zintl F, Malke H, Reimann M, Hermann J, Domula M, Dorffel W, Eggers G, Exadaktylos P, Hilgenfeld E, Kotte W, Krause I, Kunert W, Mittler U, Mobius D, Reddemann H, Weissbach G, Weinmann G (1992) Erfahrungen mit moodifizierten BFM-Protokollen bei der Behandlung von Kindern mit akuten lymphoblastischen Leukamien (ALL) in den Ostdeutschen Landern von 1981 bis 1991 (Experiences with Modified BFM-Protocols in Children with Acute Lymphoblastic Leukemia (ALL) in East German States from 1981 to 1991) *Klinische Padiatrie* **204** 221-229 256

II Articles read with the intention of using as a data source but later excluded

Note: The publication ID used in this project is given after the reference.

Aur RJA, Simone J, Hustu HO, Walters T, Borella L, Pratt C, Pinkel D, SJCRH (1971) Central Nervous System Therapy and Combination Chemotherapy of Childhood Lymphocytic Leukemia *Blood* **37** 272-281 240

Bettoni C, Reiter A, Schrappe M, Ludwig W-D, Beck J, Graf N, Ludwig R, Ebell W, Riehm H (1992) Improved Event-free Survival of Childhood T-cell Acute Lymphoblastic Leukemia (T-ALL) after introduction of high dose methotrexate in Multicentre trial ALL-BFM86. *Medical and Pediatric Oncology* **20** 369-488 39

Bleyer W Archie (1983) Central Nervous System Leukemia *Leukaemia, 4th edition (William Dameshek and Frederick Gunz's)* Chapter 35 893-894 Grune and Stratton (subsidiary of Harcourt Brace Jovanovich Publishers): New York 149

Bodey GP, Coltman CA, Freireich EJ, Bonnett JD, Gehan EA, Haut AB, Hewlett JS, McCredit KB, Saiki JH, Wilson HE, Southwest Cancer Chemotherapy Group, Houston (1974) Chemotherapy of Acute Leukemia: Comparison of Cytarabine Alone and in Combination with Vincristine, Prednisolone & Cyclophosphamide *Archives of Internal Medicine* **133** 260-266 248

Cancer and Leukaemia Group B: Henderson ES, Scharlau C, Cooper MR, Haurani FI, Silver RT, Brunner K, Carey RW, Falkson G, Blom J, Nawabi IV, Levine AS, Bank A, Cuttner J, Cornwell III GG, Henry P, Nissen NI, Wiernik PH, Leone L, Wohl H, Rai K, James GW, Weinberg V, Glidewell O, Holland JF (1979) Combination Chemotherapy and Radiotherapy for Acute Lymphocytic Leukemia in Adults: Results of CALGB Protocol 7113 *Leukaemia Research* **3 no 6** 395-407 70

Crist W, Boyett J, Pullen J, van Eys J, Vietti T (1986) Clinical and Biologic Features Predict Poor Prognosis in Acute Lymphoid Leukemias in Children and Adolescents: A Pediatric Oncology Group Review *Medical and Pediatric Oncology* **14** 135-139 178

Crist W, Pullen J, Boyett J, Falletta J, van Eys J, Borowitz M, Jackson J, Dowell B, Russell C, Quddus F, Ragab A, Vietti T (1988b) Acute Lymphoid Leukemia in Adolescents: Clinical and Biologic Features Predict a Poor Prognosis - A Pediatric Oncology Group Study *Journal of Clinical Oncology* **6 no 1** 34-43 180

Cuttner J, Mick R, Budman DR, Mayer RJ, Lee EJ, Henderson ES, Weiss RB, Paciucci PA, Sobol R, Davey F, Bloomfield C, Schiffer C (1991) Phase III Trial of Brief Intensive treatment of Adult Acute Lymphocytic Leukemia Comparing Daunorubicin and Mitoxantrone. A CALGB Study. *Leukemia* **5 no 5 (May)** 425-431 68

Darbyshire PJ, Pinkerton CR, Stevens RF, Oakhill A (1990) Treatment of acute lymphoblastic leukaemia after relapse *Lancet* **335** 733 26

Durrant IJ, Prentice HG, Richards SM for Medical Reseach Council Working Party on Leukaemia in Adults (1997) Intensification of treatment for adults with acute lymphoblastic leukaemia: results of U.K. Medical Research Council randomized trial UKALL XA *British Journal of Haematology* **99** 84-92 25

Ellison RR, Mick R, Cuttner J, Schiffer C, Sobol R for Cancer and Leukemia Group B, Brookline MA (1986) Prognostic factors affecting response and survival in adults with Acute Lymphocytic Leukemia (ALL) treated on CALGB 8011 *Proceedings of the American Society of Clinical Oncology (ASCO)* **5** 156 no 609 66

Ellison RR, Mick R, Cuttner J, Schiffer CA, Silver RT, Henderson ES, Woliver T, Royston I, Davey FR, Glicksman AS, Bloomfield CD (1991) The Effects of Post induction Intensification Treatment with Cytarabine and Daunorubicin in Adult Acute Lymphocytic Leukaemia: A Prospective Randomized Clinical Trial by Cancer and Leukemia Group B *Journal of Clinical Oncology* **9 no 11 (November)** 2002-2015 67

Enno A, Darrell J, Hows J, Catovsky D, Goldman JM, Galton DAG (1978) Co-trimoxazole for prevention of infection in acute leukaemia *Lancet* **ii** 395-7 27

Gottlieb AJ, Weinberg V, Ellison RR, Henderson ES, Terebelo H, Rafla S, Cuttner J, Silver RT, Carey RW, Levy RN, Hutchinson JL, Raich P, Cooper MR, Wiernik P, Anderson JR,Holland JF (1984) Efficacy of Daunorubicin in the Therapy of Adult Acute Lymphocytic Leukaemia: A Prospective Randomized Trial by Cancer and Leukaemia Group B *Blood* **64 Issue 1** 267-274 65

Hammond GD, Sather H, Bleyer WA, Coccia P (1987) Stratification of Prognostic Factors in the Design and Analysis of Clinical Trials for Acute Lymphoblastic Leukemia *Haematology and Blood Transfusion* Buchner, Schellong, Hiddemann, Urbanitz, Ritter (eds.) **30** pp 160-166 Springer-Verlag: Berlin, Heidelberg 109

Iarussi D, Auricchi U, Agretto A, Murano A, Giuliano M, Casale F, Indolfi P, Iacono A, Medical-Surgery Institute of Cardiology and Dept of Pediatrics, 2nd Univ of Naples, Italy (1994) Protective Effect of Coenzyme Q10 on Anthracycline Cardiotoxicity: Control Study in Children with Acute Lymphoblastic Leukemia and Non-Hodgkin's Lymphoma *Molecular Aspects Medicine* **15 (suppl)** S207-S212 275

Jacobs P, Wood L, Novitsky N (1990) Treatment of Adult Acute Lymphoblastic Leukaemia *Haematology and Blood Transfusion* **33 Acute Leukaemias II Prognostic Factors and Treatment Strategies** Buchner, Schellong, Hiddemann, Ritter (eds.) pp428-431 Springer-Verlag: Berlin, Heidelberg 73

Jacobs P, Wood L (1992) Treatment of Acute Lymphoblastic Leukaemia (ALL) *European Journal of Haematology* **49** 53-58 74

Levine AS, Siegel SE, Schreiber AD, Hauser J, Preisler H, Gildstein IM, Seidler F, Simon R, Perry S, Bennett JE, Henderson ES (1973) Protected Environments and Prophylactic Antibiotics. A Prospective Controlled Study of their Utility in the Therapy of Acute Leukemia *N England J Med* **288 no 10** 477-483 173

Mahoney DH Jr, Shuster J, Nitschke R, Lauer SJ, Winick N, Steuber CP, Camitta B. (1998) Intermediate-Dose Intravenous Methotrexate with Intravenous Mercaptopurine for Children with Lower-Risk B-Lineage Acute Lymphoblastic Leukaemia: A Pediatric Oncology Group Phase III Trial. *Journal of Clinical Oncology* **16 no 1 (Jan)** 246-254 278

Maschmeyer G, EORTC (1989) Ciprofloxacin versus Cotrimoxazole/Colistin for Infection Prevention in Acute Leukemia (Meeting Abstract) *Journal of Chemotherapy Infectious Disease Malignancies* **suppl 1** a84. 49 140

Medical Research Council Working Party (1963) Treatment of acute leukaemia in adults: Comparison of steroid therapy at high and low dosage in conjunction with 6-mercaptopurine. First report to the Medical Research Council of the Working Party on the evaluation of different methods of therapy in leukaemia. *British Medical Journal* **1** 7-14 7

Medical Research Council Working Party (1966) Treatment of Acute Leukaemia in Adults: Comparison of Steroid and Mercaptopurine Therapy, Alone and in Conjunction. Second report to the Medical Research Council of the Working Party on the evaluation of Different Methods of Therapy in Leukaemia *British Medical Journal* **2** 1383-1389 6

Petersdorf S, Kopecky KJ, Head D, Boldt D, Appelbaum F, SWOG (1993) A Comparison of Two Consolidation Regimens in the Treatment of Adult Acute Lymphoblastic Leukemia (ALL): The Results of the Southwest Oncology Group Studies (SWOG) 8417/19 *Blood* **82** 193a a759 249

Reiter A, Sauter S, Kabisch H, Ritter J, Harbott J, Gadner H, Reihm H (1989) Probability for Cure as Related to Therapy in Childhood B-Type Acute Lymphoblastic Leukemia (B-ALL) in Three Consecutive BFM Trials *Medical and Pediatric Oncology: International Society of Pediatric Oncology (SIOP) XXI Meeting-Abstracts* **17** 321 abstract 167 49

Riehm H, Reiter A, Schrappe M, Berthold F, Dopfer R, Gerein V, Ludwig R, Ritter J, Stollmann B, Henze G (1986) The in vivo response on cortico steroid therapy as an additional prognostic factor in childhood acute lymphoblastic leukaemia (therapy study ALL-BFM 83. (Die Corticosteroid-abhangige Dezimierung der Leukamiezellzahl im Blut als prognosefaktor bei der akuten lymphoblastischen Leukamie im Kindesalter (Therapiestudie ALL-BFM 83). *Klinische Padiatrie* **199** 151-160 36

Risseeuw-Appel IM, Dekker I, Hop WCJ, Hahlen K (1994) Minimal Effects of E. coli and Erwinia Asparaginase on the Coagulation System in Childhood Acute Lymphoblastic Leukemia: A Randomized Study *Medical and Pediatric Oncology* **23** 335-343 274

Rivera GK, Crist WM, Hancock ML, Simone JV, St Jude Childrens Research Hospital and the University of Tennessee College of Medicine, Memphis TN 38105 (1989) Curing children with Acute Lymphocytic Leukemia (ALL): Twenty-five years of experience in St Jude total therapy studies *Proceedings of the American Society of Clinical Oncology (ASCO)* **8 (Mar 89)** 212 a822 208

Sackmann Muriel F, Pavlovsky S, Penalver JA, Hidalgo G, Bonesana AC, Eppinger-Helft M, de Macchi GH, Pavlovsky A (1974) Evaluation of Induction of Remission, Intensification and Central Nervous System Prophylactic Treatment in Acute Lymphoblastic Leukemia *Cancer* **34 (2)** 418-426 151

Willoughby MLN on behalf of the MRC Working Party for Childhood Leukemia (1974) Treatment of Overt Meningeal Leukaemia *Lancet* **1** 363 202

Yang CP, Lin ST, Liang DC, Hung IJ, Yang YM, Chen HN, Hsieh YL, Law KL, Lin MT, Twu BH, Lin KH, Tsai YM, Peng CT, Chen PM (1993) Treatment of Childhood Acute Lymphoblastic Leukemia with Protocol TCL-842 in Taiwan: The Taiwan Children's Cancer Study Group *Journal of the Formosan Medical Association* **92 no 5** 431-439 271

III Articles eligible for use as a data source but found after the cut-off date for analysis

Freeman AI, Boyett JM, Glicksman AS, Brecher ML, Leventhal BG, Sinks LF, Holland JF (1997) Intermediate-Dose Methotrexate Versus Cranial Irradiation in Childhood Acute Lymphoblastic Leukemia: A Ten-Year Follow-Up *Medical and Pediatric Oncology* **28** 98-107

Gelber RD, Sallan SE, Cohen HJ, Donnelly M, Dalton V, Tobia F, Clavell LA, Tarbell NJ (1993) Central Nervous System Treatment in Childhood Acute Lymphoblastic Leukemia Long-term Follow-up of Patients Diagnosed between 1973 and 1985 *Cancer* **72 (1)** 261-70

Koizumi S, Fujimoto T, Oka T, Watanabe S, Kikuta A, Tsuchiya T, Matsushita T, Asami K, Yanase T, Mimaya J, Ohta S, Miyake M, Nishikawa K, Furuyama T, Yamamura Y, Takaue Y, Ninomiya T, Shimokawa T, Iwai A, Ishida Y, Ariyoshi N, Kimura K, Kawakami K, Gushiken T, Sekine I The Children's Cancer and Leukemia Study Group (CCLSG), Japan (1997) Overview of clinical studies of childhood acute lymphoblastic leukemia for more than ten years by the Japanese Children's Cancer and Leukemia Study Group *Pediatric Hematology and Oncology* **14** 17-28

Miller DR (1981) Childhood Leukemias *Cancer Achievements, Challenges, and Prospects for the 1980s* **vol 2** Burchenal JH, Oettgen HF (eds.) pp 319-330 Grune & Stratton (A Subsidiary of Harcourt Brace Jovanovich, Publishers): New York

Miller DR, Leikin S, Albo V, Sather H, Karon M, Hammond D (1981) Intensive Therapy and Prognostic Factors in Acute Lymphoblastic Leukemia of Childhood: CCG 141 A report from the Children's Cancer Study Group *Haematology and Blood Transfusion* **Vol. 26 Modern Trends in Human Leukemia IV** Neth, Gallo, Graf, Mannweiler, Winkler (eds.) pp 77-86 Springer-Verlag: Berlin Heidelberg

Nishimura K (1993) Progress in the treatment of childhood acute lymphoblastic leukemia (ALL): The results of the Tokyo Children's Cancer Study Group (TCCSG) Studies *Cancer Chemotherapy: Challenges for the Future* **8** 268-281Excerpta Medica Ltd: Tokyo

Schrappe M, Reiter A, Henze G, Niemeyer C, Bode U, Kuhl J, Gadner H, Havers W, Pluss H, Kornhuber B, Zintl F, Ritter J, Urban C, Niethammer D, Riehm H, for the ALL-BFM study group. (1998) Prevention of CNS recurrence in childhood ALL: Results with reduced radiotherapy combined with CNS-directed chemotherapy in four consecutive ALL-BFM trials *Klinische Padiatrie* **210** 192-199

IV References obtained from the literature search

Cheng K, Preston C, Ashby D, O'Hea U, Smyth RL (1998): Time to publication as full reports of abstracts of randomized controlled trials in cystic fibrosis *Pediatric Pulmonology* **26 (2)** 101-105

Chew FS (1991): Fate of manuscripts rejected for publication in the AJR *American Journal of Roentgenology* **156 (3)** 627-632

Dickersin K, Min YI (1993): NIH clinical trials and publication bias *Online Journal of Current Clinical Trials* Doc No 50

Handysides S (1996): CDR Review's editorial process: a survey of papers published in 1995 *Communicable Disease Report CDR Review* **6 (12)** R176-R178

Ioannidis JPA (1998): Effect of the Statistical Significance of Results on the Time to Completion and Publication of Randomized Efficacy Trials *Journal of the American Medical Association* **279 (4)** 281-286

Liebeskind DS, Kidwell CS, Saver JL (1999): Empiric evidence of publication bias affecting acute stroke clinical trials *Stroke* **30 (1)** 92

Misakian AL, Bero LA (1998): Publication Bias and Research on Passive Smoking. Comparison of Published and Unpublished Studies *Journal of the American Medical Association* **280 (3)** 250-253

Stern JM, Simes RJ (1997): Publication bias: evidence of delayed publication in a cohort study of clinical research projects *British Medical Journal* **315** 640-5

V Other references

Burrett JA, Clarke MJ (2002) A descriptive study of randomized trials of treatments for childhood acute lymphoblastic leukaemia *British Journal of Haematology* **118** 986-990

Childhood ALL Collaborative Group (1996) Duration and intensity of maintenance chemotherapy in acute lymphoblastic leukaemia: overview of 42 trials involving 12 000 randomized children. *Lancet* **347** 1783-1788

Childhood ALL Collaborative Group (2001): *International Register of Current Randomized Trials in Childhood Acute Lymphoblastic Leukaemia*. Clinical Trial Service Unit: Oxford

Clarke M and Halsey J (2001) DICE 2: A further investigation of the effects of chance in life, death and subgroup analyses *International Journal of Clinical Practice* **55 (4)** 240-242

Clarke MJ, Stewart LA (1994) Obtaining data from randomized controlled trials: how much do we need to perform reliable and informative meta-analyses? *British Medical Journal* **309** 1007-1010

The Cochrane Library 2002/3, National Electronic Library for Health
www.nelh.nhs.uk/cochrane.asp

Counsell CE, Clarke MJ, Slattery J, Sandercock PAG (1994) The miracle of DICE therapy for acute stroke: fact or fictional product of subgroup analysis? *British Medical Journal* **309** 1677-1681

Diggle PJ, Liang K-Y, Zeger SL (1994) *Analysis of longitudinal data* Clarendon press: Oxford

Dobson AJ (1990) *An introduction to generalized linear models* Chapman and Hall: London

Djulbegovic B, Clarke M (2001) Scientific and ethical issues in equivalence trials *Journal of the American Medical Association* **285 (9)** 1206-1208

Early Breast Cancer Trialists' Collaborative Group (1990) *Treatment of Early Breast Cancer. Volume 1. Worldwide Evidence 1985-1990.* Oxford University Press: Oxford

Institute for Scientific Information: Journal Citation Reports 1995 Science Edition http://wos.mimas.ac.uk/jcrweb/

Juni P, Altman DG, Egger M (2001) Assessing the quality of randomized trials. *Systematic Reviews in Health Care* (eds. Egger M, Davey Smith G, and Altman DG) pp 87-108 British Medical Journal Publications: London

Kish L (1965) *Survey Sampling* Wiley: New York

McCullagh P, Nelder JA (1986) *Generalized Linear Models*, 2nd edn., Chapman and Hall: London

McNeil D (1996) *Epidemiological Research Methods* John Wiley & Sons: Chichester

Machin D, Campbell MJ (1987) *Statistical Tables for the Design of Clinical Trials* Blackwell Scientific Publications: Oxford

Manly BFR (1986) *Multivariate Statistical Methods A Primer* Chapman & Hall: London

Rao JNK, Scott AJ (1992) A Simple Method for the Analysis of Clustered Binary Data *Biometrics* **48** 577-585

Stewart LA, Clarke MJ (1995) Practical Methodology of Meta-analyses (Overviews) using Updated Individual Patient Data *Statistics in Medicine* **14** 2057-2079

Sutton AJ, Song F, Gilbody SM, Abrams KR (2000) Modelling publication bias in meta-analyses: a review *Statistical Methods in Medical Research* **9** 421-445

Zeger SL, Liang K (1986) Longitudinal data analysis for discrete and continuous outcomes *Biometrics* **42** 121-130

APPENDICES

APPENDIX I: SUMMARY OF TRIALS, RANDOMIZATIONS AND ARTICLES USED AS A SOURCE OF DATA

Trial group number and name	Trial	Rand	Question	Accrual period	Size	No. Articles	Publication IDs (see key)
11 AIEOP, Italy							
AIEOP 7901/7902	1101	1101	duration	early 81 – summer 83	177	1	30
AIEOP 7903	1102	1102	combination	early 79 - summer 81	210	1	30
Second CR	1103	1103	post-relapse	May 80 - Jun 83	84	1	31
12 BFM – children Group, Germany							
ALL-BFM-76	1201	1201	intensification	Oct 76 - Mar 79	158	4	33 32 37 48
ALL-BFM-79	1202	1202	intensification	1 Apr 79 – Mar 81	199	4	34 32 37 48
ALL-BFM-79	1203	1203	intensification	1 Apr 79 – 31 Mar 81	126	4	34 32 37 48
ALL-BFM-81	1204	1204	duration	1 Apr 81 – 30 Sep 83	395	7	35 34 32 37 46 48 38
ALL-BFM-81	1205	1205	CNS prophylaxis	1 Apr 81 – 30 Sep 83	277	6	35 32 46 47 48 38
ALL-BFM-83	1206	1206	duration	1983 – Sep 86	351	6	35 32 41 47 48 281
ALL-BFM-83	1206	1207	intensification	1983 – Sep 86	126	3	32 41 46
ALL-BFM-83	1206	1208	CNS prophylaxis	late 1983- Sep 86	143	4	32 46 47 48
ALL-REZ-BFM-85	1209	1209	post-relapse	Apr 85 – 31 Mar 86	46	4	40 42 43 44
ALL-BFM-??	1210	1210	immunotherapy	1982 - ?	?	0	
ALL-BFM-86	1211	1211	intensification	Oct 86 – Mar 90	128	1	45
ALL-REZ-BFM-87	1215	1215	post-relapse	spring 87 – end 88	41	0	
13 Bombay, India							
Bombay TAT TMC 1	1301	1301	maintenance	Jul 79 – Dec 80	?	0	
14 Cancer and Leukemia Group B (CALGB), USA							
ALGB 1965	1401	1401	maintenance	1960s	116	1	50
ALGB 6601	1402	1402	intensification	1966 - 68	267	3	51 52 53
ALGB 6601	1402	1403	maintenance	1966 - 68	211	3	51 52 53
ALGB 6601	1402	1499	intensification	1966 - ?	?	2	52 53
ALGB 6801	1404	1404	induction	Feb 68 – Feb 71	514	5	51 52 53 55 282
ALGB 6801	1404	1405	intensification	Feb 68 – Feb 71	225	7	51 52 53 54 55 60 282
ALGB 6801	1404	1406	maintenance	Feb 68 – Feb 71	339	5	51 52 53 55 282
ALGB 6801	1404	1498	duration	1973 - ?	?	1	53

CLB 7111	1407	1407	induction	5 Feb 71 – 18 Mar 74	646	5	52 56 57 61 5
CLB 7111	1407	1408	CNS prophylaxis	5 Feb 71 – 18 Mar 74	467	7	52 54 56 57 60 58 61
CLB 7111	1407	1409	maintenance	12 Jul 71 – Mar 74	493	4	52 56 58 61
CLB 7112 relapse	1410	1410	post-relapse	1971 – 71	88	2	57 59
CLB 7112 relapse	1410	1411	post-relapse	1971 – 71	45	2	57 59
CLB 7211 relapse	1412	1412	post-relapse	1972 – 72	135	2	57 59
CLB 7211 relapse	1412	1413	post-relapse	1972 – 72	91	2	57 59
CLB 7411	1414	1414	CNS prophylaxis	1974 –77	339	3	54 57 60
CLB 7411	1414	1415	maintenance	1974 – 77	339	1	57
CLB 7611	1416	1416	CNS prophylaxis	12 Nov 76 – late 79	506	8	54 57 60 62 63 64 72 280
CLB 7611R	1417	1417	intensification	Jan 80 – 81	?	1	72
CLB 7811, relapse	1418	1418	post-relapse	Dec 78 – Nov 79	?	0	
CLB 6911 4-way	1423	1423	post-relapse	Mar 69 – Feb 70	?	2	69 56
CLB 6911 4-way	1423	1424	post-relapse	Feb 70 – Sep 70	?	2	69 56
CLB 6911 4-way	1423	1425	maintenance	Mar 69 – Sep 70	87	1	69
ALGB ????	1428	1428	post-relapse	1960s	57	1	71
ALGB ????	1428	1497	maintenance	1960s	19	1	71
CALGB 6611	1429	1429	unknown	1966 - ?	96	1	288
CALGB 6611	1429	1496	maintenance	1966 - ?	?	1	288

16 Children's Cancer Group (CCG), USA

CCG-101	1601	1601	CNS prophylaxis	Jun 72 – summer 74	590	9	111 112 113 114 115 116 117 118 119
CCG-101	1601	1602	duration	autumn 75 - summer 78	244	6	75 80 111 112 115 116
CCG-105	1603	1603	combination	May 83 – Apr 89	1606	4	78 82 84 85
CCG-105	1603	1604	CNS prophylaxis	May 83 – Apr 89	1389	6	78 79 81 82 84 85
CCG-106 - 3 arm	1605	1605	combination	May 83 – Nov 84	214	5	86 87 88 89 90
CCG-106 - 3 arm	1605	1606	combination	8 Nov 84 – Mar 87	328	4	86 87 88 89
CCG-121	1607	1607	post-relapse	Jan 82 – Dec 85	119	0	
CCG-123 - first stratum 3-way	1608	1608	combination	Apr 83 – Nov 85	260	5	89 91 92 93 276
CCG-123 - first stratum 3-way	1608	1609	combination	Nov 85 – Dec 85	14	4	93 91 92 276
CCG-123 - first stratum 3-way	1608	1610	combination	Dec 85 – Apr 87	183	4	93 91 92 276
CCG-123 - first stratum 3-way	1608	1635	combination	Apr 87 – Apr 89	209	4	93 91 92 276
CCG-123 - first stratum 3-way	1608	1636	combination	Apr 83 – Oct 85	272	4	93 91 92 276
CCG-139	1611	1611	combination	autumn 83 – Jan 89	148	2	94 95
CCG-141	1612	1612	combination	Feb 75 – spring 77	306	5	77 89 96 108 114
CCG-141	1612	1615	maintenance	Mar 78 – Jun 80	287	6	77 89 96 97 98 108
CCG-141	1612	1698	CNS prophylaxis	1975 - ?	?	1	97
CCG-141A I	1613	1613	intensification	spring 77 – summer 78	366	2	77 97

CCG-141A I	1613	1614	maintenance	spring 78 – summer	301	1	108
CCG-143	1616	1616	CNS prophylaxis	Aug 74 – spring 75	179	3	111 112 113
CCG-143	1617	1617	duration	autumn 77 – summer 78	72	4	75 80 111 112
CCG-161	1618	1618	CNS prophylaxis	1 Apr 78 – autumn 82	529	4	97 101 99 100
CCG-161	1618	1619	maintenance	1 Apr 78 – May 83	625	5	97 101 99 100 108
CCG-161	1618	1621	duration	1980 – 85	285	8	77 90 100 101 102 103 99 75
CCG-162	1622	1622	maintenance	1 Apr 78 – autumn 81	1058	3	97 101 110
CCG-162	1622	1624	duration	1980 – Nov 84	716	5	90 101 102 103 75
CCG-162A	1623	1623	maintenance	Jun 81 – Feb 83	541	0	
CCG-163D	1625	1625	maintenance	1 Apr 78 – Apr 81	321	4	89 101 102 97
CCG-163D	1625	1626	duration	1981 – 85	120	5	90 101 102 103 75
CCG-171	1627	1627	post-relapse	Apr 79 – Feb 82	97	0	
CCG - relapse	1633	1633	post-relapse	1970s	164	1	104
CCG - relapse	1633	1640	post-relapse	1970s	50	1	104
Relapsed Patients in New CR	1634	1634	post-relapse	1980s (?)	50	1	105
CCG 1970a	1637	1637	maintenance	1970 – 71	350	1	106
CCG 1970a	1637	1638	maintenance	1970 – 71	153	1	106
CCG-098	1639	1639	post-relapse	May 86 - Mar 88	49	1	120
CCG - (1967)	1645	1645	duration	1967 – 1967	15	1	107
CCG - (1967)	1645	1699	unknown	1963 – 67	165	1	107
CCG-144	1652	1652	unknown	1984 – 88	?	0	
17 Dana Farber Cancer Institute (DFCI), USA							
DFCI 73001	1701	1701	induction	May 73 – Dec 74	45	3	121 122 126
DFCI 77001	1702	1702	intensification	1 Jun 77 – Oct 79	64	2	123 126
DFCI 80001	1703	1703	intensification	1 Jan 80 – 31 Dec 80	22	1	126
DFCI 81001pre-induction	1704	1704	induction	May 81 – Dec 83	77	6	124 126 127 128 129 38
DFCI 85001	1705	1705	induction	spring 85 - Dec 87	210	0	
DFCI 87001 pre-induction	1706	1706	induction	Oct 87 - 91	?	0	
DFCI 87001 pre-induction	1706	1707	maintenance	Nov 87 - summer 91	353	1	130
DFCI 87001 pre-induction	1706	1708	CNS prophylaxis	Nov 87 – Jul 91	109	2	130 131
SFCC ???	1709	1709	CNS prophylaxis	Jul 72 – 197?	20	1	125
18 Dutch Childhood Leukaemia Study Group, The Netherlands							
DCLSG-ALL-V/EORTC 99801	1801	1801	induction	4 May 79 – Dec 82	240	3	132 133 134
19 European Organisation for Research on Treatment of Cancer (EORTC)							
EORTC 58741	1901	1901	intensification	May 71 – Jan 79	224	1	135

EORTC 58741	1901	1902	immunotherapy	autumn 72 – spring 80	123	1	135
EORTC 58791	1903	1903	induction	May 80 – Dec 85	108	3	136 137 138
EORTC 58791	1903	1904	intensification	1980 – 88	82	3	136 137 138
EORTC 58831	1905	1905	intensification	Jul 83 – summer 89	388	1	139
EORTC 58832	1906	1906	CNS prophylaxis	Jul 83 – end 88	191	2	142 143
EORTC ????	1912	1912	miscellaneous	1980s (?)	51	1	141
20 French ALL Cooperative Group (FRALLE) / French Society of Pediatric Hematology (SHIP), France							
FRALLE 87	2004	2004	CNS prophylaxis	summer 87- early 89	188	1	147
FRALLE 83 LR	2005	2001	duration	Jun 83 - Apr 87	66	2	144 145
FRALLE 83 LR	2005	2005	induction	Jun 83 – Apr 87	113	1	145
FRALLE 83 IR	2006	2002	combination	1983 – 87	250	1	146
FRALLE 83 IR	2006	2003	testicular XRT	1983 – 86	63	1	144
FRALLE 83 IR	2006	2006	induction	1983 - 87	?	0	
21 GATLA, Argentina							
GATLA 72	2101	2101	intensification	Oct 72 – Dec 75	465	2	148 284
GATLA 72	2101	2102	maintenance	Oct 72 – Dec 75	465	2	148 284
GATLA 72	2101	2103	immunotherapy	1972 - 75	?	1	148
GATLA 1 LLA-79	2104	2104	immunotherapy	Jan 79 – Jan 84	602	1	284
GATLA 7 LLA-87	2105	2105	maintenance	Jul 87 – Dec 89	272	2	150 284
Protocol 1-ALL-76	2198	2197	immunotherapy	Jan 76 – Dec 78	336	1	284
Protocol 1-ALL-76	2198	2198	maintenance	Jan 76 – Dec 78	336	1	284
Protocol 11-ALL-67	2199	2199	maintenance	Nov 67 – Sep 70	38	1	284
22 Great Ormond Street Hospital, London, UK							
GOS-79	2201	2201	maintenance	Jun 79 - Dec 82	144	1	152
23 Gunma University, Japan							
Japan KLSG 1	2301	2301	immunotherapy	Mar 79 – Jul 81	15	1	153
24 Cooperative Acute Lymphoblastic Leukaemia Group (COALL), Germany							
COALL 80	2401	2401	maintenance	Nov 78 – Nov 82	105	2	155 154
COALL 80	2401	2402	maintenance	Nov 78 – Nov 82	39	2	154 155
COALL 85/89	2403	2403	intensification	1 Jan 85 – end 87	143	4	156 157 158 159
COALL 85/89	2403	2404	intensification	1 Jan 85 – spring 90	58	4	157 158 159 156
COALL 82	2405	2405	intensification	1982 – Nov 83	?	1	155

25 Instituto Nacional de Enfermedades Neoplasicas (INEN), Peru

INEN-7902	2502	2502	maintenance	Jan 79 – summer 83	48	0	
INEN-P83	2503	2503	CNS prophylaxis	Jan 83 – summer 85	59	0	
INEN-P85	2504	2504	CNS prophylaxis	summer 85 – summer 88	73	0	

26 Japanese Children's Cancer and Leukaemia Group, Japan

JCCLSG S-811	2601	2601	maintenance	late 80 – early 84	115	6	160 161 162 163 166 83
JCCLSG H-811	2602	2602	maintenance	early 81 – summer 82	88	1	83
JCCLSG I-841	2603	2603	combination	early 84 - spring 87	78	2	164 83
JCCLSG I-841	2603	2604	induction	early 84 – summer 87	66	3	164 166 83
JCCLSG H-851	2605	2605	maintenance	late 84 – summer 87	93	1	83
JCCLSG 874	2606	2606	CNS prophylaxis	April 87 – early 91	87	3	166 76 83
JCCLSG 874	2606	2607	CNS prophylaxis	April 87 – autumn 91	114	4	165 166 76 83
JCCLSG 874	2606	2608	intensification	April 87 – Dec 90	164	3	166 76 83

27 Medical Research Council, UK

UKALL I	2701	2701	CNS prophylaxis	Aug 70 - Jan 72	26	4	1 12 4 5
UKALL I	2701	2702	duration	start 72 - Autumn 73	82	4	15 4 1 2
UKALL II	2703	2703	CNS prophylaxis	1 Jan 72 – Dec 72	39	2	12 3
UKALL II	2703	2704	maintenance	1 Jan 72 – Mar 73	185	2	12 3
UKALL II	2703	2705	duration	Autumn 73 – summer 75	207	3	15 12 3
UKALL III	2706	2706	maintenance	1 Sep 73 – 12 Nov 74	135	1	16
UKALL III	2706	2707	maintenance	autumn 74 – Dec 75	110	1	16
UKALL III	2706	2708	duration	summer 75 - Dec 77	249	2	15 16
UKALL IV	2709	2709	induction	start 75 – start 78	431	1	18
UKALL IV	2709	2710	maintenance	early 75 – Dec 77	336	1	18
UKALL V	2711	2711	CNS prophylaxis	Jan 76 – summer 79	334	1	9
UKALL V	2711	2712	maintenance	autumn 75 – summer 79	524	4	9 18 15 285
UKALL V	2711	2713	duration	late 77 – summer 81	305	4	9 18 15 285
UKALL VI (children)	2714	2714	intensification	summer 77 – spring 80	219	2	17 18
UKALL VI (children)	2714	2716	CNS prophylaxis	1 Jan 78 – Nov 78	62	0	
UKALL VI (children)	2714	2718	CNS prophylaxis	Nov 78 – spring 80	25	0	
UKALL VI (children)	2714	2720	testicular XRT	1 Jan 78 - spring 80	73	1	17
UKALL VII	2722	2722	induction	1 Mar 79 – Mar 80	82	1	14
UKALL VII	2722	2723	CNS prophylaxis	1 Mar 79 – Mar 80	82	2	13 14
UKALL VII	2722	2724	CNS prophylaxis	1 Mar 79 – Mar 80	82	1	14
UKALL VII	2722	2725	testicular XRT	1 Mar 79 – Mar 80	43	2	14 17
UKALL VII	2722	2726	maintenance	1 Mar 79 – Mar 80	82	2	13 14

UKALL VIII	2727	2727	induction	1 Sep 81 – 31Dec 84	630	5	21 22 8 18 19
UKALL VIII	2727	2728	duration	Feb 83 – early 87	408	6	20 21 22 18 19 8
UKALL X	2730	2730	intensification	Jan 85 – Sep 90	1171	6	23 10 11 18 24 283
MRC meningeal	2740	2740	post-relapse	1970s (?)	38	2	28 29
MRC Concord Trial	2741	2741	intensification	Jan 69 – Aug 70	122	1	277
28 US National Cancer Institute, USA							
NCI-immuno	2801	2801	immunotherapy	1970's (?)	49	1	167
NCI 72-1	2802	2802	CNS prophylaxis	1971 - 74	32	2	168 169
NCI 77-02	2803	2803	combination	Feb 80 – Dec 83	181	3	172 170 171
NCI-82-C-199	2804	2804	maintenance	Nov 82 – Apr 86	?	0	
NCI-84-C-153A/CCG-144	2805	2805	CNS prophylaxis	Jun 84 – Nov 88	181	1	172
29 Pediatric Oncology Group (POG), USA							
POG 7420 / SWOG 7420/ALinC11	2901	2901	combination	10 Sep 74 – 29 Oct 76	408	2	175 176
POG 7623 / SWOG 7623/ALinC12	2902	2902	maintenance	late 76 – end 80	434	1	177
POG 7623 / SWOG 7623/ALinC12	2902	2903	combination	late 76 - late 80	430	1	179
POG 7623 / SWOG 7623/ALinC12	2902	2904	miscellaneous	autumn 80 – summer 81	126	1	181
POG 7837	2905	2905	combination	Apr 79 – Mar 81	59	3	182 183 204
POG 7866	2906	2907	combination	Mar 79 – Sep 79	?	0	
POG 7866	2906	2917	combination	Mar 79 – Feb 80	?	0	
POG 8035 / POG 8036/ALinC13	2908	2908	combination	Jun 81 – Jan 86	1504	5	184 185 186 187 204
POG 8698	2911	2911	intensification	Feb 86 – 1988	20	0	
POG 8704	2913	2913	maintenance	May 87 – Jan 91	363	2	192 204
POG 8710 / SIMAL 5	2914	2914	post-relapse	Dec 87 -	277	1	188
POG 8602 / ALinC 14	2918	2918	intensification	Feb 86 – Aug 86	?	7	192 193 194 189 190 191 204
POG 8602 / ALinC 14	2918	2919	intensification	Aug 86 – May 87	?	7	190 191 192 193 204 189 194
POG 8602 / ALinC 14	2918	2920	intensification	May 87 – Jan 88	?	7	190 191 192 193 204 189 194
POG 8602 / ALinC 14	2918	2922	intensification	Feb 86 – Aug 86	?	7	192 189 191 190 193 194 204
POG 8602 / ALinC 14	2918	2923	intensification	Aug 86 – May 87	428	8	279 190 191 192 193 189 194 204
POG 8602 / ALinC 14	2918	2924	intensification	May 87 – Jan 88	?	7	189 191 192 193 194 190 204
POG 8602 / ALinC 14	2918	2926	intensification	Feb 86 – Aug 86	?	7	192 189 191 190 194 193 204
POG 8602 / ALinC 14	2918	2927	intensification	May 87 – 1991	?	7	193 189 190 191 192 194 204
POG 8304	2928	2928	post-relapse	Apr 83 – Nov 89	104	2	195 196
POG 8303	2929	2929	post-relapse	Apr 83 – Dec 87	258	1	197
POG 7834	2930	2930	post-relapse	Jan 79 – Apr 83	113	1	205
POG 7712	2933	2933	post-relapse	Jun 78 – Nov 82	87	2	174 198
POG 7818	2934	2934	post-relapse	Aug 79 – Aug 81	67	1	199

POG 7919	2935	2935	post-relapse	Jul 79 – Feb 82	74	1	200
POG 8022	2936	2936	post-relapse	Mar 81 – Jul 83	43	1	206
POG 8594	2939	2939	post-relapse	Oct 85 – Dec 87	19	1	201
POG CNS 1	2952	2952	CNS prophylaxis	1960s (?)	31	1	203
POG CNS 2	2953	2953	CNS prophylaxis	1960s (?)	31	1	203
30 St Joseph's Hospital, Phoenix, USA							
NCI-D79-053-088	3001	3001	post-relapse	Mar 79 – Jan 80	?	0	
31 St Jude Children's Research Hospital, Memphis, USA							
St Jude IV	3101	3101	maintenance	summer 65 – summer 67	42	3	207 209 210
St Jude VI	3103	3102	CNS prophylaxis	Jul 68 – May 70	94	7	209 210 214 215 211 212 213
St Jude VI	3103	3103	intensification	Jul 68 – May 70	94	4	209 210 215 213
St Jude VI	3103	3124	CNS prophylaxis	Jul 68 – summer 70	49	7	209 210 211 212 213 214 215
St Jude VII	3104	3104	CNS prophylaxis	May 70 – early 72	94	5	209 214 215 216 210
St Jude VII	3104	3105	maintenance	May 70 – early 72	94	4	209 215 216 210
St Jude VII	3104	3125	CNS prophylaxis	May 70 – early 72	47	5	209 210 214 215 216
SJCRH VIII	3106	3106	maintenance	Jan 72 – May 73	79	6	209 215 217 218 210 219
SJCRH VIII	3106	3113	maintenance	May 73 – Nov 75	149	3	215 218 219
SJCRH X	3107	3107	combination	May 79 – Jan 84	309	13	220 221 222 223 224 225 226 227 228 229 230 231 235
SJCRH XI	3108	3108	maintenance	Feb 84 – Sep 88	108	6	232 233 234 235 236 237
SJCRH XI	3108	3109	maintenance	Feb 84 – Sep 88	233	6	233 235 236 234 237 232
SJCRH R VIII	3112	3112	post-relapse	Sep 79 - 1982	?	0	
SJCRH TOT IX	3114	3114	combination	autumn 75 – May 79	256	2	227 238
SJCRH TOT IX	3114	3116	combination	late 75 – May 79	27	2	227 238
St Jude V	3115	3115	duration	end 67 – summer 68	20	3	209 210 239
St Jude P. carinii	3117	3117	miscellaneous	Oct 74 – Oct 76	136	1	241
32 Southwest Oncology Group (SWOG), USA							
SWOG 690/691 / ALinC 9	3201	3201	induction	Jul 71 – Mar 73	226	1	242
SWOG 690/691 / ALinC 9	3201	3202	CNS prophylaxis	Jul 71 – Mar 73	194	2	243 244
SWOG 690/691 / ALinC 9	3201	3203	maintenance	Jul 71 – Mar 73	?	1	242
SWOG 7220 / ALinC 10	3204	3204	induction	Feb 73 – Aug 73	73	1	245
SWOG 7220 / ALinC 10	3204	3205	maintenance	Feb 73 – Aug 74	154	1	246
SWOG ????	3206	3206	induction	Jul 67 – Nov 69	19	1	247
SWOG 8612	3214	3214	post-relapse	Dec 86 - ?	?	0	
SWOG 663/664 / ALinC 6	3215	3215	induction	1965 – 67	68	1	250

SWOG 663/664 / ALinC 6	3215	3216	maintenance	Oct 65 – Jun 67	68	1	250
SWOG 7420	3217	3217	CNS prophylaxis	1 Oct 74 – 1978	?	2	251 252
SWOG 7623 / ALinC 12	3218	3218	CNS prophylaxis	1976 – 1 Jun 78	357	2	251 252
33 Wroclaw, Poland							
Poland	3301	3301	immunotherapy	1970s	37	1	253
34 Jena University, Germany							
ALL VII 81	3401	3401	CNS prophylaxis	spring 83 – spring 86	244	2	254 255
ALL VII 81	3401	3403	CNS prophylaxis	spring 86 – early 88	98	1	255
ALL VII 81	3401	3405	unknown	Jul 83 – Dec 83	30	0	
ALL VII 81	3402	3402	maintenance	1 Sep 81 – 31 Dec 87	381	3	254 255 256
35 PETHEMA, Spain							
LAL 7/78	3501	3501	CNS prophylaxis	Apr 78 – Dec 83	65	5	259 257 258 286 287
LAL 7/78	3501	3502	CNS prophylaxis	Apr 78 – Dec 83	22	5	258 259 257 286 287
LAL 17/84	3503	3503	CNS prophylaxis	Oct 83 – early 90	124	3	260 286 287
LAL 17/84	3503	3504	CNS prophylaxis	spring 84 – early 89	130	3	260 286 287
36 Institut de Cancerologie et d'Immunogenetique (INSERM), France							
ICIG-ALL 9	3601	3601	immunotherapy	1970 -	22	2	261 262
ICIG-ALL 10	3602	3602	immunotherapy	1970 -	14	2	261 262
ICIG pilot	3603	3603	immunotherapy	1962 -	30	3	263 264 265
84 Memorial and Sloan Kettering Cancer Centre (MSKCC), USA							
MSK-NY-II	8401	8401	induction	Nov 86 – Feb 91	44	1	266
125 Vienna, St Anna Kinderspital, Austria							
Austrian-BFM-86	12501	12501	maintenance	autumn 86 – spring 89	10	0	
127 Tokyo Children's Cancer Study Group, Japan							
TCSLG L84-11	12701	12701	CNS prophylaxis	spring 84 – Feb 89	190	2	267 268
TCSLG L84-11	12701	12702	CNS prophylaxis	Jun 84 – Feb 89	263	1	267
TCLSG L81-10	12703	12703	CNS prophylaxis	early 80 – spring 84	68	1	267
TCLSG L81-10	12703	12704	combination	1981 – 84		1	267
305 Australasian Childhood Leukaemia Study Group							
ANZCCSG ALL V	30501	30501	maintenance	1986 - 92	600	1	169

338 Brazilian Cooperative ALL Group, Brazil

GBTLI-80	33802	33802	CNS prophylaxis	Jul 80 – Jul 82	203	1	270
GBTLI-82	33803	33803	maintenance	Aug 82 – Jul 85	360	1	270

341 Israel National Study, Israel

INS 84	34101	34101	CNS prophylaxis	autumn 84 – summer 89	75	2	272 273
INS 84	34101	34102	induction	late 84 – spring 87	13	0	

Key

1	Medical Research Council (1975)	35	Schrappe et al (1987)	72	Freeman et al (1992)
2	Medical Research Council (1977)	37	Riehm et al (1984)	75	Nesbit et al (1983)
3	Medical Research Council (1978)	38	Niemeyer et al (1991b)	76	Tsurusawa et al (1999)
4	Campbell et al (1973)	40	Henze et al (1991)	77	Miller et al (1989)
5	Medical Research Council (1973)	41	Henze et al (1990b)	78	Tubergen et al (1990)
8	Lilleyman et al (1986)	42	Wolfrom et al (1993)	79	Gilchrist et al (1990)
9	Chessells et al (1986)	43	Henze et al (1990a)	80	Nesbit et al (1982b)
10	Chessells et al (1992a)	44	Henze et al (1987)	81	Tubergen et al (1988)
11	Chessells et al (1991)	45	Reiter et al (1994)	82	Tubergen et al (1993a)
12	MacLennan et al (1976)	46	Riehm et al (1987)	83	Koizumi and Fujimoto (1994)
13	Chessells et al (1990)	47	Buhrer et al (1990)	84	Tubergen et al (1993b)
14	Lilleyman et al (1985)	48	Ritter et al (1990)	85	MacLean et al (1995)
15	Medical Research Council (1982b)	50	Frei et al (1965)	86	Gaynon et al (1993)
16	Medical Research Council (1982a)	51	Glidewell and Holland (1973)	87	Gaynon et al (1988)
17	Eden et al (1990)	52	Holland and Glidewell (1972a)	88	Gaynon et al (1989)
18	Peto et al (1986)	53	Holland (1978)	89	Gaynon et al (1987)
19	Eden et al (1987)	54	Rowland et al (1984)	90	Bleyer (1989)
20	Clayton et al (1988)	55	Holland and Glidewell (1972b)	91	Steinherz et al (1988)
21	Eden et al (1991)	56	Jones et al (1977)	92	Cherlow et al (1990)
22	Eden et al (1992)	57	Bleyer (1990)	93	Cherlow et al (1993)
23	Chessells et al (1992b)	58	Shuster and Holland (1984)	94	Meadows et al (1991)
24	Chessells et al (1995)	59	Kung et al (1978)	95	Lange et al (1996)
28	Willoughby (1976)	60	Brecher et al (1985)	96	Miller et al (1990)
29	Willoughby (1983)	61	Jones et al (1991)	97	Bleyer et al (1983)
30	Paolucci et al (1989)	62	Nesbit et al (1984)	98	Miller et al (1986)
31	Rossi et al (1986)	63	Freeman et al (1983)	99	Littman et al (1987)
32	Riehm et al (1990)	64	Hill et al (1994)	100	Bleyer et al (1991)
33	Henze et al (1981)	69	Rausen et al (1979)	101	Coccia et al (1984)
34	Henze et al (1982)	71	Jones et al (1972)	102	Bleyer (1987)

103 Coccia et al (1989)
104 Nesbit et al (1979)
105 Henderson et al (1990)
106 Heyn et al (1975)
107 Nesbit et al (1973)
108 Miller et al (1983)
110 Bleyer (1984)
111 Robison et al (1980)
112 Hamre et al (1987)
113 Nesbit et al (1981a)
114 Bleyer et al (1990)
115 Nesbit et al (1980)
116 Nesbit et al (1982a)
117 Robison et al (1984)
118 Ortega et al (1987b)
119 Nesbit et al (1981b)
120 Miser et al (1992)
121 Sallan et al (1977)
122 Hitchcock-Bryan (1986)
123 Sallan et al (1983)
124 Kretschmar et al (1984)
125 Dritschilo et al (1976)
126 Sallan et al (1990)
127 Waber et al (1992)
128 Niemeyer et al (1991a)
129 Niemeyer et al (1987)
130 Waber et al (1995)
131 Tarbell et al (1991)
132 van der Does-van den Berg et al (1988)
133 van der Does-van den Berg et al (1989)
134 van der Does-van den Berg et al (1987)
135 Otten et al (1988b)
136 Stryckmans et al (1986)
137 Stryckmans et al (1985)
138 Stryckmans et al (1987)
139 Rubic et al (1988)
141 Maschmeyer et al (1990)
142 Benoit et al (1988)
143 Otten et al (1988a)
144 Schaison et al (1990b)
145 Leverger et al (1988)
146 Schaison et al (1987)
147 Schaison et al (1990a)
148 Sackmann-Muriel et al (1978)
150 Riccheri et al (1994)
152 Chessells et al (1987)
153 Maekawa et al (1987)
154 Janka-Schaub et al (1986)
155 Janka et al (1987)
156 Janka et al (1991)
157 Janka-Schaub et al (1996)
158 Janka-Schaub et al (1988)
159 Janka-Schaub et al (1990)
160 Koizumi et al (1991)
161 Koizumi et al (1988a)
162 Koizumi et al (1989)
163 Koizumi and Fujimoto (1988b)
164 Children's Cancer and Leukemia Study Group (1989)
165 Nakadate et al (1992)
166 Koizumi et al (1992)
167 Poplack et al (1975)
168 Moss et al (1981)
169 Peylon-Ramu et al (1978)
170 Poplack et al (1984)
171 Brouwers et al (1988)
172 Poplack (1987)
174 Pullen et al (1982)
175 Sullivan et al (1982)
176 Hvizdala et al (1984)
177 van Eys et al (1989b)
179 van Eys et al (1989a)
181 van Eys et al (1987)
182 Shuster et al (1990)
183 Falletta et al (1992)
184 Crist et al (1988)
185 Pullen et al (1988)
186 Crist et al (1989)
187 Pullen et al (1993)
188 Graham-Pole (1989)
189 Crist et al (1991)
190 Land et al (1989)
191 Borowitz (1990)
192 Crist et al (1992)
193 Crist et al (1990a)
194 Shuster et al (1991)
195 Smith et al (1990)
196 Wofford et al (1992)
197 Buchanan et al (1988)
198 Land et al (1985)
199 Ragab et al (1986)
200 Krischer et al (1984)
201 Ochs et al (1991b)
203 Duttera et al (1973)
204 Crist et al (1990b)
205 Gilbert et al (1991)
206 Vats et al (1992)
207 Pinkel et al (1971)
209 Simone (1974)
210 Simone et al (1975)
211 Pinkel et al (1977)
212 Borella (1973)
213 Aur et al (1972)
214 Hustu et al (1973)
215 Mauer and Simone (1976)
216 Aur et al (1973a)
217 Aur et al (1973b)
218 Aur et al (1978)
219 Pinkel (1979)
220 Bowman et al (1984)
221 Ochs et al (1983)
222 Pui et al (1985)
223 Ochs et al (1986)
224 Abromowitch et al (1988b)

225 Abromowitch et al (1988a)
226 Pui et al (1989)
227 Mulhern et al (1991)
228 Williams et al (1991)
229 Ochs et al (1991a)
230 Pui et al (1992)
231 Ochs et al (1989)
232 Rivera et al (1990)
233 Rivera et al (1991b)
234 Rivera et al (1991a)
235 Pui et al (1991)
236 Mahmoud et al (1993)
237 Hughes et al (1987)
238 Pui et al (1984)
239 Aur et al (1974)
241 Hughes et al (1977)
242 Komp et al (1976)
243 Komp et al (1975)
244 Komp et al (1982)
245 Berry et al (1975)
246 Berry et al (1980)
247 Southwest Oncology Group (1974)
250 Fernbach et al (1975)
251 Wells et al (1983)
252 Whitt et al (1984)
253 Boguslawska-Jaworska and Raus (1981)
254 Zintl et al (1987)
255 Zintl et al (1990)
256 Zintl et al (1992)
257 Ortega Aramburu et al (1985)
258 Ortega et al (1987a)
259 Giralt et al (1992)
260 Ortega (1985)
261 Mathe et al (1977b)
262 Mathe et al (1979)
263 Mathe et al (1969)
264 Mathe (1970)
265 Mathe et al (1977a)
266 Steinherz et al (1993)
267 Tokyo Children's Cancer Study Group (1991)
268 Tsukada (1993)
269 Carden et al (1990)
270 Brandalise et al (1993)
272 Stark et al (1992)
273 Stark et al (1996)
276 Steinherz et al (1998)
277 Medical Research Council (1971)
279 Land et al (1994)
280 Freeman et al (1981)
281 Riehm et al (1996)
282 Halazun et al (1974)
283 Richards et al (1998)
284 Sackmann-Muriel et al (1998)
285 Dolan et al (1989)
286 Ortega (1998)
287 Ortega et al (1988)
288 Jones et al (1971)

Abbreviations used

CNS	central nervous system
XRT	radiotherapy

APPENDIX II: TRIAL PROTOCOLS USED AS A DATA SOURCE

In addition to the Cancer Overviews Group's trial listings, which are available for all randomizations, and the articles referenced, additional data were obtained for some randomizations using the electronic trial protocol database, *Clinprot,* and/or the hard copy of the protocol in full.

Trials for which this additional information was available are listed below:

Key
C=clinprot
P=trial protocol in full
E=EORTC Handbook

Trial name	Trial ID	Randomization IDs	Additional data source
Bombay TAT TMC 1	1301	1301	C
CLB 7611	1416	1416	C
CLB 7611R	1417	1417	C
CLB 7811, relapse	1418	1418	C
CCG-105	1603	1603 1604	C
CCG-106 - 3 arm	1605	1605 1606	C
CCG-121	1607	1607	C
CCG-123 - first stratum 3-way	1608	1608 1609 1610 1635 1636	C
CCG-139	1611	1611	C
CCG-161	1618	1618	C
CCG-162	1622	1622 1624	C
CCG-162A	1623	1623	C
CCG-163d	1625	1625 1626	C
CCG-171	1627	1627	C
CCG-098	1639	1639	C
DFCI 81001pre-induction	1704	1704	C
DFCI 85001	1705	1705	C
DFCI 87001 pre-induction	1706	1706 1707 1708	C
DCLSG-ALL-V/EORTC 99801	1801	1801	E
EORTC 58741	1901	1901 1902	C P
EORTC 58791	1903	1903 1904	C
EORTC 58831	1905	1905	C E
EORTC 58832	1906	1906	C E
FRALLE 87	2004	2004	P
GATLA 72 (1 ALL 72)	2101	2101 2102	P
GATLA 1 LLA 79	2104	2104	C P
GATLA 7 LLA-87	2105	2105	P
Japan KLSG 1	2301	2301	C
COALL 80	2401	2401 2402	C
INEN-7902	2502	2502	C
UKALL I	2701	2701 2702	P
UKALL II	2703	2703 2704 2705	P
UKALL III	2706	2706	P
UKALL IV	2709	2709 2710	P
UKALL V	2711	2712	P
UKALL VI (children)	2714	2714 2716 2718 2720	P
UKALL VII	2722	2722 2723 2724 2725 2726	P
UKALL VIII	2727	2727	P
UKALL X	2730	2730	P
MRC meningeal	2740	2740	P
NCI-77-02	2803	2803	C
NCI-82-C-199	2804	2804	C

NCI-84-C-153A/CCG-144	2805	2805	C
POG 7837	2905	2905	C
POG 7866	2906	2907 2917	C
POG 8035 / POG 8036 / ALinC 13	2908	2908	C
POG 8698	2911	2911	C
POG 8704	2913	2913	C
POG 8710 / SIMAL 5	2914	2914	C
POG 8602 / AlinC 14	2918	2918 2919 2920 2922 2923 2924 2926 2927	C
POG 8304	2928	2928	C
POG 8303	2929	2929	C
POG 7834	2930	2930	C
POG 7712	2933	2933	C
POG 7818	2934	2934	C
POG 7919	2935	2935	C
POG 8022	2936	2936	C
POG 8594	2939	2939	C
NCI-D79-053-088	3001	3001	C
SJCRH X	3107	3107	C
SJCRH XI	3108	3108 3109	C
SJCRH R VIII	3112	3112	C
SWOG 8612	3214	3214	C
Austrian-BFM-86	12501	12501	P

APPENDIX III: ALGORITHM FOR THE MAIN DATA MANAGEMENT PROGRAM (JMAIN.SAS)

Note: For many of categorical variables the codes used were added as needed without much pre-planning. In order for this information to be useful, new variables must be introduced to combine categories broadly and more meaningfully. The majority of the new variables created in this program are for this purpose.

Read in data

- Read in data from text file JANALYSIS.TXT

Set up option to select dataset from all records/all results/first mentions/first results

- Set up option for use in dealing with first mentions or first reportings of results, where records which tie (see Section 5.3.1) are replaced by special records containing data from both/all tying records. This is to be 'switched off' when dealing with all records or all records which contain results.

Operations on the 'initial dataset', applicable to both definitive and publication records

- Exclude the following: i) Records resulting from review articles, duplications, randomizations that failed to open or recruit any patients and those that were not properly randomized (e.g. if the treatment allocation was by date of birth) or not randomized at all. These are identifiable using the notes codes. ii)Randomizations which are known to be open to adults only
- Create numerical date variables, for use in the calculations, for date fields in both definitive records and publication records, i.e. start and close dates of the accrual period and the official (planned) start and close dates of the trial.
- Create a year only variable for start and close dates of the accrual period, for use with graphs.
- Convert numerical unknowns into missing values. e.g. where 'number of randomization arms' = 99.
- Create new variables to distinguish between the following:
  - Randomization process takes place at the correct time versus not.
  - Methods of randomization: central computer versus notification to central office versus sealed envelope method
  - Randomization designs used: simple versus block versus minimisation of imbalance.
  - A form of balancing has been used versus not.
- For dichotomous variables set blanks to zero. This applies to equivalence trial, planned method of follow-up indicated, actual method of follow-up indicated, subgroup results reported, reported at a meeting, other eligibility criteria stated, baseline characteristics given.
- Create new variable (transformation): $\log_{10}$ (number of patients)
- Create new variable (calculation): Duration of accrual period = close date – start date
- Create new variable for whether a randomization is for children only, adults only or both adults and children, using the lowest and highest age eligible. If highest age <25 then randomization is for children only, if lowest age >13 then randomization is for adults only. Randomizations in the latter group should have been excluded already, but this serves as an additional check. If neither the highest nor lowest age is given, then the age eligibility is unknown. Otherwise assume the randomization is open to both adults and children.
- Create new variable to combine categories in order to get the following:
  - Broad eligibility risk groups: any, low, low-standard, standard, standard-high and high.
  - Broad categories for type of trial: induction, central nervous system prophylaxis, intensification, maintenance, combination of more than one of these, duration of treatment, testicular radiotherapy, bone marrow transplant, immunotherapy, post-relapse treatment, miscellaneous (e.g. antibiotic, cardio-protective, unclear).
  - Combine the latter in two ways:
    - ❖ Whether the randomization includes a transplant arm, a radiotherapy arm, an immunotherapy arm, only chemotherapy treatments, antibiotic treatments, other or unknown.
    - ❖ First line therapy versus relapse/refractory versus other/unknown.
- Use the first three digits of the trial/randomization ID to split the country group of trialists into
  - Europe, North America and Other
  - 'Developing' or 'developed' country
- Create new variable to calculate whether the target number of patients has been reached using 'number of patients accrued' and 'planned size'
- Create new variable to split into single-centre, multi-centre (limited), i.e. at least 2 but less than 5, and multi-centre participation, i.e. 5 or more.

- Similarly, for what the participation was planned to be (but not necessarily achieved): create new variable to split into single-centre, multi-centre (limited), and multi-centre.
- Create new variable, using the above two new variables with categories single-centre, limited and multi-centre, to decide whether the target number of centres was reached
- Create new variable to specify whether participation was single-country, limited (a few adjacent countries took part) or international.
- Set cut-off for date of start of randomization as 1/1/88

Identify all randomizations which started after cut-off date, for exclusion.
Start date is known for all definitive records. Using the randomization ID these can then be excluded from the set of publication records, where the start date is not always known.

- Create new temporary dataset for definitive records
- Take the initial dataset
- Delete all records with a non-zero publication ID number (these are the publication records)
- Select records where start date is after 1/1/88 and sort by randomization ID.
- Count the number of publication records for each randomization. This count to be attached to the definitive record later

Create permanent dataset for definitive records

- Take the initial dataset
- Delete all records with a non-zero publication ID number (the publication records)
- Delete records for randomizations which began after 1/1/88.
- Sort records by randomization ID
- Merge with the count of publication records and with the small permanent dataset LEUKJR.QUESTION, created using the program QUESTIONS.SAS. This contains the definitive main questions and whether ever answered in any paper.
- If the number of publications is missing, set to zero.
- Remove dates that are too inaccurate to be useful i.e. those with qualifier 3 (decade alone is known) and 4 (even decade is estimated). Do this for planned start and close dates of trial, actual start and close dates of randomization accrual period. Remember to do this for all date forms i.e. character dates, those in numerical form and year only dates where applicable. Also duration of randomization period must be set to 'missing' if either start or close date of accrual period has qualifier 3 or 4. Also remove synthetic dates (qualifier 5) for date of randomization of first, middle and last patient, since these are only used to calculate intervals and were found be very inaccurate, when compared to the start and close date of accrual period.
- Create new variables only applicable to the definitive records:
  - Time from accrual of first patient to accrual of middle patient (1)
  - Time from accrual of middle patient to accrual of last patient (2)
  - Indication of wane of interest in randomization = (2) – (1)
- Where number of questions is unknown, calculate using number of randomization arms. For a randomization with n arms, number of questions = $1 + 2 + \ldots + (n\text{-}1) = n(n\text{-}1)/2$
- Sort records by randomization ID

Create permanent dataset for publication records

- Take the initial dataset
- Delete all records with publication ID number = 0 (the definitive records)
- Remove inaccurate dates, as described previously. Do this for planned open and close date of trial, actual start and close dates of randomization accrual period, duration of randomization, if one or both of the start and close dates are inaccurate, cut-off date for analysis, dates of receipt, acceptance and publication of article. Again, remember to do this for character, numeric and year only date forms where applicable.
- Create new variables which are only applicable to the publication records:
  - Numerical dates of receipt, acceptance and publication of article.
  - Combine categories to obtain the following new variables:
    - Publication type i.e. journal article, book chapter or meeting abstract
    - Whether published in full in the English language or as an abstract in English with the full paper in another language
- Using the journal code (in conjunction with the publication type in the case of meeting abstracts) attach the following information:

- Full name of journal
- Impact factor (taken from Journal Citation Reports – 1995 Science Edition)

- Create further new variables which are only applicable to the publication records:
  - Specify whether a journal/book/abstract has no impact factor associated with it, and if so, set the impact factor to zero.
  - Merge in number of co-authors for each article.
  - Transformation of above, $\log_{10}$ (number of co-authors)
  - Alternative number of co-authors, same as number of co-authors, but setting n=30 (approximate size of working party) for articles by the MRC Working Party
  - Transformation of above, $\log_{10}$ (alternative number of co-authors)
  - Using the journal code, specify whether the country group of publication is Europe, North America, Other or Unknown.
  - Country group of trialists is same as that of publisher versus not
  - New time variables to do with articles only:
    - ❖ Time from receipt of article to acceptance for publication
    - ❖ Time from acceptance to publication
    - ❖ Time from submission/receipt to publication
  - Broad main question and answer categories, in order of importance: 1=survival (or disease-related deaths), 2=Event-free survival (EFS) (including disease-free survival (DFS)), 3=treatment-related deaths, 4=achieving remission, 5=any relapse (*Note:* 3, 4 and 5 are ranked equal), 6=specific site relapses, 7=toxicity, 8=other. Classify main questions and main results in this way.
  - Statistical technique categories for coding the two main results in order of merit: Cox regression (an advanced form of longitudinal), longitudinal (survival analysis), cross-sectional, other
  - New variable, overall statistical technique, taking the 'better' of the two main techniques stated, using the hierarchy above.
  - Whether the main questions stated in a paper have been answered in that paper. For each of the two main questions: if main question category is blank, then whether answered must also be blank. Otherwise whether answered must be classified as yes/partly/no/unclear
  - New variable whether overall the two main questions in the paper have been answered in it. If both have been answered, or if there was only one main question and it was answered then overall answered =yes. If answered is blank for both, then overall answered=not reported. Otherwise overall answered=no.
  - New variable for statistical significance of the best of the two main results: The possible categories for the two results are, ranked in order: 3 star (p<0.001), 2 star (p<0.01) 1 star (p<0.05 or significance level not stated but known to be statistically significant), ½ star (no p-value stated but possibly statistically significant), 0 star (p>0.05 or said to be not statistically significant). The overall statistical significance for the record is defined as the more significant of the two.
    If the star category is blank (statistical significance is not reported), set to 0 star.
  - New variable, a transformation of the above: Take a typical p-value for each of the above categories and take minus the logarithm of it.

category	*typical value*	*- log $_e$ (typical value)*	*spacing (i.e. distance of value from that for not stated/non-sig. category)*
3 star	0.001	-6.9	5
2 star	0.01	-4.6	2.7
1 star	0.05	-3.0	1.1
½ star	0.075	-2.6	0.7
0 star	0.15	-1.9	0

  - New variable to indicate where statistical significance is not reported.
  - New variable for overall clinical significance for the two main results:
    If either is clinically significant then overall clinical significance = yes
    Otherwise, if either is possibly clinically significant then overall clinical significance = possibly
    Otherwise if either is not clinically significant or if it is too early to tell, then overall clinical significance=no.
    If clinical significance is blank for both results, then clinical significance= not reported.
  - New variable to specify overall direction of results:
    If both are positive then overall direction=positive
    If both are negative the overall direction=negative
    If both are null then overall direction=null
    If one is positive and the other is null or unknown then overall direction=positive

If one is negative and the other is null or unknown then overall direction=negative
If one is null and the other is unknown then overall direction=null
If one is positive and the other is negative then the overall direction=opposite
If both are blank then overall direction=not reported

- New variable to specify whether record contains result(s):
  If result (e.g. survival), type of result (e.g. median), statistical technique used, statistical significance, clinical significance and direction are all blank for both results, then record does not report results. Otherwise it does.

- For each publication ID, specify:
  - Number of trials reported
  - Number of randomizations reported
    Note: For the special records which replace tied records the publication ID will be missing. Specify for each special record number the number of trials and randomizations.
  - Create new transformed variable $\log_{10}$ (number of trials reported)
- Sort records by randomization ID and output to form permanent dataset.

To create a permanent 'merged dataset'
The publication records data and the definitive records data are merged so that the most accurate information, stored as the definitive record, is used for fields that will not change over time (e.g. start and close dates of randomization period, number of patients accrued) and the publication record is used for data specific to a particular article (e.g. results, details of the journal).

- Take the permanent definitive records dataset
- Where a variable is present in both the publication records dataset and in the definitive records dataset, rename it in the definitive records dataset.
- Remove the definitive records for those randomizations which have no publication records (i.e. which remain unpublished)
- Take the permanent publication records dataset
- Drop the definitive record version of the following variables: record number, trial ID, publication ID, centre number (i.e. group of trialists), country group, 'developing'/'developed', cut-off date of 1/1/88. These are necessarily identical in the definitive and publication records.
- Replace in the publication records dataset all variables for which there is a definitive records version with the latter
- Create calculated time variables which use data from both the definitive record (close date of accrual period) and publication records (dates relating to the article) in the calculation:
  - Time from close of randomization to submission/receipt of article
  - Time from close of randomization to publication of article
  - Where not given, calculate an estimate of the median length of follow-up, using cut off date for analysis – ½ (start date + close date)
    This can only be used if the type of result median, mean or average.
  - Similarly, calculate time from cut-off date for analysis to submission of article using
    Date article received – cut off date for analysis
    =Date article received – [½ (start date + close date) + number of days on follow-up]

For each randomization, to attach an 'order of publication' number to each of the publication records
Form a new permanent dataset:

- Take the merged dataset
- *Sort by randomization ID, and within that by date of publication*
- *Count the number of records for each randomization and attach this as a variable to the merged dataset, merging by randomization ID*
- *Store as a new permanent dataset*
- *If a record is the first for a randomization, set the order number to 0*
- *Add 1 to the order number*

For each randomization, to attach an 'order of publication' number to each of the publication records that contains results
Form a new permanent dataset:
Take the merged dataset, selecting only those records that contain results. Repeat steps in *Italics*.

APPENDIX IV: COMPLETE VARIABLE LIST
(a) SHORT FORM IN ALPHABETICAL ORDER (b) IN LOGICAL ORDER

Note: X* denotes two variables X1 and X2 relating to the 2 main questions or 2 main results collected

(a) In alphabetical order

AC	age eligibility
ALLORDER	order of mention for a randomization
ANSEVER	definitive main questions ever answered?
ANSWER	main questions in paper answered in paper?
ARMS	number of arms
AUTHMRC	number of authors where given, or set to approximate size of Working Party where authorship of MRC trials is attributed to Working Party
AUTHORS	number of authors
BALANCED	balancing used in randomization?
CANS*	definitive main questions from clinprot/protocol
CENTRE	centre i.e. trialists' ID assigned by COG
CGROUP	country group of trialists
CLNSG	clinical significance of best of 2 main results
CQUEST*	main definitive questions from clinprot/protocol
CTARGET	target number of centres reached?
DEVLPNG	trial conducted in a 'developing' country
DURRAN	duration of randomization period (days)
ENGLISH	published in English language?
EQUIV	equivalence trial?
FIRSTL	first-line or relapse/refractory therapy?
FUNDG	funding source
FUPDAYS	length of follow-up period from close of randomization (days)
IMPACT	impact factor of journal
INTERNL	degree of international participation
JGROUP	country group of publisher
LOGAUTH	log_{10} (AUTHORS)
LOGMRC	log_{10} (AUTHMRC)
LOGNTREP	log_{10} (NTREP)
LOGPEST	measure of statistical significance of best of 2 main results [distance of log_e (typical *p*-value for category) from that for non-significant/not reported category]
LOGPNR	*p*-values not reported?
LOGSIZE	log_{10} (NOPAT)
MENTND	randomization ever/never mentioned in an article?
MULTIC	degree of multi-centre participation
NACCP	date accepted for publication
NCLOSE	close date of accrual period
NDPUB	date published
NOIMPACT	no impact factor associated with article?
NOPAT	number of patients randomized
NOQ	number of questions
NMENT	frequency of mentions of a randomization
NRAND	number of randomizations
NRECDP	date submitted/received for publication
NRES	frequency of reporting of results of a randomization
NRREP	number of randomizations mentioned in article
NTREP	number of trials mentioned in article
NSTART	start date of accrual period
PANS *	definitive main question from or articles
POSNG	direction of 2 main results
PQUEST*	definitive main questions from papers
PRESENTD	presented at meeting?
PUBID	publication ID

PUBTYPE	publication type: journal article/book chapter/meeting abstract
QCAT*	main questions in paper
RANDMETH	method of randomization used
RANDTIME	timing of late randomization
RANID	randomization ID assigned by COG
RCAT*	main results in paper
RCOUNT	number of publications for each randomization
RDESIGN	randomization design used
RESORDER	order of reporting of results for a randomization
RESPUB	results of randomization ever/never reported in an article
RESULTS	record contains results?
RISK	risk group eligibility
SGROUP	country group of publisher same as that of trialists?
SUBGRP	subgroup results reported?
TARGET	target number of patients reached?
TCLPUB	time from close of randomization to publication (days)
TCLREC	time from close of randomization to submission for publication (days)
TECH	type of statistical technique used
TRCAT	trial category
TRECPUB	time from receipt by publisher to publication (days)
TRID	trial ID assigned by COG
TXCHEMO	type of treatment
WANE	accrual time of 2nd half of patients minus accrual time of 1st half of patients (days)

(b) In logical order

Continuous variables which do not change over time

NRAND — number of randomizations
Missing for 0/149 (0%) trials

ARMS — number of arms
Missing for 0/243 (0%) randomizations

NOQ — number of questions (in definitive record only)
Missing for 0/243 (0%) randomizations

NOPAT — number of patients randomized
Missing for 31/243 (13%) randomizations

LOGSIZE — $\log_{10}$ (NOPAT)
Missing for 31/243 (13%) randomizations

NSTART — start date of accrual period
(START (date format) and STARTYR (discrete years) are different forms of the same variable)
Missing for 12/243 (5%) randomizations

NCLOSE — close date of accrual period
(CLOSE (date format) and CLOSEYR (discrete years) are different forms of the same variable)
Missing for 24/243 (10%) randomizations

DURRAN — duration of randomization period (days)
Missing for 24/243 (10%) randomizations

WANE — accrual time of 2nd half of patients minus accrual time of 1st half of patients
Missing for 138/243 (57%) randomizations

Dichotomous variables which do not change over time and their categories

DEVLPNG trial conducted in a 'developing' country?
1=yes
0=no
Missing for 0/149 (0%) trials

EQUIV equivalence trial?
1=yes
0=no or not known
Missing for 0/243 (0%) randomizations

Categorical variables which do not change over time and their categories

FUNDG funding source
1=Government
2=drug company
3=charity
4=Government + charity
5=Government + drug company + charity
Missing for 65/149 (44%) trials

TRCAT trial category
1=induction
2=CNS prophylaxis
3=intensification
4=maintenance
5=combination (induction/intensification/maintenance)
6=duration of treatment
7=testicular radiotherapy
8=bone marrow transplant
9=immunotherapy
10=treatment after relapse
11=miscellaneous e.g. antibiotic, cardio protection
Missing for 4/243 (2% randomizations)

TXCHEMO type of treatment
C=chemotherapy only
T=transplant
R=radiotherapy
I=immunotherapy
A=antibiotic
O=other
Missing for 7/243 (3% randomizations)

FIRSTL first-line or relapse/refractory therapy?
1=first-line treatment
2=relapse/refractory
Missing for 13/243 (5%) randomizations

CQUEST1, CQUEST2, PQUEST1, PQUEST2
definitive main questions from clinprot/protocol and papers respectively
These only apply to the definitive record

Definitions:

Event Free Survival (EFS) = time from diagnosis until induction death/failure, CR death or 1st relapse

Disease-free survival (DFS) = time from achieving complete remission to relapse/death

These are ranked in order with the exception that 3,4 &5 are of equal rank:

1=survival
2=EFS (including DFS)
3=treatment-related deaths
4=achieving CR
5=any relapse
6=specific relapse sites
7=toxicity
8=other
Missing for 164/243 (67%), 173/243 (71%), 55/243 (23%) and 94/243 (39%) randomizations respectively

AC — age eligibility
B=both adults and children
C=children
Missing for 6/243 (2%) randomizations

RISK — risk group eligibility
0=any
1=low
2=low-standard
3=standard
4=standard-high
5=high
Missing for 100/243 (41%) randomizations

TARGET — target number of patients reached?
Y=yes
N=no
Missing for 206/243 (85%) randomizations

MULTIC — degree of multi-centre participation
Y=yes i.e. >5
L=limited i.e. 2-5
N=no i.e. 1
Missing for 3/243 (1%) randomizations

CTARGET — target number of centres reached?
Y=yes
N=no
Missing for 66/243 (27%) randomizations

INTERNL — degree of international participation
N=no
L=limited
Y=yes
Missing for 30/243 (12%) randomizations

CGROUP — country group of trialists
E=Europe
A=North America
O=other
Missing for 0/149 (0%) trials

RANDTIME — timing of late randomization
Y=late randomization done at correct time
N=late randomization done too early
Missing for 176/243 (72%) randomizations, although only applicable to a small proportion of randomizations

RANDMETH method of randomization used
C=central computer
N=notification to central office
E=sealed envelopes
Missing for 196/243 (81%) randomizations

RDESIGN randomization design used
S=simple randomization
B=block randomization
M=minimisation of imbalance
Missing for 230/243 (95%) randomizations

BALANCED balancing used in randomization?
Y=some attempt made to balance
N=not balanced
Missing for 149/243 (61%) randomizations

CANS1, CANS2, PANS1 and PANS2
definitive main question from clinprot/protocol or papers respectively ever answered (in any paper)?
Note: Only applicable to the definitive record
Y=yes
N=no
Missing for 164/243 (67%), 173/243 (71%), 55/243 (23%) and 95/243 (39%) randomizations respectively

Continuous variable only usable in definitive records dataset but can change

RCOUNT number of publications for each randomization
Missing for 0/243 (0%) randomizations

Continuous variables specific to a publication

These are only present in publication records, not the definitive record.

FUPDAYS length of follow-up period from close of randomization (days)
Missing for 228/394 (58%) publication records containing results

LOGPEST measure of statistical significance of best of 2 main results (distance of log_e (typical *p*-value for category) from that for non-significant/not reported category)
Note: This is an ordered categorical variable, but it will be used as if continuous
0 =if not statistically significant or not reported
0.7=if possibly statistically significant
1.1=if statistically significant $p<0.05$
2.7=if statistically significant $p<0.01$
5=if statistically significant $p<0.001$
Missing for 0/610 (0%) publication records

IMPACT impact factor of journal - 1995 version
0=n/a or n/k
Missing for 0/257 (0%) articles

NRECD date submitted/received for publication
Note: RECDP is the same variable in date format.
Missing for 175/257 (68%) articles

NACCP date accepted for publication

Note: ACCP is the same variable in date format.
Missing for 168/257 (65%) articles

NDPUB — date published
Note: DPUB is the same variable in date format and YPUB (discrete years) the format to produce graphs.
Missing for 0/257 (0%) articles

NTREP — number of trials mentioned in article
Missing for 0/257 (0%) articles

LOGNTREP — log_{10}(NTREP)
Missing for 0/257 (0%) articles

NRREP — number of randomizations mentioned in article
Missing for 0/257 (0%) articles

ALLORDER — order of mention for a randomization
Missing for 0/610 (0%) records

RESORDER — order of reporting of results for a randomization
Missing for 0/394 (0%) records containing results

AUTHORS — number of authors
Missing for 0/257 (0%) articles

AUTHMRC — number of authors where given, and set to 30 (approximate size of a working party) where authorship is stated as Working Party
Missing for 0/257 (0%) articles

LOGAUTH — log_{10}(AUTHORS)
Missing for 0/257 (0%) articles

LOGMRC — $\log_{10}$(AUTHMRC)
Missing for 0/257 (0%) articles

Dichotomous variables specific to a publication

RESULTS — record contains results?
Y=yes
N=no
Missing for 0/610 (0%) records

LOGPNR — *p*-values not reported
1=*p*-values not reported
0=*p*-values are reported
Missing for 0/610 (0%) records

SUBGRP — subgroup results reported?
1=subgroup results reported
0=not reported or not noted
Missing for 0/610 (0%) records

PRESENTD — presented at meeting?
1=yes
0=not known
Missing for 0/257 (0%) articles

NOIMPACT no impact factor associated with journal/book/abstract
1=no impact factor associated
0=is an associated impact factor
Missing for 0/257 (0%) articles

Categorical variables specific to a publication

QCAT1, QCAT2 main questions 1 and 2 in paper
Coding as *QUEST*
Missing for 361/610 (59%) and 491/610 (80%) publication records respectively

RCAT1, RCAT2 main results 1 and 2 in paper
Coding as *QUEST*
Missing for 36/394 (9%) and 146/394 (37%) publication records containing results respectively

TECH type of statistical technique used)
L=longitudinal (survival analysis method)
X=Cox-regression (an advanced form of L)
C=cross-sectional method
O=other
Missing for 193/394 (49%) publication records containing results

POSNG direction of two main results
1=positive
-1=negative
0=null
2=opposite
3=not reported
Missing for 0/610 (0%) publication records

CLNSG clinical significance of best of 2 main results
Y=clinically significant
P=possible clinically significant
N=not clinically significant
X=not reported
Missing for 0/610 (0%) publication records

ANSWER both main questions in paper answered in same paper?
Y=main question(s) answered
N=main questions(s) not answered
X=not reported
Missing for 0/610 (0%) publication records

PUBTYPE type of article
1=journal
2=book
3=meeting paper
Missing for 0/257 (0%) articles

ENGLISH published in English language?
E=full English
A=abstract English full other
Missing for 0/257 (0%) articles

JGROUP country group of journal publisher
E=Europe

A=North America
O=other
Missing for 17/257 (7%) journals

SGROUP country group of publisher same as that of trialists?
Y=yes
N=no
Missing for 17/257 (7%) journals

Continuous time periods used as response variables

TCLREC time from close of randomization to receipt of article by publisher (days)
Missing for 401/610 (66%) publication records

TRECPUB time from receipt by journal to publication (days)
Missing for 392/610 (64% publication records

TCLPUB time from close of randomization to publication (days)
Missing for 28/610 (5%) publication records

Other continuous response variables

NMENT frequency of mentions
Missing for 0/243 (0%) randomizations

NRES frequency of reporting of results
Missing for 0/243 (0%) randomizations

Dichotomous response variables

MENTND whether randomization ever mentioned in an article
Missing for 0/243 (0%) randomizations

RESPUB whether results of randomization ever published in an article
Missing for 0/243 (0%) randomizations

Categorical grouping variables

RANID randomization ID assigned by COG
Missing for 0/243 (0%) randomizations

TRID trial ID assigned by COG
Missing for 0/149 (0%) trials

CENTRE centre i.e. trialists ID assigned by COG
Missing for 0/149 (0%) trials

PUBID publication ID applicable to publication records only
Missing for 0/257 (0%) articles

APPENDIX V: CONTINUOUS EXPLANATORY AND RESPONSE VARIABLES: MEAN, STANDARD DEVIATION, RANGE

Variable	*n*/*N*	Mean	Standard deviation	Minimum	Maximum
Explanatory					
ARMS	243/243	2.321	0.736	2	8
AUTHORS	257/257	7.584	5.301	1	42
AUTHMRC	257/257	8.241	6.310	1	42
DURRAN	219/243	1010.580	577.076	0	2922
FUPDAYS	166/394	1762.292	1027.763	6	5478
IMPACT	257/257	3.991	5.574	0	22.412
LOGAUTH	257/257	0.780	0.313	0	1.623
LOGMRC	257/257	0.805	0.326	0	1.623
LOGNTREP	257/257	0.106	0.205	0	0.778
LOGPEST	610/610	0.516	1.208	0	5
LOGSIZE	212/243	2.108	0.446	1.000	3.206
NCLOSE	219/243	7862.667	2377.951	2709	11869
NOPAT	212/243	209.057	236.698	10	1606
NOQ	243/243	1.782	2.356	1	28
NRAND	149/149	1.644	1.103	1	10
NRREP	257/257	2.374	2.031	1	11
NSTART	231/243	6752.113	2249.953	911	10210
NTREP	257/257	1.475	1.031	1	6
WANE	105/243	28.390	141.293	-360	375
Response					
NMENT	243/243	2.510	2.179	0	13
NRES	243/243	1.621	1.531	0	8
TCLPUB	582/610	1792.938	1516.016	-1679	10150
TCLREC	209/610	1923.407	1640.573	-1113	9503
TRECPUB	218/610	300.417	164.380	18	859

N = total number of trials, randomizations, articles or publication records, (whichever is appropriate for the variable in question)

n = number of observations present for variable and used in the calculation of the mean and standard deviation

APPENDIX VI: JOURNALS: NUMBER OF ARTICLES USED AND IMPACT FACTOR

---------- Publication type=Journal ----------

Publication	Number of articles	Impact factor
Acta Paediatrica Japonica	1	0.000
Acta Therapeutica	1	0.071
American Journal of Clinical Oncology	1	0.754
American Journal of Diseases of Children	2	1.433
American Journal of Medicine	1	3.749
American Journal of Pediatric Hematology/Oncology	5	1.271
Anales Espanoles de Pediatria	2	0.000
Annals of Oncology	1	2.256
Archivum Immunologiae et Therapiae Experimentalis	1	0.000
Archives of Disease in Childhood	4	1.582
Archives of Internal Medicine	1	4.166
Archives of Neurology	1	4.260
Biomedicine	1	0.000
Blood	13	8.569
British Journal of Cancer	1	3.449
British Journal of Haematology	3	2.616
British Medical Journal	9	4.549
Cancer	30	2.864
Cancer Chemotherapy Reports	1	0.000
Cancer Clinical Trials	1	0.000
Cancer Drug Delivery	1	0.000
Cancer Research	4	8.206
Cancer Treatment Reviews	1	3.106
European Journal of Pediatrics	2	1.073
Haematology and Blood Transfusion	19	0.000
International Journal of Pediatric Hematology/Oncology	2	0.000
International Journal of Radiation Oncology Biology Physics	3	2.484
International Journal of Hematology	3	0.636
Investigational New Drugs	1	0.495
Japanese Journal of Clinical Haematology	1	0.000
Journal of Cancer Research and Clinical Oncology	1	1.459
Journal of Clinical Oncology	24	6.922
Journal of Clinical Pathology	1	0.000
Journal of Pediatric Psychology	1	0.000
Klinische Padiatrie	4	0.280
Lancet	12	17.490
Leukemia	7	2.350
Medical and Pediatric Oncology	12	1.543
Neoplasia	1	0.000
Neurotoxicity	1	1.363
New England Journal of Medicine	10	22.412
Pediatric Hematology and Oncology	2	0.425
Seminars in Hematology	1	2.095
Tijdschr Kindergeneesk	1	0.000

N = 44

---------- Publication type=Book ----------

Publication	Number of articles	Impact factor
Adjuvant Therapy of Cancer	2	0
Central Nervous System Leukaemia	1	0
Immunotherapy of Cancer: Present Status of Trials in Man	1	0
Leukemia Research: Advances in Cell Biology and Treatment	3	0
Leukemia	1	0
Recent Advances in Leukaemia and Lymphoma	1	0
Unifying Concepts of Leukaemia Bibliotheca Haematologica	1	0
William Dameshek and Frederick Gunz's Leukaemia	1	0

N = 8

---------- Publication type=Meeting abstract book ----------

Publication	Number of articles	Impact factor
Blood: Abstracts of American Society of Haematology (ASH)	4	8.569
British Journal of Haematology: British Society of Haematology (BSH) and European Hematology Association (EHA) Abstracts	1	2.616
European Conference on Clinical Oncology (ECCO)	1	2.095
Haematologica: Abstracts of the European Hematology Association (EHA)	3	1.200
International Workshop on ALL: Assesment of Progress and Future Directions	1	0.000
International Journal of Radiation Oncology Biology Physics: Abstracts	1	2.484
Leukaemia Research: Scientific Program of Abstracts	1	1.179
Medical and Pediatric Oncology: International Society of Pediatric Oncology (SIOP) Abstracts	6	1.543
Proceedings of the American Association for Cancer Research (AACR)	7	0.000
Proceedings of the American Association for Cancer Research (AACR) and the American Society of Clinical Oncology (ASCO)	1	0.000
Proceedings of the American Society of Clinical Oncology (ASCO)	25	0.000

N = 11

APPENDIX VII: SUMMARY OF THE VARIABLES USED IN THE TWELVE 'HOW LONG?' ANALYSES

Analysis	1st mentions			1st results			All mentions			All results		
	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3
Initial variables												
Categorical												
CGROUP	● 1	● 1#	● 2	● 1		● 1	● 1	● 2	●2#			●4#
FUNDG		●(22)		●(15)	●(18)	●(23)	●			●(14)		
MULTIC	● 2	● 3 #		● 2			●2#	● 2	●3#	● 2	●2	
INTERNL	●2(8)#	●1(13)#	● 2#			●3#	● 1	● 3		● 1	●3	
CTARGET	●(19)	●(21)	●(18)									●(9)
TXCHEMO	● 2	● 1	● 3		● 4	● 3	●2○	● 1	● 3			● 3
FIRSTL	●	●	●			●	●		●#	●		● #
EQUIV			●			●			● #			● #
AC	● #						● #					●
RESULTS		●	●					● #	●			
CLNSG †				●2(25)#	●3(25)					●1(18)	●3(18)#	●3(16)
POSNG †				●2	●2	●3(11)				●1#(11)		●3(12)
ANSWER †											●#(22)	
PUBTYPE			●1									
JGROUP	● 3	● 1	● 2	● 1	● 3	●3#	●3#	●1#			●1	
SGROUP	●	●		●				●		● #	●	●#
PRESENTD	●	●	●	●	● #	●	●	●	●	●	●	●
SUBGRP				●	● #			● #			●#	
ENGLISH				●					●	●#○		●
Continuous												
NRAND	●	●	●	●	●	●	●	●	●	●	●	●
ARMS	●	●	●	●	●	●	●	●	●	●	●	●
NOQ	●	●	●	●	●	●	●	●	●	●	●	●
NSTART	●	●	●	●	●	●	●	●	●	●	●	●
NCLOSE	●	●	●	●	●(13)	●	●	●	●	●	●	●
DURRAN	●	●(13)	●	●	●(13)	●	●	●	●	●	●	●
LOGSIZE	●	●	●	●	●(12)	●	●(11)	●(12)	●	●	●	●
LOGPEST †				●	●(12)	●(13)				● (11)	● (11)	●(12)
IMPACT †	● (16)	● (18)	●	●	●(13)		● (9)	●(10)		● (7)	● (9)	
2nd stage variables												
DEVLPNG	●	●	●	●	●	●	●	●	●	●	●	●
LOGPEST	●	●	●	●	● #	●	●	●	● #	●	●	●
LOGPNR	● #	●	●	●	● #	● #	●	●	● #	●	● #	● #
CLNSG	● 3	● 1	● 8#	● 3#	● 4	● 1	● 5#	● 6	●2#	● 7	● 6	● 8
POSNG	● 1	● 1#	● 2	● 3	● 4	● 2	●7 #		●5#	● 6#		● 2
ANSWER	● 2	● 1	● 2	● 1	●2	● 2	● 1	● 1	● 2	● 1	● 1	●1
PUBTYPE						●1			●1			●1
IMPACT	●	●	●	●	●	●	●	●	● #	●	●	●
NOIMPACT	●	● #	●	●	●	●	●	●	● #	●	● #	●
NTREP	●	●	●	●	●	●	●	●	●	●	●	●
NRREP	●	●	●	●	●	●	●	●	●	●	●	●
AUTHORS	●	●	●	●	●	●	●	●	●	●	●	●

Analysis	1st mentions			1st results			All mentions			All results		
	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3
Interactions												
CGROUP * INTERNL		● 123						● 13		●1		
CGROUP * MULTIC				● 145			● 13	● 14	● 12	●1		
CGROUP *DURRAN	●1	●1	●1	●1		●1	●1	●1	●1			
CGROUP * LOGSIZE							●12		●12			
CGROUP * LOGPEST						●124	●13					
CGROUP * CLNSG		● 1	● 1			● 1			● 12			
CGROUP * POSNG		● 1		● 1		● 123						
CGROUP * JGROUP	●1	●1	●1					●1				
CGROUP * IMPACT			● 1	● 1		● 123	● 1	● 1				
DEVLPNG * MULTIC								●1				
DEVLPNG * INTERNL								●1				
DEVLPNG * IMPACT			●1									
INTERNL * LOGPEST										●12		
INTERNL *CLNSG		●1								●12		
INTERNL * POSNG		● 1								● 1		
INTERNL * IMPACT								●1		●1	●1	
MULTIC * LOGPEST							●12			● 12		
MULTIC * CLNSG									● 12	● 13		
MULTIC * POSNG				●1						●1		
LOGSIZE * LOGPEST							● 12					● 12
LOGSIZE *CLNSG									● 12			● 12
LOGSIZE * POSNG												● 1
EQUIV * LOGPEST						●1						
EQUIV * CLNSG			●1			●1						
EQUIV * POSNG						●1						
LOGPEST * POSNG				●		●					●	●
CLNSG * POSNG					●							

Key

T1 = time from close to submission (TCLREC)
T2 = time from receipt to publication (TRECPUB)
T3 = time from close to publication (TCLPUB)

Notes

= box-plots of classes of a categorical variable only marginally suggest inclusion
† = variable used in initial stage but without 'not reported' category
(n) = variable missing for n% observations
○ = only 2 observations in the I (TXCHEMO) and A (ENGLISH) categories, so these variables should probably be excluded

Class combinations used for categorical variables:

FUNDG	1= G+C vs. G	G=Government C= charity
TXCHEMO	1 = I vs. C R vs. C 2 = I vs. R , C 3 = I vs. R, C , A 4 = R vs. I, C	I = immunotherapy R = radiotherapy C = chemotherapy A = antibiotic
MULTIC/ INTERNL	1 = Y vs. N L vs. N 2 = Y, L vs. N 3 = Y vs. L, N	Y = yes L = limited N = no
CGROUP/JGROUP	1 = E vs. A O vs. A 2 = O vs. A, E 3 = O,E vs. A 4 = E vs. A, O	A = North America E = Europe O = other
CLNSG	1 = Y, N vs. X, P 2 = Y vs.N,P X vs. N,P 3 = Y, N, P vs. X 4 = Y vs. X P vs. X N vs. X 5 = Y, P, X vs. N 6 = Y vs. N, P, X 7 = Y, P vs. N X vs. N 8 = Y vs. X N,P vs X	Y = yes P = possibly N = no X = not reported
CLNSG†	1 = Y vs. N P vs. N 2 = Y,P vs. N 3 = Y vs. P,N	
POSNG	1 = +, O, F vs.X - vs. X 2 = +, - , O vs. X F vs. X 3 = - vs. O,X F, + vs. O, X 4 = - vs. +, F, O, X 5 = +, -, O, X vs. F 6 = X vs. +, - ,O, F 7 = + vs. X - vs. X F vs. X O vs. X	+ = positive - = negative F = flat (null) O = opposite X = not reported
POSNG†	1 = F,-,O vs. + 2 = - vs. +, O, F 3 = +, -, O vs. F	
ANSWER	1 = Y vs. X N vs. X 2 = Y, N vs. X	
PUBTYPE	1= 2 vs. 1 3 vs. 1	1 = journal article 2 = book chapter 3 = meeting abstract

Class combinations used in interactions

CGROUP*INTERNL
1 CGROUP=A INTERNL=Y
2 CGROUP=E INTERNL= L, N
3 CGROUP=O INTERNL= L, N

CGROUP * MULTIC
1 CGROUP=A MULTIC=Y, L
2 CGROUP= E, O MULTIC=L, N
3 CGROUP= E, O MULTIC=N
4 CGROUP= O MULTIC=N
5 CGROUP=E MULTIC=N

CGROUP *DURRAN
1 CGROUP=A

CGROUP * LOGSIZE
1 CGROUP=A
2 CGROUP=E, O

CGROUP * LOGPEST
1 CGROUP=A
2 CGROUP=O
3 CGROUP=E,O
4 CGROUP=E

CGROUP * CLNSG
1 CGROUP=A CLNSG=Y
2 CGROUP=E,O CLNSG=N

CGROUP * POSNG
1 CGROUP=A POSNG= +
2 CGROUP=E
3 CGROUP=O

CGROUP * JGROUP
1 CGROUP=A JGROUP=A

CGROUP * IMPACT
1 CGROUP=A
2 CGROUP=E
3 CGROUP=O

DEVLPNG * INTERNL
1 DEVLPNG=Y INTERNL=L,N

DEVLPNG * MULTIC
1 DEVLPNG=Y MULTIC=N

DEVLPNG * IMPACT
1 DEVLPNG=N

INTERNL * LOGPEST
1 INTERNL=Y LOGPEST
2 INTERNL=L, N LOGPEST

INTERNL *CLNSG
1 INTERNL=Y CLNSG=Y
2 INTERNL=L,N CLNSG=Y

INTERNL * POSNG
1 INTERNL=Y POSNG= +

INTERNL * IMPACT
1 INTERNL=Y

MULTIC * LOGPEST
1 MULTIC=Y, L
2 MULTIC=N

MULTIC * CLNSG
1 MULTIC=Y, L CLNSG=Y
2 MULTIC=L,N CLNSG=N
3 MULTIC=N CLNSG=Y

MULTIC * POSNG
1 MULTIC=Y, L POSNG= +

LOGSIZE * LOGPEST
1 LOGSIZE ↑
2 LOGSIZE ↓

LOGSIZE * CLNSG
1 CLNSG=Y
2 CLNSG=N

LOGSIZE * POSNG
1 POSNG= +

EQUIV * LOGPEST
1 EQUIV=Y

EQUIV * CLNSG
1 EQUIV=Y CLNSG=Y

EQUIV * POSNG
1 EQUIV=Y POSNG= F, -

LOGPEST * POSNG
All categories of POSNG were tried

CLNSG * POSNG
All likely combinations of categories of CLNSG and POSNG were tried

APPENDIX VIII: OUTPUT FROM THE SIX 'TIME TO FIRST' ANALYSES

(I) Data used: first mentions Response variable: time from close to submission

Indicator variables

DUMCGRP =1 if CGROUP = Europe or Other
DUMCGRP =0 if CGROUP = America

The SAS System

The GLM Procedure

Number of observations 63

Dependent Variable: TCLREC

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	164045500.5	41011375.1	49.88	<.0001
Error	58	47687433.5	822197.1		
Corrected Total	62	211732933.9			

R-Square	Coeff Var	Root MSE	TCLREC Mean
0.774776	64.21893	906.7509	1411.968

Source	DF	Type III SS	Mean Square	F Value	Pr > F
DURRAN	1	31120955.83	31120955.83	37.85	<.0001
DUMCGRP	1	21546489.68	21546489.68	26.21	<.0001
NRAND	1	5463539.54	5463539.54	6.65	0.0125
DEVLPNG	1	78954694.05	78954694.05	96.03	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	2966.907107	457.9360691	6.48	<.0001
DURRAN	-1.652309	0.2685671	-6.15	<.0001
DUMCGRP	1344.488454	262.6375339	5.12	<.0001
NRAND	-303.309601	117.6622343	-2.58	0.0125
DEVLPNG	4984.237827	508.6250636	9.80	<.0001

(II) Data used: first mentions
Response variable: time from receipt to publication

Indicator variables

DUMCGRP = 1 if CGROUP = Europe or Other
DUMCGRP = 0 if CGROUP = America

DUMCLN = 1 if CLNSG = Yes or No
DUMCLN = 0 if CLNSG = Possibly or X (not reported)

DUMJGRP2 = 1 if JGROUP = Other
DUMJGRP2 = 0 if JGROUP = America or Europe

The SAS System

The GLM Procedure

Number of observations 72

NOTE: Due to missing values, only 59 observations can be used in this analysis.

Dependent Variable: TRECPUB

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	543813.3019	77687.6146	11.44	<.0001
Error	51	346342.8676	6791.0366		
Corrected Total	58	890156.1695			

R-Square	Coeff Var	Root MSE	TRECPUB Mean
0.610919	35.69268	82.40775	230.8814

Source	DF	Type III SS	Mean Square	F Value	Pr > F
NRAND	1	44515.1829	44515.1829	6.55	0.0135
DUMJGRP2	1	168459.4320	168459.4320	24.81	<.0001
NRREP	1	83170.0528	83170.0528	12.25	0.0010
DUMCLN	1	40969.9666	40969.9666	6.03	0.0175
DUMCGRP	1	23083.4897	23083.4897	3.40	0.0710
DURRAN	1	1116.0266	1116.0266	0.16	0.6869
DUMCGRP*DURRAN	1	35535.6404	35535.6404	5.23	0.0263

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	419.0929945	50.44683513	8.31	<.0001
NRAND	-30.5920928	11.94876838	-2.56	0.0135
DUMJGRP2	-242.5832614	48.70585828	-4.98	<.0001
NRREP	-16.0828577	4.59565868	-3.50	0.0010
DUMCLN	-62.0459035	25.26087235	-2.46	0.0175
DUMCGRP	-109.6591559	59.47880058	-1.84	0.0710
DURRAN	-0.0142201	0.03507777	-0.41	0.6869
DUMCGRP*DURRAN	0.1186665	0.05187571	2.29	0.0263

(III) Data used: first mentions
Response variable: time from close to publication

Indicator variables

DUMCGRP = 1 if CGROUP = Other
DUMCGRP = 0 if CGROUP = America or Europe

DUMANS = 1 if ANSWER = Yes or No
DUMANS = 0 if ANSWER = X (not reported)

The SAS System

The GLM Procedure

Number of observations 195

Dependent Variable: TCLPUB

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	211892227.7	23543580.9	19.75	<.0001
Error	185	220577998.5	1192313.5		
Corrected Total	194	432470226.2			

R-Square	Coeff Var	Root MSE	TCLPUB Mean
0.489958	86.49328	1091.931	1262.446

Source	DF	Type III SS	Mean Square	F Value	Pr > F
NSTART	1	31716866.08	31716866.08	26.60	<.0001
DURRAN	1	8498397.83	8498397.83	7.13	0.0083
DUMCGRP	1	4933861.66	4933861.66	4.14	0.0434
PRESENTD	1	16600239.79	16600239.79	13.92	0.0003
IMPACT	1	8345380.57	8345380.57	7.00	0.0089
DUMANS	1	17228289.24	17228289.24	14.45	0.0002
EQUIV	1	8204363.06	8204363.06	6.88	0.0094
DEVLPNG	1	21085514.19	21085514.19	17.68	<.0001
NTREP	1	18899066.36	18899066.36	15.85	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	2524.418924	351.1971293	7.19	<.0001
NSTART	-0.200958	0.0389632	-5.16	<.0001
DURRAN	-0.384979	0.1441995	-2.67	0.0083
DUMCGRP	644.396678	316.7780204	2.03	0.0434
PRESENTD	-645.988106	173.1261108	-3.73	0.0003
IMPACT	-42.258653	15.9730486	-2.65	0.0089
DUMANS	643.504927	169.2879442	3.80	0.0002
EQUIV	-530.311866	202.1641486	-2.62	0.0094
DEVLPNG	1928.473752	458.5818540	4.21	<.0001
NTREP	244.904974	61.5138151	3.98	<.0001

(IV) Data used: first results
Response variable: time from close to submission

Indicator variables

DUMMULT = 1 if MULTIC = Yes or Limited (multi-centre)
DUMMULT = 0 if MULTIC = No (single-centre)

DUMCGRP1 = 1 if CGROUP = Europe
DUMCGRP1 = 0 if CGROUP = America or Other

DUMCGRP2 = 1 if CGROUP = Other
DUMCGRP2 = 0 if CGROUP = America or Europe

DUMPN1 = 1 IF POSNG = Negative
DUMPN1 = 0 IF POSNG = Null, Positive, Opposite or X (not reported)

DUMPN2 = 1 IF POSNG = Null or Positive
DUMPN2 = 0 IF POSNG = Negative, Opposite or X (not reported)

The SAS System

The GLM Procedure

Number of observations 52

Dependent Variable: TCLREC

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	18203694.72	2600527.82	5.36	0.0002
Error	44	21361054.59	485478.51		
Corrected Total	51	39564749.31			

R-Square	Coeff Var	Root MSE	TCLREC Mean
0.460099	53.84086	696.7629	1294.115

Source	DF	Type III SS	Mean Square	F Value	Pr > F
NCLOSE	1	4604740.908	4604740.908	9.48	0.0036
DUMMULT	1	4002267.011	4002267.011	8.24	0.0063
DUMCGRP1	1	6287727.454	6287727.454	12.95	0.0008
DUMCGRP2	1	6427656.394	6427656.394	13.24	0.0007
IMPACT	1	2646128.874	2646128.874	5.45	0.0242
DUMPN1	1	3128268.900	3128268.900	6.44	0.0147
DUMPN2	1	3454815.205	3454815.205	7.12	0.0107

Dependent Variable: TCLREC

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1586.189073	392.7206826	4.04	0.0002
NCLOSE	-0.170891	0.0554882	-3.08	0.0036
DUMMULT	707.071790	246.2609142	2.87	0.0063
DUMCGRP1	1290.393294	358.5586877	3.60	0.0008
DUMCGRP2	1980.913783	544.4076939	3.64	0.0007
IMPACT	118.940337	50.9458341	2.33	0.0242
DUMPN1	-910.957672	358.8650514	-2.54	0.0147
DUMPN2	-791.530235	296.7154113	-2.67	0.0107

(V) Data used: first results
Response variable: time from receipt to publication

Indicator variables

DUMFUND = 1 if FUNDG = Government + Charity
DUMFUND = 0 if FUNDG = Government alone

DUMCLN1 = 1 if CLNSG = Yes
DUMCLN1 = 0 if CLNSG = No, Possibly or X (not reported)

DUMCLN2 = 1 if CLNSG = No
DUMCLN2 = 0 if CLNSG = Yes, Possibly or X (not reported)

The SAS System

The GLM Procedure

Number of observations 60

NOTE: Due to missing values, only 49 observations can be used in this analysis.

Dependent Variable: TRECPUB

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	282037.7808	47006.2968	9.81	<.0001
Error	42	201348.4641	4794.0111		
Corrected Total	48	483386.2449			

R-Square	Coeff Var	Root MSE	TRECPUB Mean
0.583463	29.90745	69.23880	231.5102

Source	DF	Type III SS	Mean Square	F Value	Pr > F
DUMFUND	1	74239.9634	74239.9634	15.49	0.0003
NTREP	1	86083.3226	86083.3226	17.96	0.0001
NRREP	1	205309.3058	205309.3058	42.83	<.0001
PRESENTD	1	46789.9137	46789.9137	9.76	0.0032
DUMCLN1	1	40920.3761	40920.3761	8.54	0.0056
DUMCLN2	1	44350.9529	44350.9529	9.25	0.0040

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	282.3021900	30.48992918	9.26	<.0001
DUMFUND	83.1143513	21.12061988	3.94	0.0003
NTREP	42.2871926	9.97927196	4.24	0.0001
NRREP	-33.5653528	5.12904388	-6.54	<.0001
PRESENTD	69.8529810	22.35930692	3.12	0.0032
DUMCLN1	-81.6344897	27.94173628	-2.92	0.0056
DUMCLN2	-84.8701916	27.90314914	-3.04	0.0040

(VI) Data used: first results
Response variable: time from close to publication

Indicator variables

DUMCGRP1 = 1 if CGROUP = Europe
DUMCGRP1 = 0 if CGROUP = America or Other

DUMCGRP2 = 1 if CGROUP = Other
DUMCGRP2 = 0 if CGROUP = America or Europe

DUMCLN = 1 if CLNSG = Yes or No
DUMCLN = 0 if CLNSG = Possibly or X (not reported)

DUMDIR3 = 1 if POSNG = Null
DUMDIR3 = 0 if POSNG = Positive, Negative, Opposite or X (not reported)

The SAS System

The GLM Procedure

Number of observations 170

Dependent Variable: TCLPUB

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	70026193.2	7002619.3	9.10	<.0001
Error	159	122324572.8	769336.9		
Corrected Total	169	192350766.0			

R-Square	Coeff Var	Root MSE	TCLPUB Mean
0.364055	66.95952	877.1185	1309.924

Source	DF	Type III SS	Mean Square	F Value	Pr > F
DURRAN	1	10960803.91	10960803.91	14.25	0.0002
DUMCGRP1	1	1002867.73	1002867.73	1.30	0.2553
DUMCGRP2	1	10931.17	10931.17	0.01	0.9053
NOIMPACT	1	13247370.13	13247370.13	17.22	<.0001
DUMCLN	1	5958769.47	5958769.47	7.75	0.0060
IMPACT	1	5925128.13	5925128.13	7.70	0.0062
LOGPEST	1	4625126.09	4625126.09	6.01	0.0153
DUMDIR3	1	4889632.03	4889632.03	6.36	0.0127
DUMCGRP1*IMPACT	1	9426163.00	9426163.00	12.25	0.0006
DUMCGRP2*DUMDIR3	1	7409723.47	7409723.47	9.63	0.0023

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	2676.188113	207.1635603	12.92	<.0001
DURRAN	-0.463742	0.1228607	-3.77	0.0002
DUMCGRP1	-215.819645	189.0285648	-1.14	0.2553
DUMCGRP2	32.017728	268.6058842	0.12	0.9053
NOIMPACT	-716.286008	172.6155531	-4.15	<.0001
DUMCLN	-468.630833	168.3878915	-2.78	0.0060
IMPACT	-49.701284	17.9092347	-2.78	0.0062
LOGPEST	-143.606012	58.5691776	-2.45	0.0153
DUMDIR3	-434.963330	172.5332678	-2.52	0.0127
DUMCGRP1*IMPACT	116.108156	33.1706169	3.50	0.0006
DUMCGRP2*DUMDIR3	1653.784946	532.8884726	3.10	0.0023

APPENDIX IX: OUTPUT FROM THE SIX PRELIMINARY ANALYSES THAT REQUIRE INCORPORATION OF REPEATED MEASURES

(I) Data used: all mentions
Response variable: time from close to submission

The SAS System

The GLM Procedure

Number of observations 209

Dependent Variable: TCLREC

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	239507659.2	39917943.2	25.17	<.0001
Error	202	320320199.2	1585743.6		
Corrected Total	208	559827858.4			

R-Square	Coeff Var	Root MSE	TCLREC Mean
0.427824	65.47045	1259.263	1923.407

Source	DF	Type III SS	Mean Square	F Value	Pr > F
DURRAN	1	26422945.4	26422945.4	16.66	<.0001
IMPACT	1	17020566.1	17020566.1	10.73	0.0012
LOGPEST	1	19038716.4	19038716.4	12.01	0.0006
DEVLPNG	1	115291241.2	115291241.2	72.70	<.0001
LOGNTREP	1	11043522.1	11043522.1	6.96	0.0090
NRREP	1	36299596.6	36299596.6	22.89	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	3491.432978	337.2134594	10.35	<.0001
DURRAN	-0.723688	0.1772870	-4.08	<.0001
IMPACT	-102.299476	31.2250317	-3.28	0.0012
LOGPEST	248.954897	71.8486016	3.46	0.0006
DEVLPNG	4404.250493	516.5237941	8.53	<.0001
LOGNTREP	1119.731011	424.3032724	2.64	0.0090
NRREP	-205.378841	42.9260847	-4.78	<.0001

(II) Data used: all mentions
Response variable: √(time from receipt to publication)

Indicator variables

DUMJGRP2 = 1 if JGROUP = Other
DUMJGRP2 = 0 if JGROUP = America or Europe

DUMCGRP1 = 1 if CGROUP = Europe
DUMCGRP1 = 0 if CGROUP = America or Other

The SAS System

The GLM Procedure

Number of observations 218

NOTE: Due to missing values, only 195 observations can be used in this analysis.

Dependent Variable: RTRECPUB

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	1936.218689	322.703115	31.70	<.0001
Error	188	1913.983806	10.180765		
Corrected Total	194	3850.202495			

R-Square	Coeff Var	Root MSE	RTRECPUB Mean
0.502887	19.61802	3.190731	16.26429

Source	DF	Type III SS	Mean Square	F Value	Pr > F
NSTART	1	297.1149815	297.1149815	29.18	<.0001
DUMJGRP2	1	536.9364177	536.9364177	52.74	<.0001
DURRAN	1	142.4349421	142.4349421	13.99	0.0002
NRREP	1	539.1025263	539.1025263	52.95	<.0001
DUMCGRP1	1	281.4252646	281.4252646	27.64	<.0001
LOGAUTH	1	65.3142155	65.3142155	6.42	0.0121

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	12.50450749	1.10525891	11.31	<.0001
NSTART	0.00058148	0.00010764	5.40	<.0001
DUMJGRP2	-12.83411876	1.76723713	-7.26	<.0001
DURRAN	0.00177204	0.00047376	3.74	0.0002
NRREP	-0.62203603	0.08548109	-7.28	<.0001
DUMCGRP1	-3.38810302	0.64441445	-5.26	<.0001
LOGAUTH	2.21211676	0.87336221	2.53	0.0121

(III) Data used: all mentions
Response variable: time from close to publication

Indicator variables

DUMMULT = 1 if MULTIC = Yes (multi-centre)
DUMMULT = 0 if MULTIC = Limited or No (single-centre)

DUMCGRP1 = 1 if CGROUP = Europe
DUMCGRP1 = 0 if CGROUP = America or Other

DUMCGRP2 = 1 if CGROUP = Other
DUMCGRP2 = 0 if CGROUP = America or Europe

DUMENG = 1 if published in full in a language other than English, with an English abstract
DUMENG = 0 if published in full in English

DUMCLN1 = 1 if CLNSG = Yes
DUMCLN1 = 0 if CLNSG = No, Possibly or X (not reported)

DUMCLN2 = 1 if CLNSG = X (not reported)
DUMCLN2 = 0 if CLNSG = Yes, No or Possibly

DUMANS = 1 if ANSWER = Yes or No
DUMANS = 0 if ANSWER = X (not reported)

PUBTYPE3 = 1 if PUBTYPE = Meeting abstract
PUBTYPE3 = 0 if PUBTYPE = Journal or Book

Interaction term

ICGMUL2 = 1 if non-US trial with less than 5 centres
ICGMUL2 = 0 otherwise

The SAS System

The GLM Procedure

Number of observations 582

NOTE: Due to missing values, only 581 observations can be used in this analysis.

Dependent Variable: TCLPUB

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	16	455689940	28480621	18.26	<.0001
Error	564	879561623	1559506		
Corrected Total	580	1335251563			

R-Square	Coeff Var	Root MSE	TCLPUB Mean
0.341276	69.63444	1248.802	1793.368

Source	DF	Type III SS	Mean Square	F Value	Pr > F
NSTART	1	43500131.20	43500131.20	27.89	<.0001
DURRAN	1	27852603.67	27852603.67	17.86	<.0001
DUMCGRP1	1	20083762.47	20083762.47	12.88	0.0004
DUMCGRP2	1	13376.87	13376.87	0.01	0.9262
DUMENG	1	12921648.07	12921648.07	8.29	0.0041
PRESENTD	1	45319683.78	45319683.78	29.06	<.0001
DUMCLN1	1	13847646.75	13847646.75	8.88	0.0030
DUMCLN2	1	14089452.51	14089452.51	9.03	0.0028
DUMANS	1	28701395.55	28701395.55	18.40	<.0001
DEVLPNG	1	41505168.94	41505168.94	26.61	<.0001
NTREP	1	34515944.19	34515944.19	22.13	<.0001
NRREP	1	14958164.53	14958164.53	9.59	0.0021
IMPACT	1	7853063.00	7853063.00	5.04	0.0252
DUMMULT	1	38602548.53	38602548.53	24.75	<.0001
DUMPUBT3	1	7847942.79	7847942.79	5.03	0.0253
ICGMUL2	1	18517354.16	18517354.16	11.87	0.0006

Dependent Variable: TCLPUB

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	2344.830636	266.1429356	8.81	<.0001
NSTART	-0.150788	0.0285506	-5.28	<.0001
DURRAN	-0.439751	0.1040560	-4.23	<.0001
DUMCGRP1	-535.114001	149.1136359	-3.59	0.0004
DUMCGRP2	23.508894	253.8332300	0.09	0.9262
DUMENG	-1048.554597	364.2719064	-2.88	0.0041
PRESENTD	-637.945194	118.3405324	-5.39	<.0001
DUMCLN1	444.477517	149.1610371	2.98	0.0030
DUMCLN2	438.225404	145.7954851	3.01	0.0028
DUMANS	523.716816	122.0783345	4.29	<.0001
DEVLPNG	2182.688289	423.0914114	5.16	<.0001
NTREP	243.977729	51.8601645	4.70	<.0001
NRREP	-91.030498	29.3928265	-3.10	0.0021
IMPACT	-25.620568	11.4172839	-2.24	0.0252
DUMMULT	813.222478	163.4537804	4.98	<.0001
DUMPUBT3	-354.652786	158.0953317	-2.24	0.0253
ICGMUL2	1226.461906	355.9247439	3.45	0.0006

(IV) Data used: all reportings of results
Response variable: time from close to submission

Indicator variables

DUMCLN = 1 if CLNSG = Yes, Possibly or X (not reported)
DUMCLN = 0 if CLNSG = No

DUMPN = 1 if POSNG = X (not reported)
DUMPN = 0 if POSNG = Null, Negative, Opposite or Positive

The SAS System

The GLM Procedure

Number of observations 129

Dependent Variable: TCLREC

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	63537508.8	10589584.8	9.71	<.0001
Error	122	133099522.0	1090979.7		
Corrected Total	128	196637030.8			

R-Square	Coeff Var	Root MSE	TCLREC Mean
0.323121	55.94419	1044.500	1867.039

Source	DF	Type III SS	Mean Square	F Value	Pr > F
NOQ	1	12227391.26	12227391.26	11.21	0.0011
PRESENTD	1	15227447.66	15227447.66	13.96	0.0003
IMPACT	1	7478041.04	7478041.04	6.85	0.0100
LOGPEST	1	9735783.61	9735783.61	8.92	0.0034
DUMCLN	1	4888080.39	4888080.39	4.48	0.0363
DUMPN	1	4370786.51	4370786.51	4.01	0.0475

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1665.308430	240.1279943	6.94	<.0001
NOQ	106.040895	31.6748679	3.35	0.0011
PRESENTD	-737.627826	197.4386016	-3.74	0.0003
IMPACT	-83.315715	31.8230158	-2.62	0.0100
LOGPEST	195.050060	65.2933519	2.99	0.0034
DUMCLN	493.791299	233.2827338	2.12	0.0363
DUMPN	610.754782	305.1373789	2.00	0.0475

(V) Data used: all reportings of results
Response variable: time from receipt to publication

Indicator variables

DUMMULT = 1 if MULTIC = Yes or Limited (multi-centre)
DUMMULT = 0 if MULTIC = No (single-centre)

DUMINT = 1 if INTERNL = Yes (international)
DUMINT = 0 if INTERNL = Limited or No (single-country)

DUMANS1 = 1 if ANSWER = Yes
DUMANS1 = 0 if ANSWER = No or X (not reported)

DUMJGRP1 = 1 if JGROUP = Europe
DUMJGRP1 = 0 if JGROUP = America or Other

DUMJGRP2 = 1 if JGROUP = Other
DUMJGRP2 = 0 if JGROUP = America or Europe

Interaction term

IINTIMP = 1 if truly international trial and published in a journal with a high impact factor (>=6)
IINTIMP = 0 otherwise

The SAS System
The GLM Procedure

Number of observations 137
NOTE: Due to missing values, only 121 observations can be used in this analysis.

Dependent Variable: TRECPUB

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	1591821.909	176869.101	13.19	<.0001
Error	111	1488785.050	13412.478		
Corrected Total	120	3080606.959			

R-Square	Coeff Var	Root MSE	TRECPUB Mean
0.516723	40.72208	115.8123	284.3967

Source	DF	Type III SS	Mean Square	F Value	Pr > F
DUMINT	1	192143.1012	192143.1012	14.33	0.0002
DUMMULT	1	242555.8449	242555.8449	18.08	<.0001
DUMJGRP1	1	338158.6972	338158.6972	25.21	<.0001
DUMJGRP2	1	347074.1169	347074.1169	25.88	<.0001
DURRAN	1	499587.4255	499587.4255	37.25	<.0001
DUMANS1	1	114434.3629	114434.3629	8.53	0.0042
PRESENTD	1	89059.0115	89059.0115	6.64	0.0113
IMPACT	1	24891.2211	24891.2211	1.86	0.1759
DUMINT*IMPACT	1	140376.0333	140376.0333	10.47	0.0016

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	146.3566220	40.48752438	3.61	0.0005
DUMINT	-364.8985895	96.40827156	-3.78	0.0002
DUMMULT	114.8628458	27.01022549	4.25	<.0001
DUMJGRP1	-171.5287034	34.16103607	-5.02	<.0001
DUMJGRP2	-388.8656479	76.44392471	-5.09	<.0001
DURRAN	0.1302510	0.02134175	6.10	<.0001
DUMANS1	66.2695873	22.68770405	2.92	0.0042
PRESENTD	-63.1438472	24.50454334	-2.58	0.0113
IMPACT	-7.1016176	5.21301162	-1.36	0.1759
DUMINT*IMPACT	45.2175710	13.97704154	3.24	0.0016

(VI) Data used: all reportings of results
Response variable: time from close to publication

Indicator variables

DUMCLN2 = 1 if CLNSG = No or Possibly
DUMCLN2 = 0 if CLNSG = Yes or X (not reported)

DUMANS = 1 if ANSWER = Yes or No
DUMANS = 0 if ANSWER = X (not reported)

DUMFL = 1 if FIRSTL = Treatment for relapse or refractory disease
DUMFL = 0 if FIRSTL = First-line treatment

DUMDIR1 = 1 if POSNG = Positive
DUMDIR1 = 0 if POSNG = Negative, Null, Opposite or X (not reported)

DUMDIR2 = 1 if POSNG = Negative
DUMDIR2 = 0 if POSNG = Positive, Null, Opposite or X (not reported)

The SAS System
The GLM Procedure

Number of observations 372

NOTE: Due to missing values, only 361 observations can be used in this analysis.

Dependent Variable: TCLPUB

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	147092257.9	12257688.2	8.00	<.0001
Error	348	533181438.3	1532130.6		
Corrected Total	360	680273696.2			

R-Square	Coeff Var	Root MSE	TCLPUB Mean
0.216225	67.65727	1237.793	1829.504

Source	DF	Type III SS	Mean Square	F Value	Pr > F
NOQ	1	8508038.85	8508038.85	5.55	0.0190
NCLOSE	1	21791464.59	21791464.59	14.22	0.0002
PRESENTD	1	16226892.84	16226892.84	10.59	0.0012
DUMCLN2	1	11063493.93	11063493.93	7.22	0.0076
DUMANS	1	12497499.37	12497499.37	8.16	0.0045
DUMFL	1	10878132.64	10878132.64	7.10	0.0081
NTREP	1	11332892.08	11332892.08	7.40	0.0069
LOGPEST	1	23867963.98	23867963.98	15.58	<.0001
DUMDIR1	1	3986651.11	3986651.11	2.60	0.1076
DUMDIR2	1	3617994.85	3617994.85	2.36	0.1253
LOGPEST*DUMDIR1	1	21729794.60	21729794.60	14.18	0.0002
LOGPEST*DUMDIR2	1	10819633.62	10819633.62	7.06	0.0082

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	2036.559165	301.8061314	6.75	<.0001
NOQ	47.859027	20.3093965	2.36	0.0190
NCLOSE	-0.101993	0.0270443	-3.77	0.0002
PRESENTD	-461.682526	141.8644709	-3.25	0.0012
DUMCLN2	-405.869017	151.0384910	-2.69	0.0076
DUMANS	416.199017	145.7261118	2.86	0.0045
DUMFL	692.588037	259.9235918	2.66	0.0081
NTREP	136.214422	50.0841841	2.72	0.0069
LOGPEST	330.632008	83.7693405	3.95	<.0001
DUMDIR1	276.154707	171.1970292	1.61	0.1076
DUMDIR2	339.431067	220.8846343	1.54	0.1253
LOGPEST*DUMDIR1	-390.871944	103.7897767	-3.77	0.0002
LOGPEST*DUMDIR2	-386.688245	145.5132636	-2.66	0.0082

APPENDIX X: ALGORITHMS AND SAS COMMANDS FOR THE REPEATED MEASURES PRELIMINARY INVESTIGATIONS AND ANALYSES

(i) Algorithm for producing the correlation matrix from the residuals resulting from the independence model

- Use one of the two publication records datasets output at the end of the main data management program, JMAIN.
  If all mentions are to be used, use dataset LEUKJR.ALLRECS. Here the order of each record (from an article) for a particular randomization is ALLORDER
  If only records containing results are to be used, use dataset LEUKJR.RESRECS. Here the order of each record for a particular randomization is RESORDER
- Run the regression as usual, outputting the residuals
- Only keep records actually used in the regression i.e. remove those where one or more X variables is missing, and therefore there is no residual, RESID
- Rename the residuals from the first publication of each randomization as RESID1, those from the second to RESID2 and so on up to and including RESID5.
- Output all the residuals from the 1st publications (RESID1) and the randomization ID (RANID) to one dataset, all the residuals from the 2nd publications (RESID2) and RANID to a 2nd dataset and so on up to 5th.
- Merge the five datasets by variable RANID producing a single dataset with five variables RESID1, RESID2 … RESID5
- Obtain correlation matrix for the five variables (PROC CORR in SAS)

(ii) Algorithm for program to calculate: correlation coefficient (ρ) for a matrix with an exchangeable correlation structure, variance inflation factor, new standard errors for the parameter (β) estimates, and hence revised *t*-statistics and *p*-values for the parameter estimates

- For each off-diagonal correlation coefficient of the matrix obtained from the previous program input as variables the row position (ROW), column position (COLUMN), value of correlation coefficient (PIJ) and number of pairs of observations used to calculate correlation coefficient (NIJ).Do this for elements above the diagonal only. (Matrix is symmetrical.)
- The formula for estimating ρ is given in Section 8.7. The following SAS commands will do this:

(Assume the dataset used is TEMP1. Here PEST is ρ.)

```
ZIJ=LOG(((1+PIJ)/(1-PIJ))**0.5);
NIJM3=NIJ-3;
NIJM3ZIJ=NIJM3*ZIJ;

CUM1+NIJM3ZIJ;
CUM2+NIJM3;
RUN;

DATA TEMP2;
  SET TEMP1 END=FINAL;

ZESTTOP=CUM1;
ZESTBOT=CUM2;
```

```
IF FINAL;
ZEST=ZESTTOP/ZESTBOT;
PEST=(EXP(2*ZEST)-1)/(EXP(2*ZEST)+1);
```

- Hence the variance inflation factor (VIF) (see Section 8.8) is calculated
  VIF= 1- ρ^2
- For each variable read in the variable name and the values of β and the standard error of β from the independence model
  For new value of SE(β), divide original value by √VIF
  For new *t*-value, divide β by new value of SE(β)
  Hence obtain new p-value. The following SAS command gives the *p*-value for a 2-sided test:

```
PROB_NEW = 2*(1-PROBT(ABS(T_NEW),d));
```

where *d*= residual degrees of freedom

(iii) Algorithm for program for the preliminary investigation imposing a stationary *m*-dependent correlation structure

- As with the exchangeable correlation structure, described in Section (ii), for each off-diagonal element of the correlation matrix formed from the residuals from the independence model, input as variables the row position (ROW), column position (COLUMN), value of correlation coefficient (PIJ) and number of observations used to calculate correlation coefficient (NIJ). Do this for elements above the diagonal only.
- Output to 1st new dataset those elements distance 1 from the diagonal i.e.(COLUMN-ROW=1). Similarly output to 2nd dataset those elements distance 2 from the diagonal and so on up to 4th dataset.
- Deal with each dataset in turn, estimating a separate correlation coefficient for elements distance 1,2, 3 and 4 from the diagonal, using the method described in the previous section.

(iv) SAS commands for running the generalised estimating equation analysis

The variables in the model are those for the analysis using the largest dataset, time from close to publication all mentions

```
PROC GENMOD;
  CLASS RANID ALLORDER;
  MODEL TCLPUB = NSTART DURRAN DUMCGRP1 DUMCGRP2 DUMENG PRESENTD
                 DUMCLN1 DUMCLN2 DUMANS DEVLPNG NTREP NRREP IMPACT
                 DUMPUBT3 DUMMULT ICGMUL2;
  REPEATED SUBJECT=RANID/WITHIN=ALLORDER TYPE=IND CORRW MODELSE;
  OUTPUT OUT=TEMP1 PREDICTED=YHAT RESCHI=RESID;
RUN;
```

Notes

1. RANID (randomization ID) is the clustering variable.
   For each randomization ALLORDER is the order of publication of the article
   (For analyses for all reportings of results, rather than all mentions, RESORDER is the order of publication of the article.)

2. In this example the independence correlation structure has been selected.

Set TYPE to impose correlation structure required

EXCH	= exchangeable
MDEP(*n*)	= stationary *m*-dependent with *m* = *n*
AUTO	= autoregressive

3. MODELSE produces model-based estimates of the standard error of each parameter (β) estimate

4. The output statement produces fitted values and residuals for use in model testing

(v) SAS commands to run a linear mixed effects model

```
PROC MIXED;
  CLASS RANID ALLORDER;
  MODEL TCLPUB = NSTART DURRAN DUMCGRP1 DUMCGRP2 DUMENG PRESENTD
                 DUMCLN1 DUMCLN2 DUMANS DEVLPNG NTREP NRREP IMPACT
                 DUMPUBT3 DUMMULT ICGMUL2/S;
  REPEATED ALLORDER/SUBJECT=RANID TYPE=CS;
RUN;
```

Notes

1. RANID (the randomization ID) is the clustering variable. For each randomization ALLORDER is the order of publication of the article.

2. Set TYPE to impose correlation structure required

CS	= exchangeable
AR(1)	= autoregressive
TOEP(*n*)	= stationary *m*-dependent with *m*=*n*

Omitting the REPEATED statement produces the independence model

3. S is the command to print beta estimates and their standard errors

APPENDIX XI: CORRELATION MATRICES FOR THE RESIDUALS FROM THE SIX ANALYSES THAT REQUIRE INCORPORATION OF REPEATED MEASURES

The SAS System
The CORR Procedure

(I) Data used: all mentions
Response variable: time from close to submission

Variable	N
RESID1	63
RESID2	32
RESID3	32
RESID4	38
RESID5	22

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	RESID1	RESID2	RESID3	RESID4	RESID5
RESID1	1.00000	0.97505	0.67907	-0.54919	0.78506
		<.0001	0.1380	0.3377	0.0122
	63	8	6	5	9
RESID2	0.97505	1.00000	0.66503	0.84975	1.00000
	<.0001		0.0183	0.0002	.
	8	32	12	13	2
RESID3	0.67907	0.66503	1.00000	0.52495	0.89496
	0.1380	0.0183		0.0655	0.0402
	6	12	32	13	5
RESID4	-0.54919	0.84975	0.52495	1.00000	0.72177
	0.3377	0.0002	0.0655		0.1686
	5	13	13	38	5
RESID5	0.78506	1.00000	0.89496	0.72177	1.00000
	0.0122	.	0.0402	0.1686	
	9	2	5	5	22

(II) Data used: all mentions
Response variable: time from receipt to publication

Variable	N
RESID1	59
RESID2	29
RESID3	29
RESID4	38
RESID5	20

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	RESID1	RESID2	RESID3	RESID4	RESID5
RESID1	1.00000	-0.46045	-0.32274	-0.64546	0.73565
		0.2509	0.5963	0.2395	0.0595
	59	8	5	5	7
RESID2	-0.46045	1.00000	0.07927	0.73546	-1.00000
	0.2509		0.8168	0.0042	.
	8	29	11	13	2
RESID3	-0.32274	0.07927	1.00000	0.93523	-0.79386
	0.5963	0.8168		<.0001	0.1088
	5	11	29	13	5
RESID4	-0.64546	0.73546	0.93523	1.00000	-0.12197
	0.2395	0.0042	<.0001		0.8451
	5	13	13	38	5
RESID5	0.73565	-1.00000	-0.79386	-0.12197	1.00000
	0.0595	.	0.1088	0.8451	
	7	2	5	5	20

(III) Data used: all mentions
Response variable: time from close to publication

Variable	N
RESID1	194
RESID2	128
RESID3	90
RESID4	68
RESID5	44

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	RESID1	RESID2	RESID3	RESID4	RESID5
RESID1	1.00000	0.35271	0.16092	0.16406	-0.05110
		<.0001	0.1297	0.1813	0.7418
	194	128	90	68	44
RESID2	0.35271	1.00000	0.46667	0.40235	0.20546
	<.0001		<.0001	0.0007	0.1809
	128	128	90	68	44
RESID3	0.16092	0.46667	1.00000	0.66755	0.45929
	0.1297	<.0001		<.0001	0.0017
	90	90	90	68	44
RESID4	0.16406	0.40235	0.66755	1.00000	0.75220
	0.1813	0.0007	<.0001		<.0001
	68	68	68	68	44
RESID5	-0.05110	0.20546	0.45929	0.75220	1.00000
	0.7418	0.1809	0.0017	<.0001	
	44	44	44	44	44

(IV) Data used: all reportings of results
Response variable: time from close to submission

Variable	N
RESID1	52
RESID2	22
RESID3	23
RESID4	17
RESID5	8

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	RESID1	RESID2	RESID3	RESID4	RESID5
RESID1	1.00000	0.48915	0.16009	1.00000	-0.49005
		0.2653	0.7970	.	0.5100
	52	7	5	2	4
RESID2	0.48915	1.00000	.	0.99999	.
	0.2653		.	0.0026	.
	7	22	1	3	0
RESID3	0.16009	.	1.00000	0.89594	0.99453
	0.7970	.		0.0397	0.0666
	5	1	23	5	3
RESID4	1.00000	0.99999	0.89594	1.00000	.
	.	0.0026	0.0397		.
	2	3	5	17	1
RESID5	-0.49005	.	0.99453	.	1.00000
	0.5100	.	0.0666	.	
	4	0	3	1	8

(V) Data used: all reportings of results
Response variable: time from receipt to publication

Variable	N
RESID1	50
RESID2	20
RESID3	23
RESID4	16
RESID5	6

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	RESID1	RESID2	RESID3	RESID4	RESID5
RESID1	1.00000	-0.63539	0.21270	1.00000	1.00000
		0.1252	0.7312	.	.
	50	7	5	2	2
RESID2	-0.63539	1.00000	.	-0.97530	.
	0.1252		.	0.1418	.
	7	20	1	3	0
RESID3	0.21270	.	1.00000	0.45448	-0.74345
	0.7312	.		0.4419	0.4664
	5	1	23	5	3
RESID4	1.00000	-0.97530	0.45448	1.00000	.
	.	0.1418	0.4419		.
	2	3	5	16	1
RESID5	1.00000	.	-0.74345	.	1.00000
	.	.	0.4664	.	
	2	0	3	1	6

(VI) Data used: all reportings of results
Response variable: time from close to publication

Variable	N
RESID1	164
RESID2	91
RESID3	54
RESID4	29
RESID5	13

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	RESID1	RESID2	RESID3	RESID4	RESID5
RESID1	1.00000	0.53121	0.56003	0.08945	-0.13473
		<.0001	<.0001	0.6445	0.6608
	164	91	54	29	13
RESID2	0.53121	1.00000	0.75606	0.38425	-0.10232
	<.0001		<.0001	0.0396	0.7394
	91	91	54	29	13
RESID3	0.56003	0.75606	1.00000	0.73189	0.63482
	<.0001	<.0001		<.0001	0.0198
	54	54	54	29	13
RESID4	0.08945	0.38425	0.73189	1.00000	0.89130
	0.6445	0.0396	<.0001		<.0001
	29	29	29	29	13
RESID5	-0.13473	-0.10232	0.63482	0.89130	1.00000
	0.6608	0.7394	0.0198	<.0001	
	13	13	13	13	13

APPENDIX XII: OUTPUT FROM EACH ANALYSIS IMPOSING AN EXCHANGEABLE CORRELATION STRUCTURE CALCULATED USING THE VARIANCE INFLATION FACTOR METHOD

Key
PEST = estimate of correlation coefficient
VIF = variance inflation factor
SQRTVIF = √ VIF

(I) Data used: all mentions
Response variable: time from close to submission

PEST	VIF	SQRTVIF
0.76024	0.42204	0.64964

Obs	VARIABLE	BETA	SE_OLD	T_OLD	PROB_OLD	SE_NEW	T_NEW	PROB_NEW
1	INTERCPT	3491.43	337.213	10.3538	0	519.077	6.72623	0.000000
2	DURRAN	-0.72	0.177	-4.0820	.000064627	0.273	-2.65184	0.008651
3	IMPACT	-102.30	31.225	-3.2762	.001241250	48.065	-2.12835	0.034538
4	LOGPEST	248.95	71.849	3.4650	.000649377	110.598	2.25100	0.025479
5	DEVLPNG	4404.25	516.524	8.5267	3.7748E-15	795.092	5.53929	0.000000
6	LOGNTREP	1119.73	424.303	2.6390	.008973789	653.136	1.71439	0.088014
7	NRREP	-205.38	42.926	-4.7845	.000003335	66.077	-3.10819	0.002158

(II) Data used: all mentions
Response variable: time from receipt to publication

PEST	VIF	SQRTVIF
0.48262	0.76708	0.87583

Obs	VARIABLE	BETA	SE_OLD	T_OLD	PROB_OLD	SE_NEW	T_NEW	PROB_NEW
1	INTERCPT	12.5045	1.10526	11.3136	0.000000	1.26196	9.90883	0.000000
2	NSTART	0.0006	0.00011	5.4033	0.000000	0.00012	4.73241	0.000004
3	DUMJGRP2	-12.8341	1.76724	-7.2623	0.000000	2.01779	-6.36050	0.000000
4	DURRAN	0.0018	0.00047	3.7408	0.000244	0.00054	3.27627	0.001254
5	NRREP	-0.6220	0.08548	-7.2769	0.000000	0.09760	-6.37332	0.000000
6	DUMCGRP	-3.3881	0.64441	-5.2576	0.000000	0.73578	-4.60480	0.000008
7	LOGAUTH	2.2121	0.87336	2.5329	0.012135	0.99718	2.21837	0.027733

(III) Data used: all mentions
Response variable: time from close to publication

PEST	VIF	SQRTVIF
0.37837	0.85684	0.92566

Obs	VARIABLE	BETA	SE_OLD	T_OLD	PROB_OLD	SE_NEW	T_NEW	PROB_NEW
1	INTERCPT	2344.83	266.143	8.81042	0.00000	287.517	8.15545	0.00000
2	NSTART	-0.15	0.029	-5.28143	0.00000	0.031	-4.88881	0.00000
3	DURRAN	-0.44	0.104	-4.22610	0.00003	0.112	-3.91193	0.00010
4	DUMCGRP1	-535.11	149.114	-3.58863	0.00036	161.089	-3.32185	0.00095
5	DUMCGRP2	23.51	253.833	0.09262	0.92624	274.219	0.08573	0.93171
6	DUMENG	1048.55	364.272	2.87849	0.00415	393.527	2.66451	0.00793
7	PRESENTD	-637.95	118.341	-5.39076	0.00000	127.844	-4.99001	0.00000
8	DUMCLN1	444.48	149.161	2.97985	0.00301	161.140	2.75833	0.00600
9	DUMCLN2	438.23	145.795	3.00575	0.00277	157.504	2.78231	0.00558
10	DUMANS	523.72	122.078	4.29001	0.00002	131.882	3.97109	0.00008
11	DEVLPNG	2182.69	423.091	5.15890	0.00000	457.070	4.77539	0.00000
12	NTREP	243.98	51.860	4.70453	0.00000	56.025	4.35480	0.00002
13	NRREP	-91.03	29.393	-3.09703	0.00205	31.753	-2.86680	0.00430
14	IMPACT	-25.62	11.417	-2.24402	0.02522	12.334	-2.07720	0.03823
15	DUMPUBT3	-354.65	158.095	-2.24328	0.02527	170.792	-2.07652	0.03830
16	DUMMULT	813.22	163.454	4.97524	0.00000	176.581	4.60538	0.00001
17	ICGMUL2	1226.46	355.925	3.44585	0.00061	384.509	3.18968	0.00150

(IV) Data used: all reportings of results
Response variable: time from close to submission

Not done: Too few observations

(V) Data used: all reportings of results
Response variable: time from receipt to publication

Not done: Too few observations

(VI) Data used: all reportings of results
Response variable: time from close to publication

PEST	VIF	SQRTVIF
0.55167	0.69566	0.83406

Obs	VARIABLE	BETA	SE_OLD	T_OLD	PROB_OLD	SE_NEW	T_NEW	PROB_NEW
1	INTERCPT	2036.56	301.806	6.74791	0.00000	361.852	5.62816	0.00000
2	NOQ	47.86	20.309	2.35650	0.01899	24.350	1.96546	0.05014
3	NCLOSE	-0.10	0.027	-3.77133	0.00019	0.032	-3.14552	0.00180
4	PRESENTD	-461.68	141.864	-3.25439	0.00125	170.089	-2.71436	0.00697
5	DUMCLN2	-405.87	151.038	-2.68719	0.00755	181.088	-2.24128	0.02563
6	DUMANS	416.20	145.726	2.85604	0.00454	174.719	2.38211	0.01774
7	DUMFL	692.59	259.924	2.66458	0.00806	311.637	2.22242	0.02689
8	NTREP	136.21	50.084	2.71971	0.00686	60.049	2.26840	0.02391
9	LOGPEST	330.63	83.769	3.94693	0.00010	100.436	3.29198	0.00110
10	DUMDIR1	276.15	171.197	1.61308	0.10762	205.257	1.34541	0.17936
11	DUMDIR2	339.43	220.885	1.53669	0.12527	264.831	1.28169	0.20079
12	LPDIR1	-390.87	103.790	-3.76600	0.00019	124.439	-3.14107	0.00183
13	LPDIR2	-386.69	145.513	-2.65741	0.00823	174.464	-2.21644	0.02730

APPENDIX XIII: FOR EACH ANALYSIS, OUTPUT FROM THE PRELIMINARY INVESTIGATION IMPOSING A STATIONARY *M*–DEPENDENT CORRELATION STRUCTURE

Key
PnEST = estimate of the correlation coefficient for elements distance *n* from the leading diagonal, *n* = 1,2,3,4 and $n \leq m$

Estimates of m-step correlation coefficients

(I) Data used: all mentions
Response variable: time from to submission

P1EST	P2EST	P3EST	P4EST
0.75833	0.83217	-0.54919	0.78506

(II) Data used: all mentions
Response variable: time from receipt to publication

P1EST	P2EST	P3EST	P4EST
0.53381	0.43775	-0.64546	0.73565

(III) Data used: all mentions
Response variable: time from close to publication

P1EST	P2EST	P3EST	P4EST
0.51840	0.31154	0.18015	-0.0511

(IV) Data used: all reportings of results
Response variable: Time from close to submission

Not done: Too few observations

(V) Data used: all reportings of results
Response variable: time from receipt to publication

Not done: Too few observations

(VI) Data used: all reportings of results
Response variable: time from close to publication

P1EST	P2EST	P3EST	P4EST
0.66706	0.52135	0.036238	-0.13473

APPENDIX XIV: PROGRAM APPLYING THE MANTEL TEST TO COMPARE AN ESTIMATED MATRIX HAVING A STATIONARY *M*–DEPENDENT CORRELATION STRUCTURE WITH THE MATRIX FORMED FROM THE RESIDUALS FROM THE INDEPENDENCE MODEL

Notes:

1. In order to use the same test to compare an estimated matrix having an autoregressive correlation structure with the matrix formed from the residuals from the independence model, replace P1EST=0.51840 P2EST=0.31154 P3EST=0.18015 P4EST=-0.0511 by P1EST=0.51840 P2EST=0.51840**2 P3EST=0.51840**3 P4EST=0.51840**4;
2. This example uses the response variable 'time from close of randomization to publication' and the 'all mentions' dataset.

```
* This program is called MANTEL_CLPUB_MSTEP.SAS;
* To apply Mantel test on the stationary m-dependent correlation structure
obtained using the first step of the GEE procedure to calculate a different
value of Rho for each diagonal i.e. all the elements distance 1 from the leading
diagonal have value P1EST, all elements distance 2 from the leading diagonal
have value P2EST, all elements distance 3 have value P3EST and those distance 4
have value P4EST;

LIBNAME LIBRARY '[LEUK]';
LIBNAME LEUKJR '[LEUK.JULIE.RESEARCH]';
*OPTIONS FORMDLIM='_' LS=132;
OPTIONS LS=80;
*OPTIONS LS=132 PS=43;
FILENAME MCLPUB '[LEUK.JULIE.RESEARCH]MANTEL_CLPUB.DAT';

* Let ENUM be the permutation reference number of the 5! = 120 matrices
obtained using the first step of the GEE procedure to estimate a value of Rho
for each diagonal (except the leading diagonal);
DATA TEMP;
 INPUT ENUM 3. +1 A 1. + 1 B 1. +1 C 1. +1 D 1. +1 E 1.;
CARDS;
  1 1 2 3 4 5
  2 1 2 3 5 4
  3 1 2 4 5 3
  4 1 2 4 3 5
  5 1 2 5 3 4
  6 1 2 5 4 3
  7 1 3 4 5 2
  8 1 3 4 2 5
  9 1 3 5 2 4
 10 1 3 5 4 2
 11 1 3 2 4 5
 12 1 3 2 5 4
 13 1 4 5 2 3
 14 1 4 5 3 2
 15 1 4 2 3 5
 16 1 4 2 5 3
 17 1 4 3 5 2
 18 1 4 3 2 5
 19 1 5 2 3 4
 20 1 5 2 4 3
 21 1 5 3 4 2
 22 1 5 3 2 4
 23 1 5 4 2 3
 24 1 5 4 3 2
 25 2 3 4 5 1
 26 2 3 4 1 5
 27 2 3 5 1 4
 28 2 3 5 4 1
 29 2 3 1 4 5
 30 2 3 1 5 4
 31 2 4 5 1 3
 32 2 4 5 3 1
 33 2 4 1 3 5
 34 2 4 1 5 3
```

```
 35 2 4 3 5 1
 36 2 4 3 1 5
 37 2 5 1 3 4
 38 2 5 1 4 3
 39 2 5 3 4 1
 40 2 5 3 1 4
 41 2 5 4 1 3
 42 2 5 4 3 1
 43 2 1 3 4 5
 44 2 1 3 5 4
 45 2 1 4 5 3
 46 2 1 4 3 5
 47 2 1 5 3 4
 48 2 1 5 4 3
 49 3 4 5 1 2
 50 3 4 5 2 1
 51 3 4 1 2 5
 52 3 4 1 5 2
 53 3 4 2 5 1
 54 3 4 2 1 5
 55 3 5 1 2 4
 56 3 5 1 4 2
 57 3 5 2 4 1
 58 3 5 2 1 4
 59 3 5 4 1 2
 60 3 5 4 2 1
 61 3 1 2 4 5
 62 3 1 2 5 4
 63 3 1 4 5 2
 64 3 1 4 2 5
 65 3 1 5 2 4
 66 3 1 5 4 2
 67 3 2 4 5 1
 68 3 2 4 1 5
 69 3 2 5 1 4
 70 3 2 5 4 1
 71 3 2 1 4 5
 72 3 2 1 5 4
 73 4 5 1 2 3
 74 4 5 1 3 2
 75 4 5 2 3 1
 76 4 5 2 1 3
 77 4 5 3 1 2
 78 4 5 3 2 1
 79 4 1 2 3 5
 80 4 1 2 5 3
 81 4 1 3 5 2
 82 4 1 3 2 5
 83 4 1 5 2 3
 84 4 1 5 3 2
 85 4 2 3 5 1
 86 4 2 3 1 5
 87 4 2 5 1 3
 88 4 2 5 3 1
 89 4 2 1 3 5
 90 4 2 1 5 3
 91 4 3 2 5 1
 92 4 3 2 1 5
 93 4 3 5 1 2
 94 4 3 5 2 1
 95 4 3 1 2 5
 96 4 3 1 5 2
 97 5 1 2 3 4
 98 5 1 2 4 3
 99 5 1 3 2 4
100 5 1 3 4 2
101 5 1 4 2 3
102 5 1 4 3 2
103 5 2 3 4 1
104 5 2 3 1 4
105 5 2 4 1 3
106 5 2 4 3 1
107 5 2 1 3 4
108 5 2 1 4 3
109 5 3 2 4 1
110 5 3 2 1 4
```

```
111 5 3 4 1 2
112 5 3 4 2 1
113 5 3 1 2 4
114 5 3 1 4 2
115 5 4 1 2 3
116 5 4 1 3 2
117 5 4 2 3 1
118 5 4 2 1 3
119 5 4 3 1 2
120 5 4 3 2 1
;

DATA TEMP1;
  SET TEMP;
ROW1=A;
ROW2=B;
ROW3=C;
ROW4=D;
ROW5=E;

COL1=A;
COL2=B;
COL3=C;
COL4=D;
COL5=E;

* Let EIJOLD be the corresponding elements of matrix E (the m-step correlation
structure produced using the first step of the GEE procedure. Here PXEST is
the estimation of the correlation coefficient between residuals from
publications X steps apart ie P1EST=0.51840 P2EST=0.31154 P3EST=0.18015
P4EST=-0.0511);

* To test;
/* E11OLD=11;
E12OLD=12;
E13OLD=13;
E14OLD=14;
E15OLD=15;
E21OLD=21;
E22OLD=22;
E23OLD=23;
E24OLD=24;
E25OLD=25;
E31OLD=31;
E32OLD=32;
E33OLD=33;
E34OLD=34;
E35OLD=35;
E41OLD=41;
E42OLD=42;
E43OLD=43;
E44OLD=44;
E45OLD=45;
E51OLD=51;
E52OLD=52;
E53OLD=53;
E54OLD=54;
E55OLD=55;*/

E11OLD=1.00000;
E22OLD=1.00000;
E33OLD=1.00000;
E44OLD=1.00000;
E55OLD=1.00000;

E12OLD=0.51840;
E21OLD=0.51840;
E23OLD=0.51840;
E32OLD=0.51840;
E34OLD=0.51840;
E43OLD=0.51840;
E45OLD=0.51840;
E54OLD=0.51840;

E13OLD=0.31154;
```

```
E31OLD=0.31154;
E24OLD=0.31154;
E42OLD=0.31154;
E35OLD=0.31154;
E53OLD=0.31154;

E14OLD=0.18015;
E41OLD=0.18015;
E25OLD=0.18015;
E52OLD=0.18015;

E15OLD=-0.0511;
E51OLD=-0.0511;

ARRAY ROWS {5} ROW1-ROW5;
ARRAY COLS {5} COL1-COL5;
ARRAY EIJNEW {5,5} E11NEW E12NEW E13NEW E14NEW E15NEW
                   E21NEW E22NEW E23NEW E24NEW E25NEW
                   E31NEW E32NEW E33NEW E34NEW E35NEW
                   E41NEW E42NEW E43NEW E44NEW E45NEW
                   E51NEW E52NEW E53NEW E54NEW E55NEW;
DO I=1 TO 5;
  DO J=1 TO 5;
 IF ROWS{I}=1 AND COLS{J}=1 THEN EIJNEW{I,J}=E11OLD;
 IF ROWS{I}=1 AND COLS{J}=2 THEN EIJNEW{I,J}=E12OLD;
 IF ROWS{I}=1 AND COLS{J}=3 THEN EIJNEW{I,J}=E13OLD;
 IF ROWS{I}=1 AND COLS{J}=4 THEN EIJNEW{I,J}=E14OLD;
 IF ROWS{I}=1 AND COLS{J}=5 THEN EIJNEW{I,J}=E15OLD;

 IF ROWS{I}=2 AND COLS{J}=1 THEN EIJNEW{I,J}=E21OLD;
 IF ROWS{I}=2 AND COLS{J}=2 THEN EIJNEW{I,J}=E22OLD;
 IF ROWS{I}=2 AND COLS{J}=3 THEN EIJNEW{I,J}=E23OLD;
 IF ROWS{I}=2 AND COLS{J}=4 THEN EIJNEW{I,J}=E24OLD;
 IF ROWS{I}=2 AND COLS{J}=5 THEN EIJNEW{I,J}=E25OLD;

 IF ROWS{I}=3 AND COLS{J}=1 THEN EIJNEW{I,J}=E31OLD;
 IF ROWS{I}=3 AND COLS{J}=2 THEN EIJNEW{I,J}=E32OLD;
 IF ROWS{I}=3 AND COLS{J}=3 THEN EIJNEW{I,J}=E33OLD;
 IF ROWS{I}=3 AND COLS{J}=4 THEN EIJNEW{I,J}=E34OLD;
 IF ROWS{I}=3 AND COLS{J}=5 THEN EIJNEW{I,J}=E35OLD;

 IF ROWS{I}=4 AND COLS{J}=1 THEN EIJNEW{I,J}=E41OLD;
 IF ROWS{I}=4 AND COLS{J}=2 THEN EIJNEW{I,J}=E42OLD;
 IF ROWS{I}=4 AND COLS{J}=3 THEN EIJNEW{I,J}=E43OLD;
 IF ROWS{I}=4 AND COLS{J}=4 THEN EIJNEW{I,J}=E44OLD;
 IF ROWS{I}=4 AND COLS{J}=5 THEN EIJNEW{I,J}=E45OLD;

 IF ROWS{I}=5 AND COLS{J}=1 THEN EIJNEW{I,J}=E51OLD;
 IF ROWS{I}=5 AND COLS{J}=2 THEN EIJNEW{I,J}=E52OLD;
 IF ROWS{I}=5 AND COLS{J}=3 THEN EIJNEW{I,J}=E53OLD;
 IF ROWS{I}=5 AND COLS{J}=4 THEN EIJNEW{I,J}=E54OLD;
 IF ROWS{I}=5 AND COLS{J}=5 THEN EIJNEW{I,J}=E55OLD;
  END;
END;
FILE MCLPUB;
  PUT
                   @10 E11NEW @20 E12NEW @30 E13NEW @40 E14NEW @50 E15NEW /
                   @10 E21NEW @20 E22NEW @30 E23NEW @40 E24NEW @50 E25NEW /
@1 'E' ENUM ' = ' @10 E31NEW @20 E32NEW @30 E33NEW @40 E34NEW @50 E35NEW /
                   @10 E41NEW @20 E42NEW @30 E43NEW @40 E44NEW @50 E45NEW /
                   @10 E51NEW @20 E52NEW @30 E53NEW @40 E54NEW @50 E55NEW //;

* Let MIJ be the elements of matrix M (the original original correlation
matrix produced from SAS);

M11=1.00000;
M12=0.35271;
M13=0.16092;
M14=0.16406;
M15=-0.05110;
M21=0.35271;
M22=1.00000;
M23=0.46667;
M24=0.40235;
M25=0.20546;
M31=0.16092;
```

```
M32=0.46667;
M33=1.00000;
M34=0.66755;
M35=0.45929;
M41=0.16406;
M42=0.40235;
M43=0.66755;
M44=1.00000;
M45=0.75220;
M51=-0.05110;
M52=0.20546;
M53=0.45929;
M54=0.75220;
M55=1.00000;

* Mantel's test statistic Z= (sum i=2 to n, sum j=1 to i-1) MijEij. In this
case Z = (M21*E21) + (M31*E31) + (M32*E32) + (M41*E41) + (M42*E42) + (M43*E43)
+ (M51*E51) + (M52*E52) + (M53*E53) + (M54*E54);

Z21=M21*E21NEW;
Z31=M31*E31NEW;
Z32=M32*E32NEW;
Z41=M41*E41NEW;
Z42=M42*E42NEW;
Z43=M43*E43NEW;
Z51=M51*E51NEW;
Z52=M52*E52NEW;
Z53=M53*E53NEW;
Z54=M54*E54NEW;

Z=Z21+Z31+Z32+Z41+Z42+Z43+Z51+Z52+Z53+Z54;

* Alternative test statistic r=correlation between lower diagonal elements of M
and E. Note: This is Pearson's correlation coefficient (parametric test, a
opposed to Spearman's which is a non-parametric test. PROC CORR would provide
the same value of r);

MBAR=(M21+M31+M32+M41+M42+M43+M51+M52+M53+M54)/10;
EBAR=(E21NEW+E31NEW+E32NEW+E41NEW+E42NEW+E43NEW+E51NEW+E52NEW+E53NEW+E54NEW)/10;
RTOP=Z-(10*MBAR*EBAR);
RBOTM=(M21**2)+ (M31**2) + (M32**2) + (M41**2) + (M42**2) + (M43**2) +(M51**2) +
(M52**2) + (M53**2) + (M54**2) -10*(MBAR**2);
RBOTE=(E21NEW**2) + (E31NEW**2) + (E32NEW**2) + (E41NEW**2) + (E42NEW**2) +
(E43NEW**2) + (E51NEW**2) + (E52NEW**2) + (E53NEW**2) + (E54NEW**2)
-10*(EBAR**2);
R=RTOP/((RBOTM*RBOTE)**0.5);

PROC SORT;
  BY Z;
PROC PRINT N;
  VAR ENUM Z R;
RUN;
ENDSAS;
```

APPENDIX XV: OUTPUT FROM GEE ANALYSIS APPLIED TO THE LARGEST DATASET WITH EACH CORRELATION STRUCTURE IMPOSED, GIVING THE WORKING CORRELATION MATRIX, PARAMETER (β) ESTIMATES AND ESTIMATES OF THEIR STANDARD ERRORS, *Z*- AND *P*-VALUES

Note: A list of the indicator variables used is given at the end of this appendix.

Correlation structure: independence

The GENMOD Procedure
Algorithm converged

Working Correlation Matrix

	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9
Row1	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Row2	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Row3	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Row4	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Row5	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
Row6	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
Row7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000
Row8	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000
Row9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000
Row10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Row11	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Row12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Row13	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Col10	Col11	Col12	Col13
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
1.0000	0.0000	0.0000	0.0000
0.0000	1.0000	0.0000	0.0000
0.0000	0.0000	1.0000	0.0000
0.0000	0.0000	0.0000	1.0000

Analysis Of GEE Parameter Estimates
Model-Based Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	2344.831	266.1429	1823.200	2866.461	8.81	<.0001
NSTART	-0.1508	0.0286	-0.2067	-0.0948	-5.28	<.0001
DURRAN	-0.4398	0.1041	-0.6437	-0.2358	-4.23	<.0001
DUMCGRP1	-535.114	149.1136	-827.371	-242.857	-3.59	0.0003
DUMCGRP2	23.5089	253.8332	-473.995	521.0129	0.09	0.9262
DUMENG	-1048.55	364.2719	-1762.51	-334.595	-2.88	0.0040
PRESENTD	-637.945	118.3405	-869.888	-406.002	-5.39	<.0001
DUMCLN1	444.4775	149.1610	152.1273	736.8278	2.98	0.0029
DUMCLN2	438.2254	145.7955	152.4715	723.9793	3.01	0.0026
DUMANS	523.7168	122.0783	284.4477	762.9860	4.29	<.0001
DEVLPNG	2182.688	423.0914	1353.444	3011.932	5.16	<.0001
NTREP	243.9777	51.8602	142.3337	345.6218	4.70	<.0001
NRREP	-91.0305	29.3928	-148.639	-33.4216	-3.10	0.0020
IMPACT	-25.6206	11.4173	-47.9980	-3.2431	-2.24	0.0248
DUMPUBT3	-354.653	158.0953	-664.514	-44.7916	-2.24	0.0249
DUMMULT	813.2225	163.4538	492.8590	1133.586	4.98	<.0001
ICGMUL2	1226.462	355.9247	528.8622	1924.062	3.45	0.0006
Scale	1248.802	.	.	.	.	.

NOTE: The scale parameter for GEE estimation was computed as the square root of the normalized Pearson's chi-square.

Correlation structure: exchangeable

Algorithm converged.

Working Correlation Matrix

	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9
Row1	1.0000	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831
Row2	0.0831	1.0000	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831
Row3	0.0831	0.0831	1.0000	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831
Row4	0.0831	0.0831	0.0831	1.0000	0.0831	0.0831	0.0831	0.0831	0.0831
Row5	0.0831	0.0831	0.0831	0.0831	1.0000	0.0831	0.0831	0.0831	0.0831
Row6	0.0831	0.0831	0.0831	0.0831	0.0831	1.0000	0.0831	0.0831	0.0831
Row7	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831	1.0000	0.0831	0.0831
Row8	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831	1.0000	0.0831
Row9	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831	1.0000
Row10	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831
Row11	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831
Row12	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831
Row13	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831	0.0831

Col10	Col11	Col12	Col13
0.0831	0.0831	0.0831	0.0831
0.0831	0.0831	0.0831	0.0831
0.0831	0.0831	0.0831	0.0831
0.0831	0.0831	0.0831	0.0831
0.0831	0.0831	0.0831	0.0831
0.0831	0.0831	0.0831	0.0831
0.0831	0.0831	0.0831	0.0831
0.0831	0.0831	0.0831	0.0831
0.0831	0.0831	0.0831	0.0831
1.0000	0.0831	0.0831	0.0831
0.0831	1.0000	0.0831	0.0831
0.0831	0.0831	1.0000	0.0831
0.0831	0.0831	0.0831	1.0000

Analysis Of GEE Parameter Estimates

Model-Based Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	2401.406	285.4231	1841.987	2960.825	8.41	<.0001
NSTART	-0.1541	0.0313	-0.2154	-0.0927	-4.92	<.0001
DURRAN	-0.4351	0.1145	-0.6595	-0.2107	-3.80	0.0001
DUMCGRP1	-512.433	162.0403	-830.026	-194.840	-3.16	0.0016
DUMCGRP2	46.4225	274.0935	-490.791	583.6359	0.17	0.8655
DUMENG	-1123.73	359.4823	-1828.30	-419.160	-3.13	0.0018
PRESENTD	-639.023	116.7415	-867.832	-410.213	-5.47	<.0001
DUMCLN1	461.6988	150.3223	167.0725	756.3251	3.07	0.0021
DUMCLN2	432.5392	145.8151	146.7470	718.3315	2.97	0.0030
DUMANS	454.5659	120.9839	217.4417	691.6900	3.76	0.0002
DEVLPNG	2197.779	438.8265	1337.695	3057.864	5.01	<.0001
NTREP	245.0108	54.6656	137.8683	352.1534	4.48	<.0001
NRREP	-88.0270	31.6647	-150.089	-25.9653	-2.78	0.0054
IMPACT	-28.6904	11.3770	-50.9888	-6.3920	-2.52	0.0117
DUMPUBT3	-373.240	155.5935	-678.197	-68.2818	-2.40	0.0164
DUMMULT	791.4276	185.9465	426.9793	1155.876	4.26	<.0001
ICGMUL2	1167.130	395.0436	392.8586	1941.401	2.95	0.0031
Scale	1249.511	.	.	.	.	.

NOTE: The scale parameter for GEE estimation was computed as the square root of the normalized Pearson's chi-square.

Correlation structure: stationary *m*-dependent, *m* = 3

Algorithm converged.

Working Correlation Matrix

	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9
Row1	1.0000	0.4545	0.1445	-0.1323	0.0000	0.0000	0.0000	0.0000	0.0000
Row2	0.4545	1.0000	0.4545	0.1445	-0.1323	0.0000	0.0000	0.0000	0.0000
Row3	0.1445	0.4545	1.0000	0.4545	0.1445	-0.1323	0.0000	0.0000	0.0000
Row4	-0.1323	0.1445	0.4545	1.0000	0.4545	0.1445	-0.1323	0.0000	0.0000
Row5	0.0000	-0.1323	0.1445	0.4545	1.0000	0.4545	0.1445	-0.1323	0.0000
Row6	0.0000	0.0000	-0.1323	0.1445	0.4545	1.0000	0.4545	0.1445	-0.1323
Row7	0.0000	0.0000	0.0000	-0.1323	0.1445	0.4545	1.0000	0.4545	0.1445
Row8	0.0000	0.0000	0.0000	0.0000	-0.1323	0.1445	0.4545	1.0000	0.4545
Row9	0.0000	0.0000	0.0000	0.0000	0.0000	-0.1323	0.1445	0.4545	1.0000
Row10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.1323	0.1445	0.4545
Row11	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.1323	0.1445
Row12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.1323
Row13	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Col10	Col11	Col12	Col13
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.1445	-0.1323	0.0000	0.0000
-0.1323	0.0000	0.0000	0.0000
0.4545	0.1445	-0.1323	0.0000
1.0000	0.4545	0.1445	-0.1323
0.4545	1.0000	0.4545	0.1445
0.1445	0.4545	1.0000	0.4545
-0.1323	0.1445	0.4545	1.0000

Analysis Of GEE Parameter Estimates

Model-Based Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	2592.377	296.2506	2011.737	3173.018	8.75	<.0001
NSTART	-0.1685	0.0336	-0.2345	-0.1026	-5.01	<.0001
DURRAN	-0.4256	0.1246	-0.6697	-0.1814	-3.42	0.0006
DUMCGRP1	-485.141	176.6197	-831.309	-138.973	-2.75	0.0060
DUMCGRP2	160.3201	301.2922	-430.202	750.8420	0.53	0.5947
DUMENG	-627.782	310.7131	-1236.77	-18.7951	-2.02	0.0433
PRESENTD	-522.652	102.5747	-723.695	-321.609	-5.10	<.0001
DUMCLN1	428.3798	139.4245	155.1128	701.6468	3.07	0.0021
DUMCLN2	279.7651	129.9932	24.9830	534.5471	2.15	0.0314
DUMANS	255.2884	98.0494	63.1152	447.4617	2.60	0.0092
DEVLPNG	2292.584	474.5052	1362.570	3222.597	4.83	<.0001
NTREP	183.2207	51.3591	82.5588	283.8827	3.57	0.0004
NRREP	-60.2162	30.5487	-120.091	-0.3420	-1.97	0.0487
IMPACT	-25.0378	9.7081	-44.0654	-6.0101	-2.58	0.0099
DUMPUBT3	-277.036	138.3691	-548.234	-5.8373	-2.00	0.0453
DUMMULT	805.6502	202.8144	408.1412	1203.159	3.97	<.0001
ICGMUL2	1239.207	437.8491	381.0382	2097.375	2.83	0.0047
Scale	1260.759	.	.	.	.	.

NOTE: The scale parameter for GEE estimation was computed as the square root of the normalized Pearson's chi-square.

Correlation structure: stationary *m*-dependent, *m* = 2

Algorithm converged.

Working Correlation Matrix

	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9
Row1	1.0000	0.4514	0.1436	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Row2	0.4514	1.0000	0.4514	0.1436	0.0000	0.0000	0.0000	0.0000	0.0000
Row3	0.1436	0.4514	1.0000	0.4514	0.1436	0.0000	0.0000	0.0000	0.0000
Row4	0.0000	0.1436	0.4514	1.0000	0.4514	0.1436	0.0000	0.0000	0.0000
Row5	0.0000	0.0000	0.1436	0.4514	1.0000	0.4514	0.1436	0.0000	0.0000
Row6	0.0000	0.0000	0.0000	0.1436	0.4514	1.0000	0.4514	0.1436	0.0000
Row7	0.0000	0.0000	0.0000	0.0000	0.1436	0.4514	1.0000	0.4514	0.1436
Row8	0.0000	0.0000	0.0000	0.0000	0.0000	0.1436	0.4514	1.0000	0.4514
Row9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1436	0.4514	1.0000
Row10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1436	0.4514
Row11	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1436
Row12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Row13	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Col10	Col11	Col12	Col13
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.1436	0.0000	0.0000	0.0000
0.4514	0.1436	0.0000	0.0000
1.0000	0.4514	0.1436	0.0000
0.4514	1.0000	0.4514	0.1436
0.1436	0.4514	1.0000	0.4514
0.0000	0.1436	0.4514	1.0000

Analysis Of GEE Parameter Estimates

Model-Based Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	2578.432	302.0606	1986.405	3170.460	8.54	<.0001
NSTART	-0.1695	0.0343	-0.2367	-0.1022	-4.94	<.0001
DURRAN	-0.4244	0.1270	-0.6733	-0.1754	-3.34	0.0008
DUMCGRP1	-478.640	179.4447	-830.346	-126.935	-2.67	0.0076
DUMCGRP2	155.8623	304.6515	-441.244	752.9683	0.51	0.6089
DUMENG	-667.346	316.2724	-1287.23	-47.4632	-2.11	0.0349
PRESENTD	-528.536	102.2570	-728.956	-328.116	-5.17	<.0001
DUMCLN1	419.5579	141.1564	142.8964	696.2194	2.97	0.0030
DUMCLN2	317.6258	132.9789	56.9919	578.2598	2.39	0.0169
DUMANS	287.9122	103.5654	84.9277	490.8967	2.78	0.0054
DEVLPNG	2263.555	475.4807	1331.630	3195.480	4.76	<.0001
NTREP	199.5583	53.4102	94.8763	304.2402	3.74	0.0002
NRREP	-66.3383	31.6630	-128.397	-4.2799	-2.10	0.0362
IMPACT	-25.1684	10.1114	-44.9864	-5.3504	-2.49	0.0128
DUMPUBT3	-301.566	140.2622	-576.475	-26.6576	-2.15	0.0316
DUMMULT	788.2816	207.5016	381.5858	1194.977	3.80	0.0001
ICGMUL2	1216.753	447.7949	339.0914	2094.415	2.72	0.0066
Scale	1258.024	.	.	.	.	.

NOTE: The scale parameter for GEE estimation was computed as the square root of the normalized Pearson's chi-square.

Correlation structure: stationary *m*-dependent, *m* = 1

Algorithm converged.

Working Correlation Matrix

	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9
Row1	1.0000	0.4484	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Row2	0.4484	1.0000	0.4484	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Row3	0.0000	0.4484	1.0000	0.4484	0.0000	0.0000	0.0000	0.0000	0.0000
Row4	0.0000	0.0000	0.4484	1.0000	0.4484	0.0000	0.0000	0.0000	0.0000
Row5	0.0000	0.0000	0.0000	0.4484	1.0000	0.4484	0.0000	0.0000	0.0000
Row6	0.0000	0.0000	0.0000	0.0000	0.4484	1.0000	0.4484	0.0000	0.0000
Row7	0.0000	0.0000	0.0000	0.0000	0.0000	0.4484	1.0000	0.4484	0.0000
Row8	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.4484	1.0000	0.4484
Row9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.4484	1.0000
Row10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.4484
Row11	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Row12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Row13	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Col10	Col11	Col12	Col13
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.4484	0.0000	0.0000	0.0000
1.0000	0.4484	0.0000	0.0000
0.4484	1.0000	0.4484	0.0000
0.0000	0.4484	1.0000	0.4484
0.0000	0.0000	0.4484	1.0000

Analysis Of GEE Parameter Estimates

Model-Based Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	2586.432	290.1848	2017.680	3155.184	8.91	<.0001
NSTART	-0.1692	0.0332	-0.2343	-0.1041	-5.09	<.0001
DURRAN	-0.4393	0.1229	-0.6801	-0.1984	-3.58	0.0004
DUMCGRP1	-461.283	173.7127	-801.753	-120.812	-2.66	0.0079
DUMCGRP2	192.4554	295.9115	-387.521	772.4313	0.65	0.5154
DUMENG	-614.185	294.0555	-1190.52	-37.8470	-2.09	0.0367
PRESENTD	-461.486	91.7993	-641.409	-281.563	-5.03	<.0001
DUMCLN1	406.5547	134.8488	142.2559	670.8535	3.01	0.0026
DUMCLN2	319.6508	126.8129	71.1021	568.1995	2.52	0.0117
DUMANS	293.9733	98.9132	100.1070	487.8396	2.97	0.0030
DEVLPNG	2264.908	471.2528	1341.270	3188.547	4.81	<.0001
NTREP	184.1264	51.4340	83.3175	284.9353	3.58	0.0003
NRREP	-69.6618	30.5387	-129.517	-9.8070	-2.28	0.0225
IMPACT	-22.4326	9.6879	-41.4206	-3.4446	-2.32	0.0206
DUMPUBT3	-314.162	133.9088	-576.618	-51.7054	-2.35	0.0190
DUMMULT	798.5519	199.4591	407.6192	1189.485	4.00	<.0001
ICGMUL2	1226.130	433.1791	377.1145	2075.145	2.83	0.0046
Scale	1259.627	.	.	.	.	.

NOTE: The scale parameter for GEE estimation was computed as the square root of the normalized Pearson's chi-square.

Correlation structure: autoregressive [AR(1)]

Algorithm converged.

Working Correlation Matrix

	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10	Col11	Col12	Col13
Row1	1.0000	0.4507	0.2031	0.0915	0.0413	0.0186	0.0084	0.0038	0.0017	0.0008	0.0003	0.0002	0.0001
Row2	0.4507	1.0000	0.4507	0.2031	0.0915	0.0413	0.0186	0.0084	0.0038	0.0017	0.0008	0.0003	0.0002
Row3	0.2031	0.4507	1.0000	0.4507	0.2031	0.0915	0.0413	0.0186	0.0084	0.0038	0.0017	0.0008	0.0003
Row4	0.0915	0.2031	0.4507	1.0000	0.4507	0.2031	0.0915	0.0413	0.0186	0.0084	0.0038	0.0017	0.0008
Row5	0.0413	0.0915	0.2031	0.4507	1.0000	0.4507	0.2031	0.0915	0.0413	0.0186	0.0084	0.0038	0.0017
Row6	0.0186	0.0413	0.0915	0.2031	0.4507	1.0000	0.4507	0.2031	0.0915	0.0413	0.0186	0.0084	0.0038
Row7	0.0084	0.0186	0.0413	0.0915	0.2031	0.4507	1.0000	0.4507	0.2031	0.0915	0.0413	0.0186	0.0084
Row8	0.0038	0.0084	0.0186	0.0413	0.0915	0.2031	0.4507	1.0000	0.4507	0.2031	0.0915	0.0413	0.0186
Row9	0.0017	0.0038	0.0084	0.0186	0.0413	0.0915	0.2031	0.4507	1.0000	0.4507	0.2031	0.0915	0.0413
Row10	0.0008	0.0017	0.0038	0.0084	0.0186	0.0413	0.0915	0.2031	0.4507	1.0000	0.4507	0.2031	0.0915
Row11	0.0003	0.0008	0.0017	0.0038	0.0084	0.0186	0.0413	0.0915	0.2031	0.4507	1.0000	0.4507	0.2031
Row12	0.0002	0.0003	0.0008	0.0017	0.0038	0.0084	0.0186	0.0413	0.0915	0.2031	0.4507	1.0000	0.4507
Row13	0.0001	0.0002	0.0003	0.0008	0.0017	0.0038	0.0084	0.0186	0.0413	0.0915	0.2031	0.4507	1.0000

Analysis Of GEE Parameter Estimates

Model-Based Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	2589.813	309.0583	1984.070	3195.557	8.38	<.0001
NSTART	-0.1714	0.0352	-0.2404	-0.1024	-4.87	<.0001
DURRAN	-0.4207	0.1301	-0.6757	-0.1656	-3.23	0.0012
DUMCGRP1	-474.045	183.3097	-833.325	-114.764	-2.59	0.0097
DUMCGRP2	151.9198	310.1821	-456.026	759.8655	0.49	0.6243
DUMENG	-709.793	318.7329	-1334.50	-85.0883	-2.23	0.0260
PRESENTD	-546.968	102.7414	-748.337	-345.598	-5.32	<.0001
DUMCLN1	423.8141	141.6190	146.2460	701.3822	2.99	0.0028
DUMCLN2	330.8883	133.8130	68.6197	593.1569	2.47	0.0134
DUMANS	293.6093	105.0567	87.7019	499.5167	2.79	0.0052
DEVLPNG	2256.449	478.0560	1319.476	3193.421	4.72	<.0001
NTREP	204.5176	54.3113	98.0695	310.9657	3.77	0.0002
NRREP	-67.2755	32.3408	-130.662	-3.8887	-2.08	0.0375
IMPACT	-25.1378	10.1943	-45.1182	-5.1574	-2.47	0.0137
DUMPUBT3	-316.564	139.8623	-590.689	-42.4390	-2.26	0.0236
DUMMULT	777.0006	214.1923	357.1915	1196.810	3.63	0.0003
ICGMUL2	1189.320	460.4376	286.8788	2091.761	2.58	0.0098
Scale	1257.180	.	.	.	.	.

NOTE: The scale parameter for GEE estimation was computed as the square root of the normalized Pearson's chi-square.

Indicator variables

```
DUMMULT = 1 if MULTIC = Yes (multi-centre)
DUMMULT = 0 if MULTIC = Limited or No (single-centre)

DUMCGRP1 = 1 if CGROUP = Europe
DUMCGRP1 = 0 if CGROUP = America or Other

DUMCGRP2 = 1 if CGROUP = Other
DUMCGRP2 = 0 if CGROUP = America or Europe

DUMENG = 1 if published in full in a language other than English, with an English abstract
DUMENG = 0 if published in full in English

DUMCLN1 = 1 if CLNSG = Yes
DUMCLN1 = 0 if CLNSG = No, Possibly or X (not reported)

DUMCLN2 = 1 if CLNSG = X (not reported)
DUMCLN2 = 0 if CLNSG = Yes, No or Possibly

DUMANS = 1 if ANSWER = Yes or No
DUMANS = 0 if ANSWER = X (not reported)

PUBTYPE3 = 1 if PUBTYPE = Meeting abstract
PUBTYPE3 = 0 if PUBTYPE = Journal or Book
```

Interaction term

```
ICGMUL2 = 1 if non-US trial with less than 5 centres participating
ICGMUL2 = 0 otherwise
```

APPENDIX XVI: OUTPUT FOR THE LARGEST DATASET COMPARING HOW FOR EACH VARIABLE, THE *P*-VALUE CHANGES WHEN DIFFERENT CORRELATION STRUCTURES ARE IMPOSED

---------- Parameter number=1 ----------

Parameter	Correlation structure	Beta estimate	SE of beta estimate	Z value	Prob>Z
INTERCPT	INDP	2344.83	266.143	8.81	<.0001
INTERCPT	EXCH	2401.41	285.423	8.41	<.0001
INTERCPT	M3	2592.38	296.251	8.75	<.0001
INTERCPT	M2	2578.43	302.061	8.54	<.0001
INTERCPT	M1	2586.43	290.185	8.91	<.0001
INTERCPT	AUTO	2589.81	309.058	8.38	<.0001

---------- Parameter number=2 ----------

Parameter	Correlation structure	Beta estimate	SE of beta estimate	Z value	Prob>Z
NSTART	INDP	-0.1508	0.0286	-5.28	<.0001
NSTART	EXCH	-0.1541	0.0313	-4.92	<.0001
NSTART	M3	-0.1685	0.0336	-5.01	<.0001
NSTART	M2	-0.1695	0.0343	-4.94	<.0001
NSTART	M1	-0.1692	0.0332	-5.09	<.0001
NSTART	AUTO	-0.1714	0.0352	-4.87	<.0001

---------- Parameter number=3 ----------

Parameter	Correlation structure	Beta estimate	SE of beta estimate	Z value	Prob>Z
DURRAN	INDP	-0.4398	0.1041	-4.23	<.0001
DURRAN	EXCH	-0.4351	0.1145	-3.80	0.0001
DURRAN	M3	-0.4256	0.1246	-3.42	0.0006
DURRAN	M2	-0.4244	0.1270	-3.34	0.0008
DURRAN	M1	-0.4393	0.1229	-3.58	0.0004
DURRAN	AUTO	-0.4207	0.1301	-3.23	0.0012

---------- Parameter number=4 ----------

Parameter	Correlation structure	Beta estimate	SE of beta estimate	Z value	Prob>Z
DUMCGRP1	INDP	-535.114	149.114	-3.59	0.0003
DUMCGRP1	EXCH	-512.433	162.040	-3.16	0.0016
DUMCGRP1	M3	-485.141	176.620	-2.75	0.0060
DUMCGRP1	M2	-478.640	179.445	-2.67	0.0076
DUMCGRP1	M1	-461.283	173.713	-2.66	0.0079
DUMCGRP1	AUTO	-474.045	183.310	-2.59	0.0097

---------- Parameter number=5 ----------

Parameter	Correlation structure	Beta estimate	SE of beta estimate	Z value	Prob>Z
DUMCGRP2	INDP	23.509	253.833	0.09	0.9262
DUMCGRP2	EXCH	46.423	274.094	0.17	0.8655
DUMCGRP2	M3	160.320	301.292	0.53	0.5947
DUMCGRP2	M2	155.862	304.652	0.51	0.6089
DUMCGRP2	M1	192.455	295.912	0.65	0.5154
DUMCGRP2	AUTO	151.920	310.182	0.49	0.6243

---------- Parameter number=6 ----------

Parameter	Correlation structure	Beta estimate	SE of beta estimate	Z value	Prob>Z
DUMENG	INDP	-1048.55	364.272	-2.88	0.0040
DUMENG	EXCH	-1123.73	359.482	-3.13	0.0018
DUMENG	M3	-627.78	310.713	-2.02	0.0433
DUMENG	M2	-667.35	316.272	-2.11	0.0349
DUMENG	M1	-614.19	294.056	-2.09	0.0367
DUMENG	AUTO	-709.79	318.733	-2.23	0.0260

---------- Parameter number=7 ----------

Parameter	Correlation structure	Beta estimate	SE of beta estimate	Z value	Prob>Z
PRESENTD	INDP	-637.945	118.341	-5.39	<.0001
PRESENTD	EXCH	-639.023	116.742	-5.47	<.0001
PRESENTD	M3	-522.652	102.575	-5.10	<.0001
PRESENTD	M2	-528.536	102.257	-5.17	<.0001
PRESENTD	M1	-461.486	91.799	-5.03	<.0001
PRESENTD	AUTO	-546.968	102.741	-5.32	<.0001

---------- Parameter number=8 ----------

Parameter	Correlation structure	Beta estimate	SE of beta estimate	Z value	Prob>Z
DUMCLN1	INDP	444.478	149.161	2.98	0.0029
DUMCLN1	EXCH	461.699	150.322	3.07	0.0021
DUMCLN1	M3	428.380	139.425	3.07	0.0021
DUMCLN1	M2	419.558	141.156	2.97	0.0030
DUMCLN1	M1	406.555	134.849	3.01	0.0026
DUMCLN1	AUTO	423.814	141.619	2.99	0.0028

---------- Parameter number=9 ----------

Parameter	Correlation structure	Beta estimate	SE of beta estimate	Z value	Prob>Z
DUMCLN2	INDP	438.225	145.796	3.01	0.0026
DUMCLN2	EXCH	432.539	145.815	2.97	0.0030
DUMCLN2	M3	279.765	129.993	2.15	0.0314
DUMCLN2	M2	317.626	132.979	2.39	0.0169
DUMCLN2	M1	319.651	126.813	2.52	0.0117
DUMCLN2	AUTO	330.888	133.813	2.47	0.0134

---------- Parameter number=10 ----------

Parameter	Correlation structure	Beta estimate	SE of beta estimate	Z value	Prob>Z
DUMANS	INDP	523.717	122.078	4.29	<.0001
DUMANS	EXCH	454.566	120.984	3.76	0.0002
DUMANS	M3	255.288	98.049	2.60	0.0092
DUMANS	M2	287.912	103.565	2.78	0.0054
DUMANS	M1	293.973	98.913	2.97	0.0030
DUMANS	AUTO	293.609	105.057	2.79	0.0052

---------- Parameter number=11 ----------

Parameter	Correlation structure	Beta estimate	SE of beta estimate	Z value	Prob>Z
DEVLPNG	INDP	2182.69	423.091	5.16	<.0001
DEVLPNG	EXCH	2197.78	438.827	5.01	<.0001
DEVLPNG	M3	2292.58	474.505	4.83	<.0001
DEVLPNG	M2	2263.56	475.481	4.76	<.0001
DEVLPNG	M1	2264.91	471.253	4.81	<.0001
DEVLPNG	AUTO	2256.45	478.056	4.72	<.0001

---------- Parameter number=12 ----------

Parameter	Correlation structure	Beta estimate	SE of beta estimate	Z value	Prob>Z
NTREP	INDP	243.978	51.8602	4.70	<.0001
NTREP	EXCH	245.011	54.6656	4.48	<.0001
NTREP	M3	183.221	51.3591	3.57	0.0004
NTREP	M2	199.558	53.4102	3.74	0.0002
NTREP	M1	184.126	51.4340	3.58	0.0003
NTREP	AUTO	204.518	54.3113	3.77	0.0002

---------- Parameter number=13 ----------

Parameter	Correlation structure	Beta estimate	SE of beta estimate	Z value	Prob>Z
NRREP	INDP	-91.0305	29.3928	-3.10	0.0020
NRREP	EXCH	-88.0270	31.6647	-2.78	0.0054
NRREP	M3	-60.2162	30.5487	-1.97	0.0487
NRREP	M2	-66.3383	31.6630	-2.10	0.0362
NRREP	M1	-69.6618	30.5387	-2.28	0.0225
NRREP	AUTO	-67.2755	32.3408	-2.08	0.0375

---------- Parameter number=14 ----------

Parameter	Correlation structure	Beta estimate	SE of beta estimate	Z value	Prob>Z
IMPACT	INDP	-25.6206	11.4173	-2.24	0.0248
IMPACT	EXCH	-28.6904	11.3770	-2.52	0.0117
IMPACT	M3	-25.0378	9.7081	-2.58	0.0099
IMPACT	M2	-25.1684	10.1114	-2.49	0.0128
IMPACT	M1	-22.4326	9.6879	-2.32	0.0206
IMPACT	AUTO	-25.1378	10.1943	-2.47	0.0137

---------- Parameter number=15 ----------

Parameter	Correlation structure	Beta estimate	SE of beta estimate	Z value	Prob>Z
DUMPUBT3	INDP	-354.653	158.095	-2.24	0.0249
DUMPUBT3	EXCH	-373.240	155.594	-2.40	0.0164
DUMPUBT3	M3	-277.036	138.369	-2.00	0.0453
DUMPUBT3	M2	-301.566	140.262	-2.15	0.0316
DUMPUBT3	M1	-314.162	133.909	-2.35	0.0190
DUMPUBT3	AUTO	-316.564	139.862	-2.26	0.0236

---------- Parameter number=16 ----------

Parameter	Correlation structure	Beta estimate	SE of beta estimate	Z value	Prob>Z
DUMMULT	INDP	813.223	163.454	4.98	<.0001
DUMMULT	EXCH	791.428	185.947	4.26	<.0001
DUMMULT	M3	805.650	202.814	3.97	<.0001
DUMMULT	M2	788.282	207.502	3.80	0.0001
DUMMULT	M1	798.552	199.459	4.00	<.0001
DUMMULT	AUTO	777.001	214.192	3.63	0.0003

---------- Parameter number=17 ----------

Parameter	Correlation structure	Beta estimate	SE of beta estimate	Z value	Prob>Z
ICGMUL2	INDP	1226.46	355.925	3.45	0.0006
ICGMUL2	EXCH	1167.13	395.044	2.95	0.0031
ICGMUL2	M3	1239.21	437.849	2.83	0.0047
ICGMUL2	M2	1216.75	447.795	2.72	0.0066
ICGMUL2	M1	1226.13	433.179	2.83	0.0046
ICGMUL2	AUTO	1189.32	460.438	2.58	0.0098

APPENDIX XVII: OUTPUT FROM THE LINEAR MIXED EFFECTS ANALYSES

Note: The name of the structure as used in this thesis is followed by that used in the SAS procedure in brackets where this differs.

Covariance structure: independence (diagonal)

The Mixed Procedure
Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	2344.83	266.14	564	8.81	<.0001
NSTART	-0.1508	0.02855	564	-5.28	<.0001
DURRAN	-0.4398	0.1041	564	-4.23	<.0001
DUMCGRP1	-535.11	149.11	564	-3.59	0.0004
DUMCGRP2	23.5089	253.83	564	0.09	0.9262
DUMENG	-1048.55	364.27	564	-2.88	0.0041
PRESENTD	-637.95	118.34	564	-5.39	<.0001
DUMCLN1	444.48	149.16	564	2.98	0.0030
DUMCLN2	438.23	145.80	564	3.01	0.0028
DUMANS	523.72	122.08	564	4.29	<.0001
DEVLPNG	2182.69	423.09	564	5.16	<.0001
NTREP	243.98	51.8602	564	4.70	<.0001
NRREP	-91.0305	29.3928	564	-3.10	0.0021
IMPACT	-25.6206	11.4173	564	-2.24	0.0252
DUMPUBT3	-354.65	158.10	564	-2.24	0.0253
DUMMULT	813.22	163.45	564	4.98	<.0001
ICGMUL2	1226.46	355.92	564	3.45	0.0006

Covariance structure: exchangeable (compound symmetry)

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	2415.19	289.73	186	8.34	<.0001
NSTART	-0.1547	0.03192	186	-4.85	<.0001
DURRAN	-0.4341	0.1167	186	-3.72	0.0003
DUMCGRP1	-509.13	164.89	186	-3.09	0.0023
DUMCGRP2	51.1239	278.64	186	0.18	0.8546
DUMENG	-1138.99	358.57	378	-3.18	0.0016
PRESENTD	-639.94	116.45	378	-5.50	<.0001
DUMCLN1	467.20	150.59	378	3.10	0.0021
DUMCLN2	433.02	145.83	378	2.97	0.0032
DUMANS	441.39	120.79	378	3.65	0.0003
DEVLPNG	2201.15	442.99	186	4.97	<.0001
NTREP	245.14	55.2231	378	4.44	<.0001
NRREP	-87.9852	32.1093	378	-2.74	0.0064
IMPACT	-29.1786	11.3668	378	-2.57	0.0106
DUMPUBT3	-377.39	155.14	378	-2.43	0.0155
DUMMULT	786.06	190.52	186	4.13	<.0001
ICGMUL2	1154.85	403.65	186	2.86	0.0047

Covariance structure: autoregressive

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	2679.02	335.96	186	7.97	<.0001
NSTART	-0.1804	0.03914	186	-4.61	<.0001
DURRAN	-0.4158	0.1446	186	-2.88	0.0045
DUMCGRP1	-442.88	202.09	186	-2.19	0.0297
DUMCGRP2	203.47	341.55	186	0.60	0.5521
DUMENG	-610.83	294.97	378	-2.07	0.0391
PRESENTD	-508.62	94.1749	378	-5.40	<.0001
DUMCLN1	420.11	133.93	378	3.14	0.0018
DUMCLN2	310.96	125.01	378	2.49	0.0133
DUMANS	231.08	95.9865	378	2.41	0.0165
DEVLPNG	2274.07	514.36	186	4.42	<.0001
NTREP	184.81	53.3658	378	3.46	0.0006
NRREP	-58.9015	32.4150	378	-1.82	0.0700
IMPACT	-23.3669	9.4001	378	-2.49	0.0134
DUMPUBT3	-298.02	128.63	378	-2.32	0.0210
DUMMULT	760.73	242.27	186	3.14	0.0020
ICGMUL2	1171.89	518.39	186	2.26	0.0249

Covariance structure: stationary *m*–dependent, *m* = 3 (banded Toeplitz)

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	2798.35	346.96	186	8.07	<.0001
NSTART	-0.1962	0.04215	186	-4.66	<.0001
DURRAN	-0.4524	0.1554	186	-2.91	0.0040
DUMCGRP1	-390.28	216.94	186	-1.80	0.0736
DUMCGRP2	381.91	374.11	186	1.02	0.3087
DUMENG	-181.19	206.64	378	-0.88	0.3811
PRESENTD	-294.76	63.2705	378	-4.66	<.0001
DUMCLN1	308.29	96.1685	378	3.21	0.0015
DUMCLN2	180.88	83.1247	378	2.18	0.0302
DUMANS	135.47	59.6526	378	2.27	0.0237
DEVLPNG	2302.79	584.62	186	3.94	0.0001
NTREP	81.8924	38.9355	378	2.10	0.0361
NRREP	-33.5013	24.3684	378	-1.37	0.1700
IMPACT	-5.9639	5.9422	378	-1.00	0.3162
DUMPUBT3	-114.50	86.1129	378	-1.33	0.1845
DUMMULT	879.92	260.81	186	3.37	0.0009
ICGMUL2	1317.19	568.10	186	2.32	0.0215

Covariance structure: stationary *m*–dependent, *m* = 2 (banded Toeplitz)

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	2656.70	289.62	186	9.17	<.0001
NSTART	-0.1789	0.03399	186	-5.27	<.0001
DURRAN	-0.4402	0.1263	186	-3.49	0.0006
DUMCGRP1	-397.22	177.40	186	-2.24	0.0263
DUMCGRP2	274.81	304.30	186	0.90	0.3677
DUMENG	-412.62	255.00	378	-1.62	0.1065
PRESENTD	-345.69	75.3982	378	-4.58	<.0001
DUMCLN1	370.90	121.08	378	3.06	0.0023
DUMCLN2	319.05	112.67	378	2.83	0.0049
DUMANS	272.23	84.9148	378	3.21	0.0015
DEVLPNG	2274.04	486.10	186	4.68	<.0001
NTREP	139.47	48.0213	378	2.90	0.0039
NRREP	-58.5385	29.4615	378	-1.99	0.0476
IMPACT	-17.8250	8.5695	378	-2.08	0.0382
DUMPUBT3	-307.83	114.98	378	-2.68	0.0077
DUMMULT	782.77	205.68	186	3.81	0.0002
ICGMUL2	1202.49	447.82	186	2.69	0.0079

Covariance structure: stationary *m*–dependent, *m* = 1 (banded Toeplitz)

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	2344.83	266.14	186	8.81	<.0001
NSTART	-0.1508	0.02855	186	-5.28	<.0001
DURRAN	-0.4398	0.1041	186	-4.23	<.0001
DUMCGRP1	-535.11	149.11	186	-3.59	0.0004
DUMCGRP2	23.5089	253.83	186	0.09	0.9263
DUMENG	-1048.55	364.27	378	-2.88	0.0042
PRESENTD	-637.95	118.34	378	-5.39	<.0001
DUMCLN1	444.48	149.16	378	2.98	0.0031
DUMCLN2	438.23	145.80	378	3.01	0.0028
DUMANS	523.72	122.08	378	4.29	<.0001
DEVLPNG	2182.69	423.09	186	5.16	<.0001
NTREP	243.98	51.8602	378	4.70	<.0001
NRREP	-91.0305	29.3928	378	-3.10	0.0021
IMPACT	-25.6206	11.4173	378	-2.24	0.0254
DUMPUBT3	-354.65	158.10	378	-2.24	0.0255
DUMMULT	813.22	163.45	186	4.98	<.0001
ICGMUL2	1226.46	355.92	186	3.45	0.0007

APPENDIX XVIII: OUTPUT FROM GEE ANALYSIS APPLIED TO THE FIVE OTHER DATASETS USING THE FINAL CHOICE OF CORRELATION STRUCTURE, STATIONARY *M*-DEPENDENT, *M*=2

(I) Data used: all mentions
Response variable: time from close to submission

The GENMOD Procedure

Algorithm converged.

Working Correlation Matrix

	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
Row1	1.0000	0.5276	0.4617	0.0000	0.0000	0.0000	0.0000	0.0000
Row2	0.5276	1.0000	0.5276	0.4617	0.0000	0.0000	0.0000	0.0000
Row3	0.4617	0.5276	1.0000	0.5276	0.4617	0.0000	0.0000	0.0000
Row4	0.0000	0.4617	0.5276	1.0000	0.5276	0.4617	0.0000	0.0000
Row5	0.0000	0.0000	0.4617	0.5276	1.0000	0.5276	0.4617	0.0000
Row6	0.0000	0.0000	0.0000	0.4617	0.5276	1.0000	0.5276	0.4617
Row7	0.0000	0.0000	0.0000	0.0000	0.4617	0.5276	1.0000	0.5276
Row8	0.0000	0.0000	0.0000	0.0000	0.0000	0.4617	0.5276	1.0000
Row9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.4617	0.5276
Row10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.4617
Row11	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Row12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Col9	Col10	Col11	Col12
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.4617	0.0000	0.0000	0.0000
0.5276	0.4617	0.0000	0.0000
1.0000	0.5276	0.4617	0.0000
0.5276	1.0000	0.5276	0.4617
0.4617	0.5276	1.0000	0.5276
0.0000	0.4617	0.5276	1.0000

Analysis Of GEE Parameter Estimates

Model-Based Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	3178.243	340.8240	2510.240	3846.246	9.33	<.0001
DURRAN	-0.6232	0.1956	-1.0066	-0.2398	-3.19	0.0014
IMPACT	-104.060	30.8414	-164.508	-43.6115	-3.37	0.0007
LOGPEST	257.3812	68.4086	123.3028	391.4596	3.76	0.0002
DEVLPNG	4370.561	516.3852	3358.465	5382.657	8.46	<.0001
LOGNTREP	871.3414	414.7244	58.4965	1684.186	2.10	0.0356
NRREP	-147.811	45.7156	-237.412	-58.2102	-3.23	0.0012
Scale	1265.495	.	.	.	.	.

NOTE: The scale parameter for GEE estimation was computed as the square root of the normalized Pearson's chi-square.

(II) Data used: all mentions
Response variable: √(time from receipt to publication)

Algorithm converged

Working Correlation Matrix

	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
Row1	1.0000	0.0006	-0.2442	0.0000	0.0000	0.0000	0.0000	0.0000
Row2	0.0006	1.0000	0.0006	-0.2442	0.0000	0.0000	0.0000	0.0000
Row3	-0.2442	0.0006	1.0000	0.0006	-0.2442	0.0000	0.0000	0.0000
Row4	0.0000	-0.2442	0.0006	1.0000	0.0006	-0.2442	0.0000	0.0000
Row5	0.0000	0.0000	-0.2442	0.0006	1.0000	0.0006	-0.2442	0.0000
Row6	0.0000	0.0000	0.0000	-0.2442	0.0006	1.0000	0.0006	-0.2442
Row7	0.0000	0.0000	0.0000	0.0000	-0.2442	0.0006	1.0000	0.0006
Row8	0.0000	0.0000	0.0000	0.0000	0.0000	-0.2442	0.0006	1.0000
Row9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.2442	0.0006
Row10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.2442
Row11	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Row12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Col9	Col10	Col11	Col12
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
-0.2442	0.0000	0.0000	0.0000
0.0006	-0.2442	0.0000	0.0000
1.0000	0.0006	-0.2442	0.0000
0.0006	1.0000	0.0006	-0.2442
-0.2442	0.0006	1.0000	0.0006
0.0000	-0.2442	0.0006	1.0000

Indicator variables

DUMJGRP2 = 1 if JGROUP = Other
DUMJGRP2 = 0 if JGROUP = America or Europe

DUMCGRP1 = 1 if CGROUP = Europe
DUMCGRP1 = 0 if CGROUP = America or Other

Analysis Of GEE Parameter Estimates

Model-Based Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	12.3444	1.0641	10.2589	14.4300	11.60	<.0001
NSTART	0.0006	0.0001	0.0004	0.0008	6.07	<.0001
DUMJGRP2	-12.8175	1.7665	-16.2798	-9.3551	-7.26	<.0001
DURRAN	0.0017	0.0005	0.0008	0.0026	3.73	0.0002
NRREP	-0.6469	0.0851	-0.8136	-0.4801	-7.60	<.0001
DUMCGRP1	-3.3324	0.6392	-4.5852	-2.0797	-5.21	<.0001
LOGAUTH	2.2717	0.8689	0.5686	3.9748	2.61	0.0089
Scale	3.1934	.	.	.	.	.

NOTE: The scale parameter for GEE estimation was computed as the square root of the normalized Pearson's chi-square.

(III) Data used: all mentions
Response variable: time from close to publication

The output for this analysis has been given in Appendix XV.

(IV) Data used: all reportings of results
Response variable: time from close to submission

Algorithm converged.

Working Correlation Matrix

	Col1	Col2	Col3	Col4	Col5	Col6
Row1	1.0000	0.5657	0.0650	0.0000	0.0000	0.0000
Row2	0.5657	1.0000	0.5657	0.0650	0.0000	0.0000
Row3	0.0650	0.5657	1.0000	0.0000	0.5657	0.0650
Row4	0.0000	0.0650	0.5657	1.0000	0.5657	0.0650
Row5	0.0000	0.0000	0.0650	0.5657	1.0000	0.5657
Row6	0.0000	0.0000	0.0000	0.0650	0.5657	1.0000
Row7	0.0000	0.0000	0.0000	0.0000	0.0650	0.5657
Row8	0.0000	0.0000	0.0000	0.0000	0.0000	0.0650

Col7	Col8
0.0650	0.0000
0.5657	0.0650
1.0000	0.5657
0.5657	1.0000

Indicator variables

DUMPN = 1 if POSNG = X (not reported)
DUMPN = 0 if POSNG = Null, Negative, Opposite or Positive

Analysis Of GEE Parameter Estimates

Model-Based Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	2011.066	185.8980	1646.712	2375.419	10.82	<.0001
NOQ	95.0424	32.5245	31.2956	158.7893	2.92	0.0035
PRESENTD	-599.944	190.4868	-973.291	-226.597	-3.15	0.0016
IMPACT	-93.3368	32.4989	-157.033	-29.6402	-2.87	0.0041
LOGPEST	236.0925	60.5592	117.3987	354.7863	3.90	<.0001
DUMPN	624.4268	311.3094	14.2717	1234.582	2.01	0.0449
Scale	1059.887	.	.	.	.	.

NOTE: The scale parameter for GEE estimation was computed as the square root of the normalized Pearson's chi-square.

(V) Data used: all reportings of results
Response variable: time from receipt to publication

Algorithm converged.

Working Correlation Matrix

	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
Row1	1.0000	-0.5996	0.0805	0.0000	0.0000	0.0000	0.0000	0.0000
Row2	-0.5996	1.0000	-0.5996	0.0805	0.0000	0.0000	0.0000	0.0000
Row3	0.0805	-0.5996	1.0000	-0.5996	0.0805	0.0000	0.0000	0.0000
Row4	0.0000	0.0805	-0.5996	1.0000	-0.5996	0.0805	0.0000	0.0000
Row5	0.0000	0.0000	0.0805	-0.5996	1.0000	-0.5996	0.0805	0.0000
Row6	0.0000	0.0000	0.0000	0.0805	-0.5996	1.0000	-0.5996	0.0805
Row7	0.0000	0.0000	0.0000	0.0000	0.0805	-0.5996	1.0000	-0.5996
Row8	0.0000	0.0000	0.0000	0.0000	0.0000	0.0805	-0.5996	1.0000

Indicator variables

DUMMULT = 1 if MULTIC = Yes or Limited (multi-centre)
DUMMULT = 0 if MULTIC = No (single-centre)

DUMINT = 1 if INTERNL = Yes (international)
DUMINT = 0 if INTERNL = Limited or No (single-country)

DUMANS1 = 1 if ANSWER = Yes
DUMANS1 = 0 if ANSWER = No or X (not reported)

DUMJGRP1 = 1 if JGROUP = Europe
DUMJGRP1 = 0 if JGROUP = America or Other

DUMJGRP2 = 1 if JGROUP = Other
DUMJGRP2 = 0 if JGROUP = America or Europe

Analysis Of GEE Parameter Estimates

Model-Based Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	130.8163	35.3986	61.4363	200.1962	3.70	0.0002
DUMINT	-417.780	99.9902	-613.758	-221.803	-4.18	<.0001
DUMMULT	132.8854	24.7516	84.3732	181.3977	5.37	<.0001
DUMJGRP1	-160.416	32.3132	-223.749	-97.0835	-4.96	<.0001
DUMJGRP2	-385.418	72.1713	-526.871	-243.965	-5.34	<.0001
DURRAN	0.1198	0.0178	0.0850	0.1545	6.75	<.0001
DUMANS1	60.9252	21.8332	18.1328	103.7176	2.79	0.0053
IMPACT	-8.0550	4.8317	-17.5250	1.4150	-1.67	0.0955
DUMINT*IMPACT	48.4723	13.9574	21.1163	75.8283	3.47	0.0005
Scale	119.6587	.	.	.	.	.

NOTE: The scale parameter for GEE estimation was computed as the square root of the normalized Pearson's chi-square.

(VI) Data used: all reportings of results
Response variable: time from close to publication

Algorithm converged.

Working Correlation Matrix

	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
Row1	1.0000	0.5526	0.1911	0.0000	0.0000	0.0000	0.0000	0.0000
Row2	0.5526	1.0000	0.5526	0.1911	0.0000	0.0000	0.0000	0.0000
Row3	0.1911	0.5526	1.0000	0.5526	0.1911	0.0000	0.0000	0.0000
Row4	0.0000	0.1911	0.5526	1.0000	0.5526	0.1911	0.0000	0.0000
Row5	0.0000	0.0000	0.1911	0.5526	1.0000	0.5526	0.1911	0.0000
Row6	0.0000	0.0000	0.0000	0.1911	0.5526	1.0000	0.5526	0.1911
Row7	0.0000	0.0000	0.0000	0.0000	0.1911	0.5526	1.0000	0.5526
Row8	0.0000	0.0000	0.0000	0.0000	0.0000	0.1911	0.5526	1.0000

Indicator variables

DUMANS = 1 if ANSWER = Yes or No
DUMANS = 0 if X (not reported)

DUMDIR1 = 1 if POSNG = Positive
DUMDIR1 = 0 if POSNG = Negative, Null, Opposite or X (not reported)

DUMDIR2 = 1 if POSNG = Negative
DUMDIR2 = 0 if POSNG = Positive, Null, Opposite or X (not reported)

Analysis Of GEE Parameter Estimates

Model-Based Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	2334.615	304.2938	1738.210	2931.019	7.67	<.0001
NCLOSE	-0.1281	0.0333	-0.1934	-0.0629	-3.85	0.0001
PRESENTD	-355.426	111.2406	-573.454	-137.399	-3.20	0.0014
DUMANS	302.9117	120.3815	66.9682	538.8551	2.52	0.0119
NTREP	108.7966	40.3946	29.6246	187.9685	2.69	0.0071
LOGPEST	313.6706	66.0366	184.2413	443.0999	4.75	<.0001
DUMDIR1	241.5361	147.3340	-47.2332	530.3055	1.64	0.1011
DUMDIR2	377.0421	183.5213	17.3469	736.7373	2.05	0.0399
LOGPEST*DUMDIR1	-328.093	85.6308	-495.926	-160.260	-3.83	0.0001
LOGPEST*DUMDIR2	-325.913	118.8067	-558.769	-93.0558	-2.74	0.0061
Scale	1262.351	.	.	.	.	.

NOTE: The scale parameter for GEE estimation was computed as the square root of the normalized Pearson's chi-square.

APPENDIX XIX: THE RESULTS OF THE 'HOW LONG?' ANALYSES IN THE CONTEXT OF OTHER STUDIES IDENTIFIED IN A LITERATURE SEARCH

REFERENCE SET OF TRIALS USED	**OUTCOME MEASURE(S)** TRIAL CHARACTERISTICS COLLECTED	RESULTS
Liebeskind et al (1999) (abstract) All controlled clinical trials in acute ischaemic stroke reported in English from 1957-1997, compiled from a computerized MEDLINE search and international trials registers. 127 reports describing 52 interventions.	**# Start of accrual period to publication** **Submission of article to acceptance for publication** Combination of statistical significance and direction of result such that *Positive = significant result in favour of experimental arm *Negative = non significant result or significant result in favour of control arm Number of patients Timing of publication Funding: charity vs. corporate vs. not known	Time from start to publication shorter for positive than for negative trials (median 3.5 vs. 4.4 yrs) log rank p=0.079 For subgroup of corporate funded trials difference is greater (median 3.5 vs. 4.7 yrs) Time from submission to acceptance also shorter for positive than for negative trials (median 0.17 vs. 0.44 yrs) p=0.011
Misakian and Bero (1998) (full paper) Studies on passive smoking in humans and animals funded between 1981 and 1995. Source: 78 investigators contacted, 65 responded, who had conducted 61 studies, of which 47 had been published.	**# Year funding began to year of publication of first results (funding start date used as completion date difficult to define)** Statistically significant results vs. not *Statistically significant = statistically significant association between passive smoking and a harmful health effect *Not statistically significant = results do not indicate a statistically significant association *Mixed=multiple primary outcomes measured and at least one was statistically significant Statistical significance defined as p<0.05 or odds ratio with confidence interval not overlapping 1. Experimental design vs. observational Animal vs. human studies Number of patients <=500 vs. >500 Health outcome measure Funding source: government vs. private vs. tobacco industry vs. other Funding: external vs. internal, <=5 yrs vs. >5yrs	Median time to publication 5 yrs (95% CI, 4-7 yrs) for non-significant results and 3 yrs (95% CI 3-5 yrs) for statistically significant results. p=0.004 Statistically significant results (p=0.004), experimental study design (p=0.01), study size <=500 (p=0.01) and animals as subjects (p=0.03) were predictive of time to publication (proportional hazards models or nonparametric Wilcoxon test). Studies of human subjects only; only statistical significance is predictive of publication (p=0.007) Multivariate analysis of all studies: statistical significance (p=0.001) and study design (p=0.01) were only independent predictors.

Cheng et al (1998) (abstract) The 199 abstracts describing 180 randomized controlled trials in cystic fibrosis listed in Cochrane CF Group's register at end of 1995. Abstracts initially identified from hand-searching 44 meeting abstract books over 30 yrs. Register searched for subsequent full reports. Only 32% of 178 abstracts analysed were subsequently published in full	**Time from publication as abstract to publication as full report** Experimental treatment as effective or better than control vs. not Number of patients	No difference in time to publication when abstracts stratified according to conclusions or sample size. No significant association (p>0.05) between time to publication and both sample size and conclusions together. Method: log rank test.
Stern and Simes (1997) (full paper) Cohort of 748 eligible studies submitted to Royal Prince Alfred Hospital Ethics Committee, Australia, between 1979 and 1988 were examined retrospectively.	**# Time from approval by ethics committee to first publication** For quantitative studies: *Statistical significance of results defined as positive (p<0.05) vs. negative (p>=0.10) versus indefinite conclusions (0.05<=p<0.10) For qualitative studies: Subjective assessment by the principal investigator: striking, important & definite vs. unimportant & negative findings Research design using trial vs. non-trial design Science importance rating of study by investigator Funding: pharmaceutical vs. non-pharmaceutical and external vs. internal vs. none Studies with non-comparable study groups Clinical trials randomized or not Study is part of degree or not Single- vs. multi-centre data collection sites Sample size <100 vs. >=100 Research department undertaking the study Year of study approval Outcome qualitative vs. quantitative For clinical trials placebo control & blinding vs. not	Of the 218 studies analysed with test of significance, those with positive results had shorter time to publication than those with negative results, median 4.8 vs. 8.0 yrs. Finding stronger for the subgroup of 130 clinical trials, with median times 4.7 and 8.0 yrs. Results not materially changed after adjusting for other significant predictors. Studies with indefinite conclusions took even longer time to publication than studies with negative results (median not yet reached). For 103 studies where outcome rated qualitatively: No clear evidence. Method: Cox regression.
Handysides (1996) (abstract) The 48 papers published in the Communicable Disease Report Review in 1995.	**Time from receipt by journal to publication**	Mean time from receipt to publication is 5 mths. Abstract does not compare by trial characteristics. Full paper not available.

Dickersin and Min (1993) (abstract) Meta-analysis of the authors' study plus 3 others found in literature search. All have prospective design, followed cohort of health-related research projects, started around same time, used similar study design and data collection forms.	Study findings: *Results 'significant' if judged by investigator to be either statistically significant or 'of great importance' Publication status: publication as abstract, journal article, book chapter or letter to the editor.	No results given on time to publication. However refers to **Stern and Simes (1997)** previously described and to **Ioannidis (1998)** (see below), stating " … it takes longer for 'negative' studies to be published than 'positive' studies."
Chew (1991) (abstract) Rejected manuscripts that were originally submitted to the American Journal of Roentgenology (AJR) during the first 5 months of 1986. MEDLINE searches conducted 45-54 months after dates of rejection by AJR located 162 (64%) published papers out of a consecutive series of 254 manuscripts rejected by AJR (69% rejected major papers and 62% rejected case reports). The papers had been published in 30 different radiologic and 27 non-radiologic journals.	**Time from rejection from AJR to publication elsewhere**	Mean time lapse between rejection by AJR and publication elsewhere is 15 mths. The delay in publication was greater for papers published in non-radiologic and foreign journals than for papers published in radiologic and American journals. Most of the journals of eventual publication published fewer papers, had smaller circulations and had lower impact factors than the AJR.
Ioannidis (1998) (full paper) Prospective cohort of randomized phase 2 and 3 trials conducted by two multi-centre trialist groups from 1986 to 1996. 109 efficacy trials in human immunodeficiency virus infection sponsored by National Institutes of Health of which 101 were eligible for analysis.	**Time from start of enrolment to completion of follow-up** **Time from completion of follow-up to publication** **# Time from start of enrolment to publication** *Statistically significant findings in favour of an experimental arm vs. of control arm vs. of neither arm vs. results pending Sample size >1000 vs. 200-1000 vs. <200 Accrual to target ratio <0.5 vs. >=0.5 Trialist group CPCRA vs. ACTG Population paediatric vs. adult Domain: Antiretroviral treatment vs. Complications of HIV Double-blind design vs. not Data management: Pharmaceutical industry vs. Other federally sponsored	Median time from start of enrolment to publication was 5.5 yrs, substantially longer for negative trials than for results favouring an experimental arm (6.5 vs. 4.3 yrs p<0.001, hazard ratio for time to publication for positive vs. negative trials 3.7;95% CI 1.8-7.7), mostly attributable to differences in time from completion of follow-up to publication (median 3.0 vs. 1.7 yrs p<0.001). On average, trials with significant results favouring any arm completed follow-up slightly earlier than trials with non significant results (median 2.3 vs. 2.5 yrs p=0.045), but long-protracted trials often had low event rates and failed to reach statistical significance, while trials that were terminated early had significant results. Positive trials were submitted for publication significantly more rapidly after completion than were negative trials (median 1.0 vs. 1.6 yrs p=0.001) and were published more rapidly after submission (median 0.8 vs. 1.1 yrs p=0.04).

This project All properly randomized trials which began before 1 January 1988** and all publications referring to these trials, as identified by the Cancer Overviews Group, which were published prior to 1/1/2000. (**This was the set of trials included in the Second International Collaborative Workshop on Childhood ALL Studies at the end of 1992) Trials still open to randomization are included, as well as those that have closed.	**For article first reporting results:** **Time from close to submission of article** **Time from receipt to publication** **Time from close to publication** The following characteristics had few enough missing values to be included and preliminary graphs suggested they may have an effect: Funding source Number of randomizations Number of arms Number of questions Number of patients Start and close dates of accrual period Duration of accrual period Type of treatment (chemotherapy, radiotherapy, transplant, antibiotic) First-line vs. relapse/refractory Equivalence trial vs. not Age eligibility: Children vs. children + adults Multi-centre (>=5) vs. limited (2-4) vs. single-centre Target number of centres reached vs. not International vs. limited vs. single-country Country group of trialist Conducted in 'developing' country vs. 'developed' Statistical significance (numerical measure based on spacings between $\log_e$ (typical value for category) Statistical significance not reported vs. reported Direction of results: + vs. – vs. null vs. opposite (i.e. + and -) vs. not reported Clinical significance of results: Yes vs. possibly vs. no vs. not reported Main questions answered vs. not Subgroup results reported vs. not Reported in journal, book or as meeting abstract Impact factor of journal	Time from close to submission is longer for randomizations with the following characteristics: Earlier close date (p=0.0036) Multi-centre participation (p=0.0063) Conducted in a European country (p=0.0008) Conducted in a country other than North America or Europe (p=0.0007), the latter having a greater effect than the former. Published in a high impact factor journal (p=0.0242) Direction of result not negative (p=0.0147) Direction of result not positive or null (p=0.0107), i.e. negative are published faster than positive/ null results, with opposite results and randomizations where results are not reported taking longest. (F statistic=5.36, *p*-value=0.0002, R^2=0.460099 based on all 52 observations) Time from receipt to publication is longer for randomizations with the following characteristics: Funded by charity as well as Government money (p=0.0003) Reported in an article which mentions a **larger** number of trials (p=0.0001) Reported in an article which mentions a **smaller** number of randomizations (p<0.0001) Has been presented at a major meeting (p=0.0032) Results not clearly reported as clinically significant (p=0.0056) Results not clearly reported as **not** clinically significant (p=0.0040) i.e. results not clinically significant are published faster than those which are, with those randomizations for which clinical significance is not reported taking longest. (F statistic=9.81, *p*-value<0.0001, R^2=0.583463 based on 49 out of 60 observations)

	Journal/book/abstract has no impact factor vs. has Country group of publisher Country group of trialists and publisher same vs. not Article published in English vs. not Results presented at major meeting? Number of trials mentioned in article Number of randomizations mentioned in article Number of authors These variables were considered but ruled out due to a high proportion of missing values and/or on the basis of preliminary graphs: Length of follow-up period Eligibility risk group Target accrual reached vs. not Correct timing of a late randomization vs. not Method of randomization: central computer vs. notification central office vs. sealed envelopes Randomization design: simple vs. block vs. minimization of imbalance Attempt to balance vs. not Statistical technique used	Time from close to publication is longer for randomizations with the following characteristics: Main effects: Short duration of accrual period (p=0.0002) No clear indication of whether clinically significant or not is given (p=0.0060) Reported in a journal with an impact factor associated with it (p<0.0001) Results less statistically significant (0.0153) Interactions: Conducted by European trialists and reported in a publication with a high impact factor (p=0.0006) Conducted outside Europe and reported in a publication with a lower impact factor (p=0.0062) Conducted in North America or Europe and results not null (p=0.0127) Conducted elsewhere and results are null (p=0.0023) (F statistic=9.10, *p*-value<0.0001, R^2=0.364055 based on all 170 observations)

Note: * definition of statistical significance differs from that used in this project
definition of 'time to publication' differs from that used in this project (i.e. time from close of accrual period to publication)

APPENDIX XX: ALGORITHMS AND SAS COMMANDS FOR PERFORMING THE 'HOW WIDE?' ANALYSES

(i) Algorithm for obtaining a suitable dataset for addressing the 'How wide?' questions i.e. ever/never reported and frequency of reporting

- The last part of the main data management program (JMAIN.SAS) outputs 2 datasets – one containing all records, the other all records which contain results. In each of these, there is an order number of the publication record relating to the randomization to which it belongs (ALLORDER in the all records dataset, and RESORDER in the results dataset). In each case retain only the randomization ID (RANID) and the order number. Records are ordered by RANID and within that by ALLORDER/RESORDER.
- From each database keep only the last record for each randomization. Rename ALLORDER as NMENT (the number of articles in which each randomization is mentioned) and RESORDER as NRES (the number of articles in which the results of each randomization are reported)
- Merge both of the above with the definitive records database, by variable RANID
- If NMENT is missing set NMENT to zero
  If NRES is missing set NRES to zero
- Create new variables MENTND (ever mentioned) and RESPUB (results ever published) using NMENT and RESPUB

(ii) SAS commands for performing logistic regression

```
PROC LOGISTIC DESCENDING;
  CLASS DEVLPNG FIRSTL;
  MODEL MENTND= NSTART FIRSTL DEVLPNG DURRAN/EXPB;
RUN;
```

Notes
1. Select response level of interest to be reported=yes by using DESCENDING
2. Command EXPB outputs the exponentiated values of estimates

(iii) SAS commands for performing Poisson or negative binomial regression

```
PROC GENMOD;
  CLASS DEVLPNG DUMMULT3 EQUIV FIRSTL DUMCGRP3;
  MODEL NMENT= NSTART DEVLPNG DUMMULT3 LOGSIZE EQUIV FIRSTL
    DUMCGRP3/DIST=NEGBIN OFFSET= LNSTCUT LINK=LOG;
  OUTPUT OUT=TEMP3 P=YHAT RESCHI=RESID STDRESCHI=STRESID;
RUN;
```

Notes
1. Set DIST = NEGBIN for the negative binomial model
   DIST = POISSON for the Poisson distribution
   For both, the link function, LINK=LOG
2. Offset variable LNSTCUT = ln (CUTANAL - NSTART)
   i.e. natural log of the time between the date the randomization opened for recruitment and the cut-off date for analysis (28/11/00). For the frequency of mentions analysis, start date chosen rather than close date so that mentions before randomization has closed will fall within this period. For the frequency of reportings of results analysis close date was used.
3. Output statement provides fitted values (YHAT), residuals (RESID) and standardised residuals (STRESID) for testing model assumptions

APPENDIX XXI: OUTPUT FROM NEGATIVE BINOMIAL AND POISSON REGRESSIONS TO MODEL (a) FREQUENCY OF MENTIONS AND (b) FREQUENCY OF REPORTING RESULTS

(a) Response = frequency of mentions
Model = negative binomial

The GENMOD Procedure

Model Information

Data Set	LEUKJR.DEFIN
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	NMENT
Offset Variable	LNSTCUT
Observations Used	188
Missing Values	55

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	180	185.3164	1.0295
Scaled Deviance	180	185.3164	1.0295
Pearson Chi-Square	180	178.3175	0.9907
Scaled Pearson X2	180	178.3175	0.9907
Log Likelihood		51.8054	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	-11.6794	0.4778	-12.6158	-10.7430	597.62	<.0001
NSTART		1	0.0001	0.0000	0.0001	0.0002	30.89	<.0001
DEVLPNG	0	1	0.9101	0.3043	0.3136	1.5065	8.94	0.0028
DEVLPNG	1	0	0.0000	0.0000	0.0000	0.0000	.	.
DUMMULT3	0	1	0.4908	0.1438	0.2090	0.7725	11.66	0.0006
DUMMULT3	1	0	0.0000	0.0000	0.0000	0.0000	.	.
LOGSIZE		1	0.6256	0.1261	0.3785	0.8727	24.62	<.0001
EQUIV	0	1	-0.2538	0.1076	-0.4647	-0.0430	5.57	0.0183
EQUIV	1	0	0.0000	0.0000	0.0000	0.0000	.	.
FIRSTL	1	1	0.5924	0.2230	0.1553	1.0296	7.06	0.0079
FIRSTL	2	0	0.0000	0.0000	0.0000	0.0000	.	.
DUMCGRP3	0	1	0.2625	0.1095	0.0479	0.4771	5.75	0.0165
DUMCGRP3	1	0	0.0000	0.0000	0.0000	0.0000	.	.
Dispersion		1	0.0214	0.0363	0.0008	0.5903		

NOTE: The negative binomial dispersion parameter was estimated by maximum likelihood.

Indicator variables

DUMMULT3 = 1 if MULTIC = Yes or Limited (multi-centre)
DUMMULT3 = 0 if MULTIC = No (single-centre)

DUMCGRP3 = 1 if CGROUP = Europe or Other
DUMCGRP3 = 0 if CGROUP = America

Model = Poisson

The GENMOD Procedure

Model Information

Data Set	LEUKJR.DEFIN
Distribution	Poisson
Link Function	Log
Dependent Variable	NMENT
Offset Variable	LNSTCUT
Observations Used	188
Missing Values	55

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	180	196.2034	1.0900
Scaled Deviance	180	196.2034	1.0900
Pearson Chi-Square	180	189.3096	1.0517
Scaled Pearson X2	180	189.3096	1.0517
Log Likelihood		51.6094	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	-11.6814	0.4668	-12.5963	-10.7665	626.26	<.0001
NSTART		1	0.0001	0.0000	0.0001	0.0002	32.90	<.0001
DEVLPNG	0	1	0.9082	0.3004	0.3194	1.4970	9.14	0.0025
DEVLPNG	1	0	0.0000	0.0000	0.0000	0.0000	.	.
DUMMULT3	0	1	0.4925	0.1382	0.2217	0.7634	12.71	0.0004
DUMMULT3	1	0	0.0000	0.0000	0.0000	0.0000	.	.
LOGSIZE		1	0.6259	0.1220	0.3868	0.8650	26.32	<.0001
EQUIV	0	1	-0.2546	0.1033	-0.4571	-0.0521	6.07	0.0137
EQUIV	1	0	0.0000	0.0000	0.0000	0.0000	.	.
FIRSTL	1	1	0.5943	0.2193	0.1645	1.0241	7.34	0.0067
FIRSTL	2	0	0.0000	0.0000	0.0000	0.0000	.	.
DUMCGRP3	0	1	0.2632	0.1061	0.0552	0.4711	6.15	0.0131
DUMCGRP3	1	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

(b) Response = frequency of reporting results
Model = Poisson

The GENMOD Procedure

Model Information

Data Set	LEUKJR.DEFIN
Distribution	Poisson
Link Function	Log
Dependent Variable	NRES
Offset Variable	LNCLCUT
Observations Used	188
Missing Values	55

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	182	193.6045	1.0638
Scaled Deviance	182	193.6045	1.0638
Pearson Chi-Square	182	173.5242	0.9534
Scaled Pearson X2	182	173.5242	0.9534
Log Likelihood		-97.5575	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	-11.9120	0.5721	-13.0334	-10.7907	433.50	<.0001
NCLOSE		1	0.0001	0.0000	0.0001	0.0002	30.55	<.0001
DEVLPNG	0	1	1.5879	0.4518	0.7023	2.4734	12.35	0.0004
DEVLPNG	1	0	0.0000	0.0000	0.0000	0.0000	.	.
DUMMULT3	0	1	0.7270	0.1396	0.4533	1.0006	27.10	<.0001
DUMMULT3	1	0	0.0000	0.0000	0.0000	0.0000	.	.
LOGSIZE		1	0.4545	0.1332	0.1935	0.7154	11.65	0.0006
TXCHEMO4	B	1	0.3315	0.1120	0.1120	0.5511	8.76	0.0031
TXCHEMO4	D	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

Indicator variables

DUMMULT3 = 1 if MULTIC = Yes or Limited (multi-centre)
DUMMULT3 = 0 if MULTIC = No (single-centre)

TXCHEMO4 = B if TXCHEMO = Immunotherapy or Radiotherapy
TXCHEMO4 = D if TXCHEMO = Chemotherapy or Antibiotic

Model = negative binomial

Note:
The output is identical to that for the Poisson model except for the following:

WARNING: Negative of Hessian not positive definite.

Analysis Of Parameter Estimates

Dispersion	0	0.0000	0.0000	0.0000	0.0000

NOTE: The negative binomial dispersion parameter was estimated by maximum Likelihood.